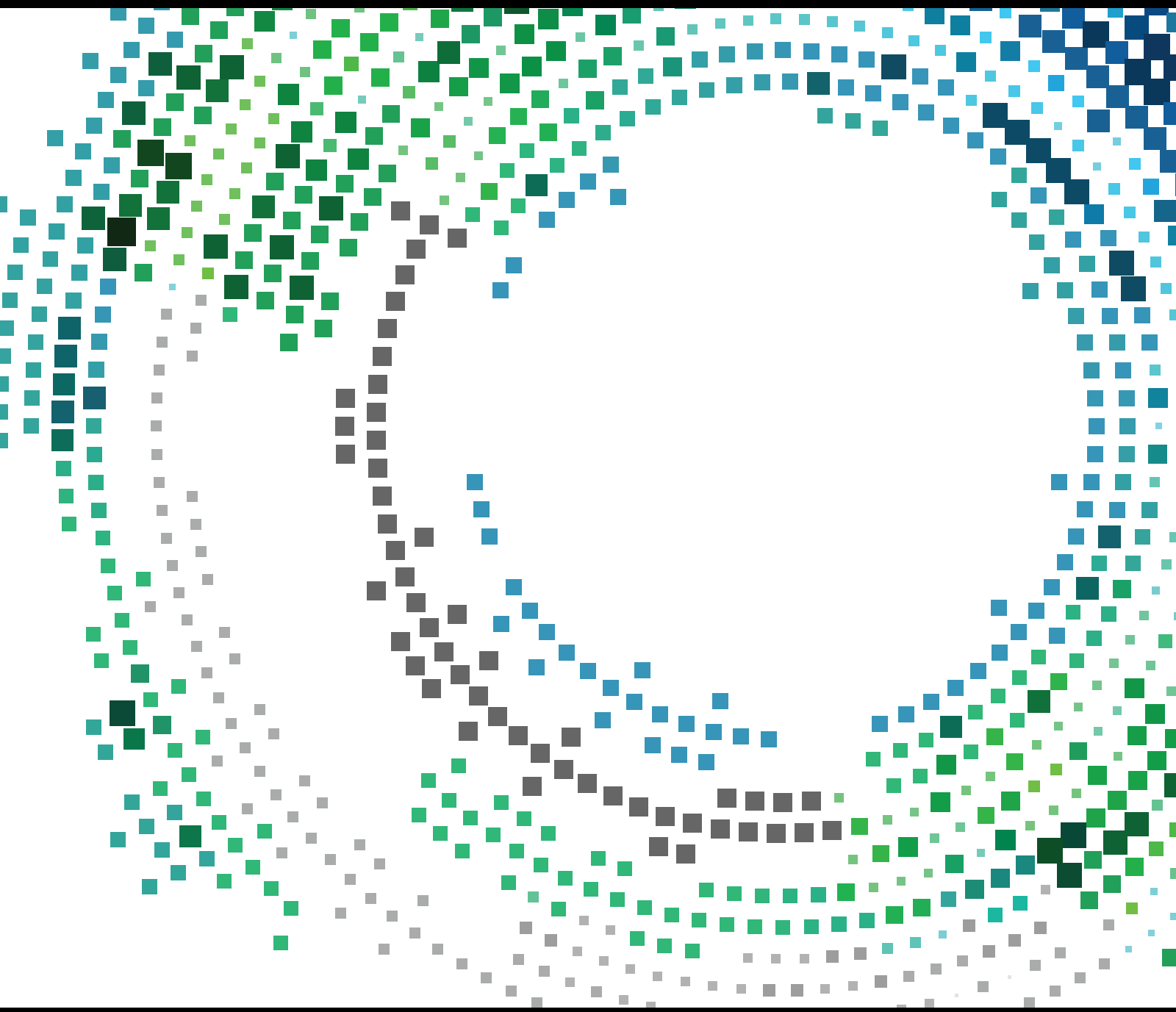
Mobile Information Systems

Wireless and Mobile Technologies for the Internet of Things

Guest Editors: Jong-Hyouk Lee, Kamal Deep Singh, Yassine Hadjadj-Aoul, and Neeraj Kumar



Wireless and Mobile Technologies for the Internet of Things

Mobile Information Systems

Wireless and Mobile Technologies for the Internet of Things

Guest Editors: Jong-Hyouk Lee, Kamal Deep Singh, Yassine Hadjadj-Aoul, and Neeraj Kumar

Contents

Hindawi Publishing Corporation
Mobile Information Systems
Volume 2016, Article ID 8206548, 2 pages
http://dx.doi.org/10.1155/2016/8206548

Editorial
Wireless and Mobile Technologies for the Internet of Things

Jong-Hyouk Lee,[1] Kamal Deep Singh,[2] Yassine Hadjadj-Aoul,[3] and Neeraj Kumar[4]

[1]*Department of Computer Science and Engineering, Sangmyung University, Cheonan, Republic of Korea*
[2]*Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai, India*
[3]*Universite de Rennes 1, Rennes, France*
[4]*Department of Computer Science and Engineering, Thapar University, Patiala, India*

Correspondence should be addressed to Jong-Hyouk Lee; jonghyouk@smu.ac.kr

Received 17 July 2016; Accepted 17 July 2016

Billions of interconnected devices that have limited capabilities in terms of computational power, memory, and battery are expected to soon form a new type of the Internet, called the Internet of Things (IoT). According to forecasts, the Internet will consist of over 50 billion connected things, that is, IoT devices, including televisions, cars, kitchen appliances, surveillance cameras, smartphones, utility meters, cardiac monitors, thermostats, and almost anything that we can imagine. This will in turn transform our ability to interact with real-world objects, process information, and make decisions in addition to saving us time and money. Already many high-tech companies over the world started developing IoT products and services and promoting their early stage of IoT products and services in all market domains. Among the most notable challenges, wireless and mobile technologies are the underlying technologies for realizing the IoT. Resource constrained devices are required to communicate with other devices in wireless networks. The devices are also required to communicate on the move. In addition to these requirements, various technical and scientific research considerations are also required.

This special issue gathers high quality research and development results of wireless and mobile technologies for the IoT. In one article, R.-S. Cheng et al. target localization and guidance in indoor settings. Such technologies are easily available for outdoor settings, but a requirement exists for indoor applications. They propose a guiding system combining GPS, Bluetooth low energy (BLE), and near field communication (NFC) technologies. In another article, A. A. Minhas et al. proposed a technique based on cognitive heterogeneous wireless grid to eliminate white spots in mobile WiMAX broadcasting regions. The idea of the technique is interesting in the sense that it uses the incorporation between mobile WiMAX and GSM, where the vertical handoff and switching towards GSM grid are triggered by mobile subscriber without suffering from the white spot problem. Y.-W. Kuo and L.-D. Chou focused on reducing power consumption of IoT devices. They proposed a fuzzy-based power saving scheduling scheme for IoT devices which communicate using LTE/LTE-A networks. J. Guan et al. propose a PMIPv6-based group binding solution to provide the mobility support for IoT devices. The proposed solution adopts the group movement character of IoT devices and modifies the registration procedure by extending the PBU and PBA messages to reduce the redundant binding operations. C. Zhang et al. proposed a new hierarchical update mining system, which can broadcast with useful, new, and timely sentence-length updates about a developing event. The proposed solution incorporates techniques from topic-level and sentence-level summarization, which allows extracting efficiently unexpected event. The solution has many applications in different topics like large human accident or natural disaster.

Two other papers target vehicular networks. C.-R. Dow et al. propose a taxi management system using location based services and zone queuing techniques which allow taxi drivers to look for passengers on the road or wait in the queuing zones. Y. Zou et al. focus on communication in vehicular networks by improving Direction of Arrival estimation algorithms.

In the article by D. Kim, the author proposes group interest based verification scheme for CCN using a transmission process to handle request messages at one time. The proposed verification method is designed to be suitable for the proposed transmission process. Also, this hybrid verification approach is first proposed so as to improve the computation overheads of a verification process.

Three papers focus on security and privacy issues. W. Ren et al. process a novel file access scheme for file storage services in mobile cloud computing services. S.-H. Chang et al. focus on mobile crowd sensing paradigm which leverages citizens for large-scale sensing by various mobile devices to efficiently collect and share local information. They focus on the network trustworthiness problem, namely, Sybil attacks in such network. In their article, Y. Kim et al. target privacy data leaks in mobile applications and propose a methodology and architecture for measuring user awareness of sensitive data leakage, which features runtime application analysis over timing distance between the user input event and actual privacy data leak.

In the article by J. Lee and J. Lee, they propose a duplicate name prefix detection mechanism to enhance the content source mobility in Content Centric Networking (CCN). H.-W. Wang focuses on RFID technology which is used for recognizing objects in Internet of Things (IoT). In order to improve system performance, H.-W. Wang proposes an efficient scheme to estimate the number of unidentified tags for Dynamic Framed Slotted Aloha (DFSA) based RFID system, with the view of increasing system performance. H. Kim et al. proposed a network application agent in order to overcome the issues related to the impossibility of controlling IoT devices beyond a certain threshold without using the devices' native language. The proposed solution, which is based on Cordova, is a wearable device control platform for the development of network applications, controls input/output functions of smartphones and wearable/IoT, and enables device control and information exchange by external users by offering a self-defined API.

In another article, X. Li et al. present a novel analytical model to investigate the eavesdropping attacks in Wireless Net of Things (WNoT). S. Kim proposed an effective resources' allocation technique, which maximizes the performance of dynamic environment of real-time IoT systems. The main idea behind the paper is to exploit learning based Markov game model to optimize dynamically the allocation strategy to the current system conditions. Besides, the proposed solution presents many possibilities of extension, notably in the context of uncertain system environments, in which the proposed iterative approach may help in optimizing the resources' usage. M. Chung and I. Ko propose a data sharing method among multi-smart devices at close range. The proposed method uses inaudible frequencies as a trigger signal and Wi-Fi and GPS information in order to address the shortcomings of existing data sharing methods, specifically data sharing using the Bump application. A. Ahmad et al. consider the challenge of managing mobility in M2M and IoT applications and propose a novel vertical handover scheme.

Acknowledgments

The guest editors would like to thank the significant contributors to this special issue including the authors and reviewers.

Jong-Hyouk Lee
Kamal Deep Singh
Yassine Hadjadj-Aoul
Neeraj Kumar

Hindawi Publishing Corporation
Mobile Information Systems
Volume 2016, Article ID 6506341, 10 pages
http://dx.doi.org/10.1155/2016/6506341

Research Article

Protecting Mobile Crowd Sensing against Sybil Attacks Using Cloud Based Trust Management System

Shih-Hao Chang and Zhi-Rong Chen

Department of Computer Science and Information Engineering, Tamkang University, New Taipei City 25137, Taiwan

Correspondence should be addressed to Shih-Hao Chang; shhchang@mail.tku.edu.tw

Received 7 August 2015; Accepted 16 February 2016

Academic Editor: Jong-Hyouk Lee

Mobile crowd sensing (MCS) arises as a new sensing paradigm, which leverages citizens for large-scale sensing by various mobile devices to efficiently collect and share local information. Unlike other MCS application challenges that consider user privacy and data trustworthiness, this study focuses on the network trustworthiness problem, namely, Sybil attacks in MCS network. The Sybil attack in computer security is a type of security attack, which illegally forges multiple identities in peer-to-peer networks, namely, Sybil identities. These Sybil identities will falsify multiple identities that negatively influence the effectiveness of sensing data in this MCS network or degrading entire network performance. To cope with this problem, a cloud based trust management scheme (CbTMS) was proposed to detect Sybil attacks in the MCS network. The CbTMS was proffered for performing active and passive checking scheme, in addition to the mobile PCS trustworthiness management, and includes a decision tree algorithm, to verify the covered nodes in the MCS network. Simulation studies show that our CbTMS can efficiently detect the malicious Sybil nodes in the network and cause 6.87 Wh power reduction compared with other malicious Sybil node attack mode.

1. Introduction

In recent years, mobile computing devices on the market, for example, smartphones and tablet computers, have become ubiquitous. Differing from the last century, the mobile phone of today, namely, the smartphone, usually comes with multifunction sensors, such as camera, microphone, GPS, accelerometer, digital compass, and gyroscope. These new technologies have enabled smartphone users to collect sensed data from their local information and upload these sensed data back to an application server using existing wireless communication infrastructure (such as 3G/4G/5G services or even WiMAX access points). Smartphones provide an excellent platform for mobile crowd sensing (MCS) [1]. Hence, a requester of data can create tasks that use the general public to capture geotagged images, videos, audio snippets, or all-out surveys. Participants who have installed the client apps on their smartphones can submit their data and get rewarded.

In recent years, a surfeit of novel and fascinating MCS applications have been developed, ranging from health care to multiple cultural aspects. Two examples of MCS applications are BALANCE [2] and HealthSense [3], used to collect and share data about personal health projects, which monitor the activities and behavior related to diet and encourage healthy living. MCS application provides a very open concept platform, which allows anybody to contribute their local sensing information; however, it may also leak malicious and erroneous attacks to the application. Sharing sensed data tagged with spatiotemporal information could reveal a lot of personal information, such as users' identity, personal activities, political views, and health status, thereby posing threats to the participating users. Malicious participants may unintentionally position the phone in an adverse position or deliberately contribute bad data while collecting sensor readings from mobile phones. Hence, an attacker can use these identities to act maliciously, by providing huge amount of corrupt data to either degrading data correctness or network performance via a Sybil attack in MCS application.

Despite the great number of works in network security literature, a systematic study and classification of the research problems of the Sybil attacks in MCS research domain is lacking to guide further research and development of this

emerging field. The Sybil attack was first introduced by Microsoft researcher Douceur, who described a Sybil attack, relying on the fact that a participatory sensing network possibly includes tremendous and unrealistic data from different resource parities and coordination among entities [4]. Sybil attacks relied on the fact that a MCS network data server cannot confirm that each unknown data-collecting element is a distinct mobile device. Therefore, any malicious MCS network attack can try to inject false information into the network to confuse or even collapse the network applications. Recently, some researchers have revealed Sybil attacks have compromised mobile social network [5] and Internet of Things (IoT) [6] because social network and IoT platform are both vulnerable to Sybil attacks, as Sybil attacks can manipulate pseudoidentities to compromise the effectiveness of social network and IoT system.

Cloud computing provides flexible and on-demand infrastructures which have drawn lots of attention from research and industry in recent years. Cloud computing services commonly denote functionalities such as IaaS (infrastructure as a service), SaaS (software as a service), and PaaS (platform as a service), delineated as a layered system structure for cloud computing. TaaS (trust as a service) decides which types of solutions are appropriate for their unique needs. Recently, several methodologies have been offered for trust management in cloud computing environments [7–9]. For example, CATRAC [9] has proposed security architecture related to combining web services. CATRAC combines both Role-Based Access Control and Trust-Based Access Control in order to arrive at an optimum solution. The authors in [9] described that trust levels are presented as a vector ranging from 0 to 10, indicating "fully distrusted" to "fully trusted," respectively. Five denotes an uncertainty or a neutral level, which is commonly assigned to new clients. Nevertheless, analyzing trust issues from a cloud user will normally respect their data in terms of security and privacy. Therefore, a good reputation system requires reducing alliance of user identification and his/her privacy information.

To solve this problem, a cloud based trust management scheme (CbTMS) is proposed to evaluate the Sybil attacks in MCS applications. The proposed CbTMS framework proffered active and passive checking schemes that leverage mobile PCS and base station to perform the Sybil identity detection over a period of time. Moreover, to address the trustworthiness issue in the proposed system, the CbTMS also provides a Trust Credit Assessment and analytical decision support that perform the trust management service in the cloud to evaluate mobile PCS trustworthiness level. Hence, a high credit score is an indication that a particular smartphone device has been reporting reliable PCS. To verify this idea, the OMNeT++ simulation has been used to present our CbTMS's effectiveness against Sybil attacks. The rest of this paper is organized as follows. Section 2 presents a literature review of current related works and summarizes their conclusions. Section 3 provides the detection factors motivating the need for a reputation system in the context of MCS; it presents an overview of the system architecture. In Section 4, the experimental setup and simulation results are described and Section 5 concludes the paper.

2. Background

In recent years, there have been more and more MCS applications in different fields. For example, in personal health monitoring, BALANCE [2] allows clients to monitor their activities and diet behavior, encouraging healthy living. Food calories are entered via mobile phones and an accelerator detects movement patterns and time to project the calories consumed, thereby achieving health management. HealthSense [3] automatically detects health-related events, such as pain or depression, which cannot be observed directly through current sensor technology. HealthSense analyzes sensor data from the patient by applying machine learning methods. HealthSense also utilizes patient input events to assist in data classification (such as pain or itching). Finally, the user provides feedback on the machine learning process. As mentioned, MCS applications are subject to malicious attacks.

Due to the MCS applications, participants allow anyone with an appropriate mobile device that has the application installed to register as a participant. Such human intervention entails serious security and privacy risks. The free transmission of users' sensor data could result in compromised security and privacy. For instance, users may leak their personal identity information through personal responses. In [5], the mobile Sybil detection is exploited based on mobile user's friend and foe list. Mobile users can detect Sybil attackers with profile matching when they are encountered. Liang et al. [6] explore Internet of Things (IoT) exposed to Sybil attacks where attackers can manipulate fake identities or abuse pseudoidentities to compromise the effectiveness of the IoT and even disseminate spam. Particularly, in [6], the authors also outline three types of Sybil attacks, namely, SA-1, SA-2, and SA-3, according to the Sybil attacker's capabilities. As a result, Sybil detection in research efforts is becoming more popular for the development of both online and mobile Sybil detection and defense schemes in social network and IoT system.

Douceur formalized the Sybil attack in the perspective of peer-to-peer networks [4]. He presented that there is no practical solution for this attack and indicated that Sybil attacks can overthrow the redundancy mechanisms of distributed data storage systems. Problems arise when a reputation system (such as a trusted certification) is tricked into thinking that an attacking computer has a disproportionally large influence. Grover et al. [10] proposed a scheme to protect against the Sybil attack using neighboring nodes' information. In this approach, every node will participate to detect the suspicious node in the network. Every mobile node has a different group of neighbors at different time interval. After these mobile nodes share their network tables, they will match their neighboring tables; if some nodes are simultaneously observed with the same set of neighbors at different interval of time, then these nodes are under Sybil attack. In this case, identities are neighboring nodes associated with specific trust devices. Similar to a central authority creating certificates, there are only few methods to prevent an attacker from reaching multiple devices.

Trust and reputation have been verified as influencing customers or users in selecting high quality service in multiple situations. The concept of trust and reputation is similar

in computational models that can be formally characterized based on history of past interactions. For instance, after the completion of the transaction of rating among parties, the aggregated ratings about a given party can then be used to derive its reputation score. Nevertheless, it seems that threats to users' privacy will be encountered. To solve this problem, Ries [11] intuitively allows the analysis of trust as a subjective prospect, which permits the consideration of context-dependent and individual preferences parameters. However, building up trust and reputation usually requires long duration categorizing that can be a link across numerous transactions.

In cloud computing, trust management is one of the most critical matters and has become a popular research area [7–9]. For example, Brandic et al. [7] presented compliance management using a centralized approach in cloud server-side environments. This method supports customers to select proper trust services in the cloud environment from their own viewpoint. Hwang and Li [8] proposed security-aware cloud architecture, which offers data coloring techniques, and trust negotiation to support the cloud service from a provider perspective. The cloud service consumers' perspective is supported using the trust-overlay networks to deploy reputation-based trust management. Ghali et al. [9] have proposed a security framework called CATRAC to compose web services. CATRAC combines both Role-Based Access Control and Trust-Based Access Control in order to arrive at an optimum solution. However, unlike previous research works [7–9, 12], the proposed CbTMS method utilizes data mining approach to classify Sybil attacks models. Data mining is a new technology and has widely been used by data scientists for research and business purposes. The overall goal of the proposed data mining approach is to extract information from a dataset and convert it into a comprehensible structure for further use. Among many data mining techniques, the decision tree is appropriate for use to extract models and find out how certain variables are associated with important data classes.

Decision trees offer multiple advantages; however, one of the most important advantages is that the knowledge can be extracted and represented in the form of classification (if-then) rules. Decision tree induction is the learning of decision trees from class labeled training tuples. Each rule indicates a unique path from the root to each leaf. In operations research, specifically in decision analysis, a decision tree (or tree diagram) is a decision support tool. A decision tree is a flowchart-like tree structure, where each internal node (nonleaf node) denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (or terminal node) holds a class label. There are several machine learning algorithms (MLAs) currently in use for distinguishing the normal and anomalous activities, including the ID3 [13] algorithm as well as Naive Bayes Filter, J48, and C4.5 [14–16] classification model. These classification models can be used in any point of the network, which provide very fast statistical detection of the application, to distinguish the normal and anomalous activities in a cloud. Fastidiousness of future classification by MLAs depends heavily on quality of the training data.

As described above, mechanisms and algorithms for MCS application, Sybil attack, and cloud computing trust models have been proposed. However, their approaches are not applicable to detect Sybil attacks in MCS environments by utilizing trust management system. Therefore, we attempt to identify Sybil attacks in MCS environment by utilizing a cloud based trust management system that differentiates between credible trust nodes' and malicious trust nodes' feedback through a credibility model.

3. Detection of the Sybil Attack in Mobile Crowd Sensing Factors

The mobile crowd entities in the system include smartphone or tablet PC, and the service provider will support interactions between them, that is, inquiries about environment information service. Therefore, such interaction will specify the service content. For example, a user using his/her smartphone, namely, entity A, is interacting with service providers regarding temperature information in his/her current location. Then, entity A here, an interaction initiator, will select a trustable service provider from a set of available service providers; he/she will start evaluating the trustworthiness of the available service providers from the selection list. Then, entity A will examine the direct evidence from previous interactions and recommendations from one or multiple service providers. Hereby, the trust model can be used for combining collected evidence or giving lower weight to recommendations from unreliable sources. The trust model derives trust values for the service providers and then becomes the basis for deciding whether to interact with one of the available services providers and which service provider to select.

Therefore, a cloud based service management framework has been proposed in this paper that consists of trust as a service (TaaS) using the service oriented architecture (SOA). In particular, the proposed cloud based service management framework applies web services to interact with distributed smartphones. This web service is one of the most significant empowering technologies for cloud computing; hence, its similarities to other resources (e.g., software, infrastructures, and platforms) in the cloud are unprotected as services. Therefore, when there is a trusted participant wishing to give his/her trust feedback or inquire about the current trust data in our SOA, he/she can utilize feedback message such as text messaging or multimedia messaging to deliver his/her own data or to get inquired trust data. Figure 1 represents the proposed framework; this framework contains three different layers: the cloud service provider layer, the trust management system layer, and the cloud service consumer layer.

The proposed architecture consists of a cloud service provider layer, which contains multiple service providers who provide cloud services. The minimum symbolic feature of every cloud service is at least providing infrastructure as a service (IaaS); that is, the cloud provider should have a data center that provides the storage, the processing power, and communication capability. Moreover, in the trust management system layer, the proposed layer consists of several distributed trust management system (TMS) nodes that expose interfaces so that cloud service consumers can give their trust

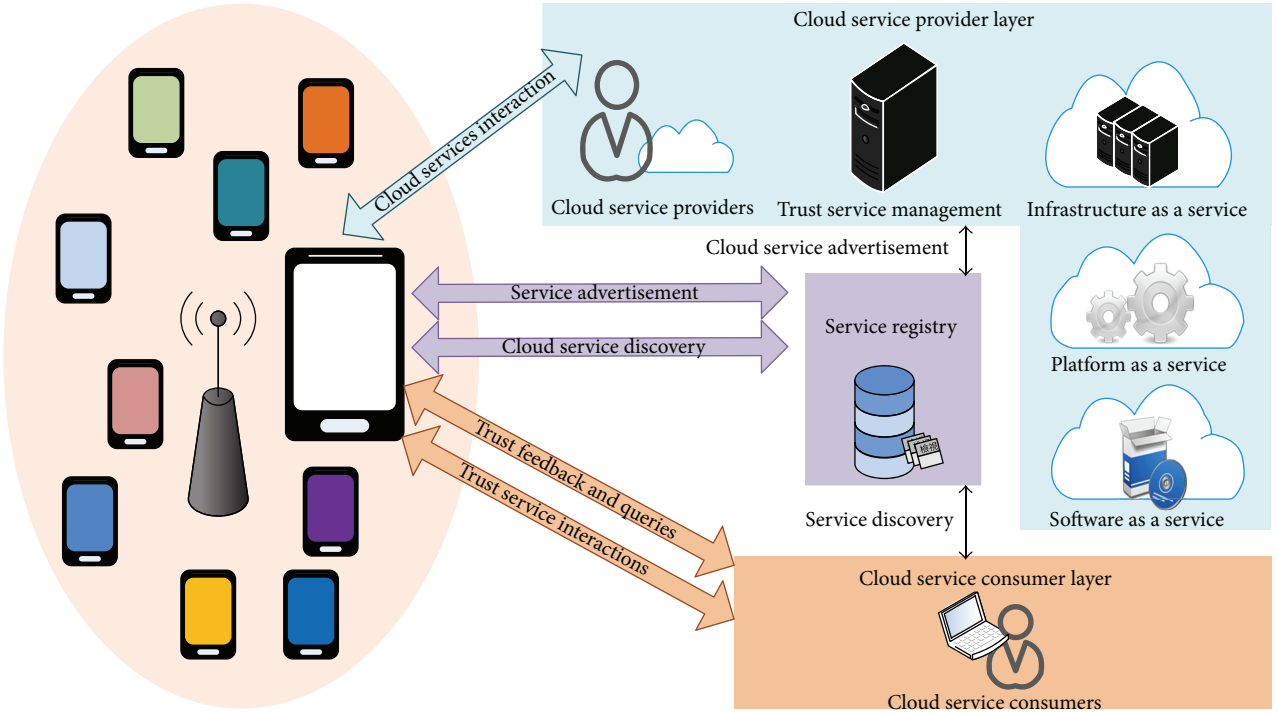


Figure 1: Architecture of the trust as a service framework.

feedback or inquire about the trust results. Furthermore, the cloud service consumer layer consists of different cloud service consumers. For example, a new startup that has limited funding can consume cloud services (e.g., hosting their services in Amazon S3). A cloud service consumer can give trust feedback of a particular cloud service by invoking the TMS.

However, MCS in the wireless environment is exposed to malicious participants deliberately contributing forged nodes and bad data. These malicious participants can also exploit these links to remove the anonymity of the volunteers and compromise their privacy. Like other networks, the security requirements in participatory sensing include services such as confidentiality, integrity, authentication, and access control to defend against malicious participants. Threats such as Sybil attack should be addressed. Therefore, identifying specific Sybil identity features in a MCS network needs to be addressed. For example, while Sybil identities compromise a MCS network, a Sybil identity will impersonate multiple identities. Hence, these Sybil identities will move in a united way because all these impersonating nodes were propagated by a single physical device. As Sybil identities move geographically, all of them will appear or disappear in the network simultaneously as the attacker moves in and out of range. This phenomenon differs from a healthy MCS network where participators are free to move at will.

Therefore, this CbTMS framework exploits Sybil attack characteristics to perform Sybil attack detection based on the following three assumptions. First, it is assumed that the MCS network traffic can be recorded in the cloud. Therefore, the normal network traffic and abnormal network traffic can be observed and analyzed. Second, it is assumed that each user and service provider who wants to participate in the system owns a unique identity, which is acquired at the bootstrapping phase from a party that is trusted by all involved parties (i.e., users, services directory provider, and service providers). Third, it is assumed that each Sybil identity uses a single-channel radio frequency; multiple Sybil identities should transmit consecutively whereas multiple independent nodes can transmit in parallel.

3.1. Passive Checking Scheme. This CbTMS framework includes a passive checking scheme (PCS) and active checking scheme (ACS) that simultaneously keep Sybil identity nodes in check, including traffic volume, signal strength, and network topology. This PCS introduces an adaptive threshold (similar to the watchdog implementation method) to identify the characteristics of Sybil attacks in MCS network. This PCS is implemented in the cloud server side and ACS is implemented in the remote client side. The PCS regularly checks the covered MCS node's conditions to decide whether the node's identity is genuine or has been compromised. The PCS will set multiple adaptive thresholds to monitor covered MCS nodes' characteristics and is implemented as part of the system operations process running on the cloud server. When a requester inquires about the trust credit of an inspector from the CbTMS framework, if the passive PCS does not detect any attack pattern on the node, it returns no attack pattern found to the requester. Otherwise, it will notify the requester to disconnect suspicious malicious node(s).

(1) Traffic Volume. Inside a base station communication range, there may be several thousand mobile devices, with

multiple applications for each device. Hence, the next step is to further classify different groups within the mobile device population with dissimilar characteristics and refine the models. Due to different devices presenting greatly different behaviors and traffic patterns, a naive extension of this model will be to develop a specialized model for every device type. The next step is to further identify groups in device population with similar characteristics and refine the models. As mentioned in our background work, once a Sybil identity has compromised a partial MCS, it will create a number of online identities and use these identities to compromise participant sensing. Therefore, by analyzing this traffic volume, signal strength, and network topology at a regular period, our CbTMS framework can infer whether the system has suspicious Sybil identities.

In our framework, the dynamic traffic of the MCS network is recorded in the cloud. It can be represented as $F = \langle F_1, F_2, \ldots, F_i, \ldots \rangle$, where F_i denotes the traffic at time i. The proposed CbTMS framework may group n entries in F into a single entry. For example, assuming $n = 2$, the new sequence for the traffic volume becomes $\langle F_1F_2, F_3F_4, F_5F_6, F_7F_8, F_9F_{10}, \ldots \rangle$. Thus, the traffic volume can be measured and analyzed with different time resolutions. Our goal is to obtain normal and abnormal traffic models from the collected sensing data. For this purpose, the k-means clustering [4], which is a well-known method for partition clustering, is applied in our framework. The k-means clustering can associate every observation with the nearest mean and hence is useful for cluster analysis, especially for a large number of variables and datasets. More specifically speaking, in this study, the k-means clustering can be used to divide the sensing data space so as to distinguish the normal and the abnormal traffic models. The intracluster heterology V has been used for measuring to select the appropriate value of k. As presented in formula (1), the value of heterology V will be calculated for increasing values of k starting from $k = 2$. Intracluster heterology is defined as

$$V = \sum_{i=1}^{k} \sum_{x_j \in S_i} \left(x_j - \mu_i \right)^2, \quad (1)$$

where x_j is a data point residing in the ith cluster, μ_i is the centroid point of the ith cluster, S_i is the collection of all the data points residing in cluster i, and k is the number of clusters. For instance, we can group the normal network traffic volume to S_c and S_r. Now, k-means clustering has been applied to analyze and divide normal and abnormal network traffic into distinct groups. In this study, we can calculate the value of V for increasing values of k. As shown in an example in Figure 2, T8's S_c and S_r ratios are obviously different from the other groups. In this situation, the PCS can analyze the network traffic volume in the cloud DB and assume that suspected Sybil identities existed in the MCS network.

(2) Signal Strength. After the suspected Sybil identities are detected using the traffic volumes as described above, the signal strength of these suspected Sybil identities is further analyzed. The signal strength is determined by considering

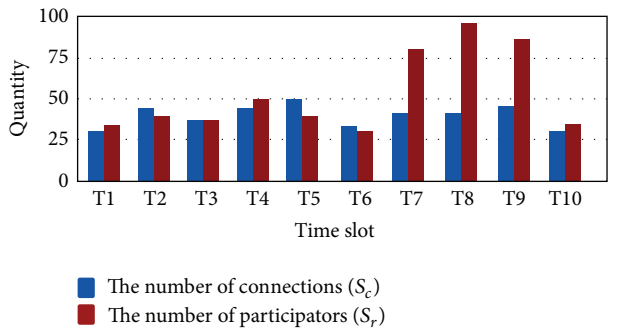


Figure 2: An example diagram of suspicious Sybil attack activity traffic volume.

the number of neighbor nodes inside a base station communication range. For example, when Sybil identities have compromised a MCS network, it will represent multiple fake identities and exchange of data among them. Fortunately, this gives our PCS an opportunity to obtain and check the signal strength of Sybil identities. However, we do not check the entire transmission signal. We only check the transmission signal from Sybil identity successfully received by its neighbor node. For example, we denoted the number S, $0 \leq S \leq 1$, as a signal-received probability that a transmission signal will be picked up by a neighbor node of a Sybil identity. Then, we denoted the number s, $0 \leq s \leq 1$, as the probability of whether this neighbor node will receive the signal. For each transmission, the transmission signal will be checked only if $s < S$.

Assume that R represents the maximum ratio difference, P_r represents received signal strength, and P_e represents expected received signal strength. Given a signal, the ratio difference r is shown in

$$r = 1 - \left(\frac{\min\left(P_r, P_e\right)}{\max\left(P_r, P_e\right)} \right). \quad (2)$$

For any signal that is received by a node, a suspicious signal can be classified if its ratio is different $r > R$. In addition, this signal strength may have precision problem because the received signal measurement result will depend on the transmitter geographical location. An example is shown in Figure 3. Figure 3(a) shows that, in an original network, there are 4 mobile nodes in the base station communication range, and Figure 3(b) shows that there are another 6 suspicious Sybil nodes when Sybil attacks occur.

(3) Network Topology. Because each Sybil group will present a similar topography map, therefore, nodes will very frequently received neighboring nodes' signals even when they are not Sybil identities and will rarely be heard apart as they normally will not move out of radio range. This phenomenon will lead to a false identification rate in topographies due to the higher density in terms of nodes per square meter. Hence, when a Sybil attacker is present in the network, the error rates for a single node spectator will be very obvious. However, in smaller topographies, as all nodes are seen as part of the same identity, there is insufficient mixing to separate

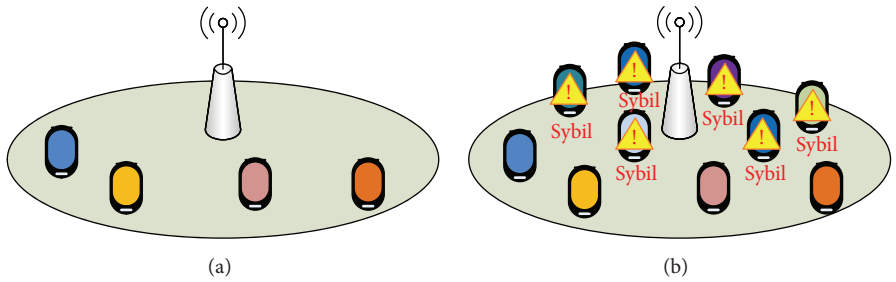


Figure 3: Illustration of suspicious Sybil attack activities in a region.

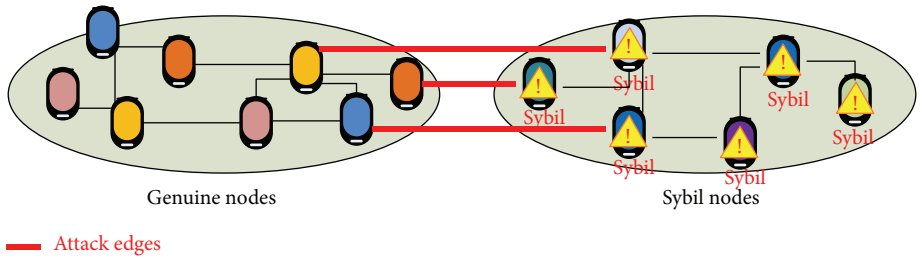


Figure 4: A conceptual network topology of Sybil attacks activities.

Sybil identities from real node, which leads to the high error rate. Once the topography size increases, the number of meaningful observations that can be made increases; and the true positive rate stays high. As the topography size increases further, the number of observations that a single node can make is reduced, as all nodes are spread far apart, and the accuracy of identifying the Sybil identities decreases. As shown in Figure 4, when Sybil attacks occur, the network topology can be conceptually divided into two parts: one consisting of all genuine identities and the other consisting of all Sybil identities. The link connecting a genuine node to a Sybil node is called an attack edge [12].

3.2. Active Checking Scheme. In this section, an active checking scheme (ACS) has been proposed to detect these Sybil nodes and eliminate them in peer-to-peer (P2P) network. The most significant feature in the P2P network is that each peer acts as both server and client. In other words, there is no central server that is used for storing the files and providing download. All nodes download files directly from other peers. Therefore, modern P2P networks suffer from the nuisance of malicious entities, such as DDoS query flooding attacks. We refer to Sybil attacks, which forge multiple identities to negatively impact or even control the entire network. A malicious Sybil identity will cheat its neighbor nodes by creating virtual nodes that are called virtual Sybil nodes. Our challenge is how to detect the Sybil identity with these virtual Sybil nodes and eliminate them to ensure the routing security while routing forwarding.

As the Sybil identity will forge fake identity and location and report its virtual location information to server nodes, it is easy for the malicious Sybil identity to forge reasonable virtual locations if the malicious node knows the location information of its neighbors. For example, we assume that a malicious Sybil identity node is O and it obtains locations of its five neighbor mobile nodes A, B, C, D, and E. Then, it can infer that the five neighbor nodes are in the concentric circles with the center O and satisfy $OA > OB > OC > OD > OE$. Nevertheless, a Sybil identity node has difficulty generating five neighboring mobile nodes' traffic volume, signal strength, and network topology from time to time. Therefore, the proposed ACS will inquire regard with its forged node identity information and hop distances from suspicious Sybil identity node. Then, ACS will pass this information to PCS scheme to do further verification. Hence, the ACS will need to cooperate with a mobile PCS that actively inquires Sybil nodes information and the PCS to verify the response information from the suspicious Sybil identity node. Once the response messages are different from PCS reservation result, then we can find out the Sybil attack area in mobile crowd sensing network and eliminate the Sybil identity in cloud server side.

3.3. Trust Credit Assessment. In the proposed framework, the trust credit of a MCS node is evaluated by Trust Credit Assessment (TCA) scheme. It is characterized by a collection of invocation history records denoted by H. Each requester node r holds a point of view regarding the trustworthiness of a mobile inspector node i in the supplication history record, which is managed by a trust management service. Each supplication history record is represented in a tuple that consists of the MCS node primary identity P, the mobile PCS identity I, a set of trust credits T, and the aggregated trust feedback weighted by the credibility T_c (i.e., $H = (P, I, T, T_c)$). Each credit in T is represented in numerical form with the

range of [0, 1], where 0, +1, and 0.5 signify negative feedback, positive feedback, and neutral feedback, respectively.

Whenever a requester node inquires the trust management service regarding the trustworthiness of a mobile PCS i, the trust result, denoted by $T_r(i)$, is calculated as

$$T_r(i) = \frac{\left(\sum_{k=1}^{l} v(k,i)\, T_c(l,i)\right)}{|v(k,i)|}, \tag{3}$$

where $v(k,i)$ is all of the feedback given to the mobile PCS i and $|v(k,i)|$ represents the length of $v(k,i)$ (i.e., the total amount of feedback given to the mobile PCS i). $T_c(l,i)$ are the trust feedback from the lth cloud consumer weighted by the credibility.

3.4. Analytical Decision Support. In this section, we applied the well-known and widely used C5.0 decision tree algorithm, which was an improved version of C4.5 [16]. In the last few decades, several decision tree learning algorithms have been proposed including Ross Quinlan who invented the Iterative Dichotomiser 3 (ID3) [13] which was used to generate a decision tree from a dataset, as well as Naive Bayes Filter [14], J48 [15], and C4.5 [16] decision tree learning models that can be applied in many applications which provide a very fast statistical method to classify data. The decision tree can use various machine learning algorithms (MLAs) for providing informative diagnosis models according to data features or rules to solve classification problems. The aim of decision tree is to identify numerous ways of splitting a dataset into branch-like segments. This branch-like segment can produce a relationship model on the basis of the data collected from different sources. One of the most noteworthy characteristics of decision trees is represented in the form of classification (if-then) rules. Each rule represents a unique path from the root to each leaf. In operations research, precisely in decision analysis, a decision diagram is a decision support tool.

Different from previous decision tree algorithms, the C5.0 classifier comprises a simple command-line interface, which can generate the decision trees and rules and finally test the classifier. In numerous applications, rule-sets are preferred because they are simpler and easier to understand than decision trees, but compared to C5.0, C4.5's rule-set methods are slow and memory-hungry. As described in [16], C5.0 [17] algorithm has been recognized as an efficient data mining technique compared with C4.5 algorithm. Because C5.0 represents a new algorithm for generating rule-sets, the improvement is substantial. C5.0 model works by splitting the sample based on the field that provides the maximum information gain [18]. The information gain is computed to estimate the gain produced by a split over an attribute.

Let S be the sample:

(i) C_i is Class I, $i = 1, 2, \ldots, m$, $I(s_1, s_2, \ldots, s_m) = -\sum p_i \log_2(p_i)$.

(ii) S_i is the number of samples in Class I. $P_i = S_i/S$; $\log_2$ is the binary logarithm.

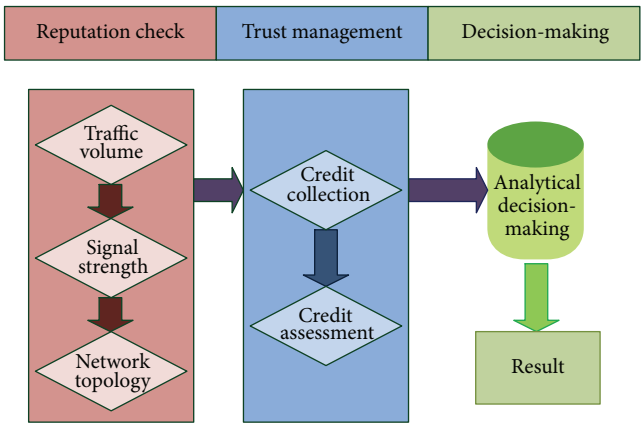


Figure 5: Hybrid reputation monitoring diagram.

(iii) Let Attribute A have v distinct values:

$$\text{Entropy} = E(A) = \sum \left\{ \frac{\left(S_{1j} + S_{2j} + \cdots + S_{mj}\right)}{S} \right\} * I\left(s_{1j}, \ldots, s_{mj}\right), \quad j = 1. \tag{4}$$

(iv) S_{ij} is samples in Class i and subset j of Attribute A:

$$I\left(S_{1j}, S_{2j}, \ldots, S_{mj}\right) = -\sum p_{ij} \log_2\left(p_{ij}\right). \tag{5}$$

(v) $\text{Gain}(A) = I(s_1, s_2, \ldots, s_m) - E(A)$.

(vi) Gain Ratio then chooses, from among the tests with at least average gain, the Gain Ratio = $P(A)$:

$$\sum_{i}^{t} \frac{S_i}{S} \log\left(\frac{S_i}{S}\right). \tag{6}$$

(vii) $\text{Gain Ratio}(A) = \text{Gain}(A)/P(A)$.

3.5. An Example of the Scenario. As this attack has no relation to the identification scheme, we do not further evaluate it. On the other hand, an attacker can utilize Sybil attacks to compromise and control a genuine node. The compromised genuine node will be considered as a Sybil node and not as a genuine node. This Sybil node will focus on creating multiple online user identities called Sybil identities and try to achieve malicious results through these identities. As shown in Figure 5, we will implement our CbTMS algorithm in three phases. In the first phase, the cloud server-side manager will record network traffic to those who participate in the system and define multiple adaptive thresholds, including traffic volume, signal strength, and network topology, to evaluate network trustworthiness.

When a Sybil identity uses a single-channel radio and has been identified as exceeding the adaptive threshold range in our PCS, the PCS module will generate a notification

to the TCA. Then, the TCA will draw these PCS history records from its database and process the credit assessment. Once the Sybil attack pattern has been preliminarily identified, it will enable the analytical decision-making (ADM) to further analyze and determine the Sybil attacks in this network. This framework will check regular network and system statistics and use an adaptive threshold to achieve network trustworthiness. To improve the completeness of the analysis by observing how a Sybil identity behaves in participatory environments, it will require cooperation with telecommunication service cloud providers. In this cloud, we can develop a subset of system calls invoked by the analyzed program in a mobile user environment and receive the result of the computation.

4. Experimental Evaluations

In this section, the proposed algorithm cloud based trust management scheme (CbTMS) has been simulated in NS2 [19]. The main focus of this paper is to simulate Sybil node mobility model, which has been well used in many different application areas. NS2 is a popular and reliable network simulator tool based on C++ and OTcl programming languages, which were developed by UC Berkeley. It provides an open source and collaborative environment to support protocol design and network traffic studies. NS2 has extended version that provides mobility model and adds more supporting features such as sending and receiving packets over wireless channel and provides radio propagation model, MAC protocols, interface queue, link layer configuration, ability to move within a given topology, and ad hoc on-demand distance vector (AODV) routing which is suitable for setting our simulation parameters. This simulation environment considers an urban road scenario with two lanes in each direction. In this scenario, vehicles are placed on the road randomly with the a minimum 5-meter intervehicle space in each lane. Vehicles travel on the road with speeds of predefined formula (miles/hour). The communication range of benign vehicles is 50 meters while the malicious vehicle may adjust its transmission power according to the situation. We inspect the detection time, which is defined as the time interval from when the malicious vehicle starts Sybil attacks to neighboring nodes and when it is identified by other vehicles.

The proposed CbTMS framework has been implemented in NS2 based on the Internet framework and utilizes AODV algorithm and interface queue type Queue/DropTail/PriQueue model for mobility of the nodes because this model can well depict a real world situation and successfully simulate work. This mobility model is based on an entity mobility model where the nodes move independently of each other. The simulation work has taken the parameters for implementation as shown in Table 1 and Figure 6.

4.1. Malicious Sybil Node and Compromised Node Selection. Based on previous sections, the described malicious Sybil identities will be exposed to like mobile malicious participants that can deliberately contribute forge nodes and bad data. In our simulation experience, Sybil identities were designed to modify packet contents and participated in route discovery and route maintenance. They will not forward packets to neighbor mobile nodes, but only to specific compromised nodes. Hence, the packet routing paths will be the same even when new formal nodes join the routing process. Moreover, when the Sybil identity has compromised its neighbor nodes, they will have the same mobility model. Furthermore, in a Sybil attack, the selection of compromised nodes based on detecting node misbehavior was done in a random manner. These compromised nodes will have a random number generator inside them so that every time they will need to see its value before overhearing the channel. If the random number was evaluated as 0, then they will turn on their compromised mode to forward the malicious message to their neighbor nodes or else they have to remain idle. This idle state will also result in a lot of power saving of the compromised nodes without affecting the fault detection.

Table 1: Simulation implementation parameter list.

Parameter	Value
Simulation environment	Wireless channel
Radio propagation model	Two-ray ground
Network interface type	WirelessPhy
MAC type	802.11
Routing protocol	AODV
Interface queue type	Queue/DropTail/PriQueue
Link layer type	LL
Antenna model	OmniAntenna
Max. packet in IFQ	50
Simulation area	1000 m∗800 m
Simulation time	150 s
Number of nodes	18

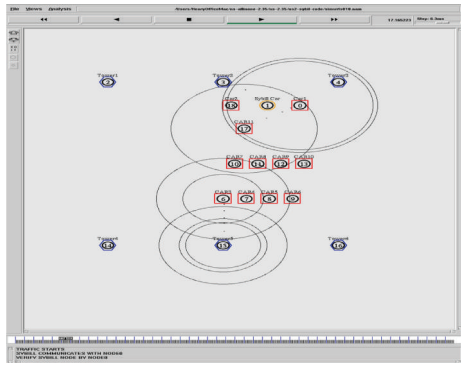


Figure 6: Simulation graphical view of nodes.

4.2. Ad Hoc On-Demand Distance Vector Routing. As described in [20], in AODV routing protocol, the routing agent will cache the packet first and broadcast a request to try to find a route. Once the packet has reached link layer, the link layer looks up ARP table to map IP address to MAC address, and then it delivers packet to interface queue. Wireless MAC will be used to avoid collision and if a

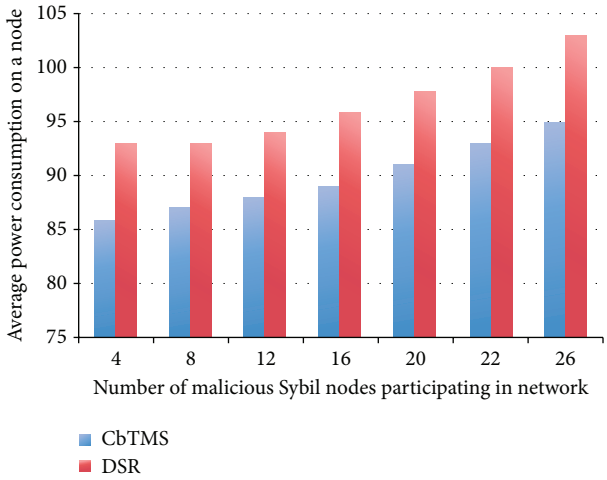


Figure 7: Average power consumption comparative diagram.

packet can be transmitted safely, it will be handed to network interface. The network interface simulates the smartphone wireless behavior in the real world. Finally, the modulated data will be transmitted over the wireless channel.

4.3. Experiment Results. Ideally, the average power consumption for a MCS node mode is 89.85 Wh as defined normal mode as shown in Figure 7. The Wh is a unit of energy equivalent to one watt of power expended for one hour of time. On the other hand, in a malicious Sybil node attack mode, the average power consumption is much higher than in a normal mode while utilizing AODV routing protocol. The simulation result shows that each node will consume 96.72 Wh on average. In the case of AODV routing protocol, it is based on the nodes having to cooperate to find a path between nodes. It allows nodes to discover and maintain routes to arbitrary destinations in the ad hoc network. Therefore, it will consume enormous power in the network in our malicious Sybil node mode. But compared with the proposed CbTMS algorithm, we will detect malicious Sybil node and compromised nodes to prevent communication overhead. In addition, in our simulation setup, there is only 1 node that has been set up as Sybil identity node. The proposed CbTMS provides lower power consumption which causes 6.87 Wh reduction compared with malicious Sybil node attack mode while utilizing AODV routing protocol.

5. Conclusion

In this paper, a cloud based trust management scheme (CbTMS) was proposed for detecting Sybil attacks in mobile crowd sensing (MCS) networks. Sybil attacks create multiple online user identities called Sybil identities and try to compromise systems with their malicious information through these identities. The proposed CbTMS framework can perform trust management and reputation checker to verify the nodes in the MCS network. It combines two schemes, namely, Characteristics Checking Scheme (PCS) and Trust Credit Assessment (TCA), to detect suspicious Sybil nodes. PCS was proposed for passively monitoring the characteristics of the suspicious Sybil nodes, including time, density, and topology in the MCS, whereas TCA was proposed for evaluating the trustworthiness of the suspicious Sybil nodes. Our simulation studies show that our CbTMS can efficiently detect the malicious Sybil nodes in the network and cause 6.87 Wh reduction compared with malicious Sybil node attack mode.

Competing Interests

The authors declare that they have no competing interests.

References

[1] H. Ma, D. Zhao, and P. Yuan, "Opportunities in mobile crowd sensing," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 29–35, 2014.

[2] T. Denning, A. Andrew, R. Chaudhri et al., "BALANCE: towards a usable pervasive wellness application with accurate activity inference," in *Proceedings of the 10th Workshop on Mobile Computing Systems and Applications (HotMobile '09)*, vol. 5, pp. 15–16, Santa Cruz, Calif, USA, February 2009.

[3] E. P. Stuntebeck, J. S. Davis II, G. D. Abowd, and M. Blount, "HealthSense: classification of health-related sensor data through user-assisted machine learning," in *Proceedings of the 9th Workshop on Mobile Computing Systems and Applications (HotMobile '08)*, pp. 1–5, February 2008.

[4] R. Douceur, "The Sybil attack," in *Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS '02)*, Cambridge, Mass, USA, March 2002.

[5] X. Liang, X. Lin, and X. S. Shen, "Enabling trustworthy service evaluation in service-oriented mobile social networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 2, pp. 310–320, 2014.

[6] K. Zhang, X. Liang, R. Lu, and X. Shen, "Sybil attacks and their defenses in the internet of things," *IEEE Internet of Things Journal*, vol. 1, no. 5, pp. 372–383, 2014.

[7] I. Brandic, S. Dustdar, T. Anstett, D. Schumm, F. Leymann, and R. Konrad, "Compliant Cloud Computing (C3): architecture and language support for user-driven compliance management in Clouds," in *Proceedings of the 3rd IEEE International Conference on Cloud Computing (CLOUD '10)*, pp. 244–251, July 2010.

[8] K. Hwang and D. Li, "Trusted cloud computing with secure resources and data coloring," *IEEE Internet Computing*, vol. 14, no. 5, pp. 14–22, 2010.

[9] C. Ghali, A. Chehab, and A. Kayssi, "CATRAC: contextaware trust- and role-based access control for composite web services," in *Proceedings of the 10th IEEE International Conference on Computer and Information Technology*, pp. 1085–1089, Bradford, UK, 2010.

[10] J. Grover, M. S. Gaur, V. Laxmi, and N. K. Prajapati, "A sybil attack detection approach using neighboring vehicles in VANET," in *Proceedings of the 4th International Conference on Security of Information and Networks (SIN '11)*, pp. 151–158, Sydney, Australia, November 2011.

[11] S. Ries, "Extending Bayesian trust models regarding context-dependence and user friendly representation," in *Proceedings of the 24th Annual ACM Symposium on Applied Computing (SAC '09)*, pp. 1294–1301, ACM Press, Honolulu, Hawaii, USA, March 2009.

[12] S.-H. Chang and T.-S. Huang, "A fuzzy knowledge based fault tolerance algorithm in wireless sensor networks," in *Proceedings of the 26th IEEE International Conference on Advanced Information Networking and Applications Workshops (WAINA '12)*, pp. 891–896, IEEE, fukuoka, Japan, March 2012.

[13] D. Jiang, *Information Theory and Coding*, Science and Technology of China University Press, 2001.

[14] S. F. Chen and Z. Q. Chen, *Artificial Intelligence in Knowledge Engineering*, Nanjing University Press, Nanjing, China, 1997.

[15] A. P. Muniyandi, R. Rajeshwari, and R. Rajaram, *Network Anomaly Detection by Cascading K-Means Clustering and C4.5 Decision Tree Algorithm*, vol. 30, Elsevier, New York, NY, USA, 2012.

[16] *Is See5/C5.0 Better Than C4.5?*, 2009, http://www.rulequest.com/see5-comparison.html.

[17] Information on See5/C5.0—RuleQuest Research Data Mining Tools, 2011, http://www.rulequest.com/see5-info.html.

[18] X. Zhu, J. Wang, S. Wu, and H. Yan, "Research and application of the improved algorithm C4.5 on decision tree," in *Proceedings of the International Conference on Test and Measurement (ICTM '09)*, pp. 184–187, IEEE, Hong Kong, December 2009.

[19] S. McCanne and S. Floyd, "Network Simulator Version 2," http://www.isi.edu/nsnam/ns.

[20] Z. Wang, Y. P. Chen, and C. Li, *Implementation of the AODV Routing Protocol in ns2 for Multi-Hop Wireless Networks*, Memorial University, 2010.

Hindawi Publishing Corporation
Mobile Information Systems
Volume 2016, Article ID 6254028, 17 pages
http://dx.doi.org/10.1155/2016/6254028

Research Article

Towards the Elimination of White Spots in Mobile WiMAX Grids through Time Efficient Cognitive Vertical Handoff Algorithm

Abid Ali Minhas,[1] Ahmed Qaddus,[2] Muhammad Saeed Khan,[3] Mohsin Iftikhar,[4] Tanveer Zia,[4] Abdulaziz Almazyad,[1,5] and Naveen Chilamkurti[6]

[1]*Al Yamamah University, Riyadh 11512, Saudi Arabia*
[2]*Wireless Research Center, Bahria University, Shangrila Road, Sector E-8, Islamabad, Pakistan*
[3]*University College of Engineering, Sciences & Technology, Lahore Leads University, Lahore, Pakistan*
[4]*Charles Sturt University, Locked Bag 588, Boorooma Street, Wagga Wagga, NSW 2678, Australia*
[5]*King Saud University, Riyadh 12372, Saudi Arabia*
[6]*La Trobe University, Plenty Road and Kingsbury Drive, Melbourne, VIC 3086, Australia*

Correspondence should be addressed to Mohsin Iftikhar; miftikhar@csu.edu.au

Received 2 October 2015; Accepted 29 December 2015

Academic Editor: Kamal Deep Singh

During the past two decades particularly, there has been incredible growth in the telecommunication industry which evolved the demand for real time applications. Mobile grids and internet service providers are offering competitive real time services to customers in order to fulfil their daily demands. Nowadays, WiMAX has become a key player in telecommunication industry for providing cost effective communication solutions to customers residing in developed and underdeveloped regions. Moreover, a significant role of WiMAX technology is to fill the gap between the first-world and third-world countries. WiMAX technology offers affordable low cost wireless multimedia services to its customers by using point to multipoint propagation methods. Over the past few years, the problem of white spots in WiMAX grids has been reported in the literature extensively. Consequently, this issue has got a lot of attention from researchers. In order to disentangle the phenomenon of white spots, a novel algorithm based on time efficient cognitive vertical handoff scheme has been presented in this paper. The proposed algorithm is based upon principles of cognitive heterogeneous wireless grids. The simulation results have shown that the proposed algorithm is a promising candidate to overcome the problem of white spots in mobile WiMAX grids.

1. Introduction

The WiMAX radio frequency technology for Wireless Metropolitan Area Network (WMAN) grid is based upon principle of IEEE 802.16-2004 standard [1]. WiMAX IEEE 802.16-2004 standard has become a good and affordable alternative against fixed lines wired grid like Public Switched Telephone Network (PSTN) and is an attractive replacement for cord and Digital Subscriber Line (DSL) utilities. WiMAX IEEE 802.16e standard is capable of providing association among static Base Transceiver Stations (BTS) along with mobile gadgets. High speed handoff is feasible for end users located in rapidly moving transport by using IEEE 802.16e standard, which entirely holds the mobility of mobile users [1–3].

Through practicing Orthogonal Frequency Division Multiple Access (OFDMA), WiMAX can accommodate multipath execution for near line of sight intercommunication. One enhanced feature of WiMAX IEEE 802.16e standard endures Scalable Orthogonal Frequency Division Multiple Access (SOFDMA), which basically restricts the interference between different end users and can accommodate high data rate applications and enhances scalability as well. Quality

of Service (QoS) and security are the two fundamental requirements in WiMAX grid. WiMAX can provide pervasive, point to point, and point to multipoint broadcasting by using flexile engineering and better encryption methods. WiMAX is very flexible in terms of requiring less time for the deployment of the grid. Moreover, WiMAX is capable of providing interoperability between different vendor's devices and portability with existing grids. The usage of higher modulation and channel bandwidth methods allow WiMAX to accommodate its users with high capacity [2–7].

2. Background of WiMAX

WiMAX forum has proposed different WiMAX standards. One of the key features of initial WiMAX standard IEEE 802.16a is that it can be executed in licensed and unlicensed frequency bands and has a frequency band spectrum range from 2 to 11 GHz. In IEEE 802.16a WiMAX standard transceiver antennas of source and destination devices do not require direct line of sight between each other as the devices operate in low frequency ranges. WiMAX forum introduced IEEE 802.16b standard for providing good quality to real time voice and data services customers. It has operating frequency band range from 5 to 6 GHz. WiMAX IEEE 802.16c standard allows interoperability between different vendor devices and gadgets and has an operating frequency band range from 10 to 66 GHz [3, 8].

For compliance testing amongst different vendor devices, WiMAX forum proposed much improved and well equipped version of IEEE 802.16a standard known as IEEE 802.16d. Lastly, IEEE 802.16e standard by WiMAX forum suggests nonline of sight communication in a wider broadcasting area. It also provides the capability to mobile users to communicate moving at vehicular speeds [3].

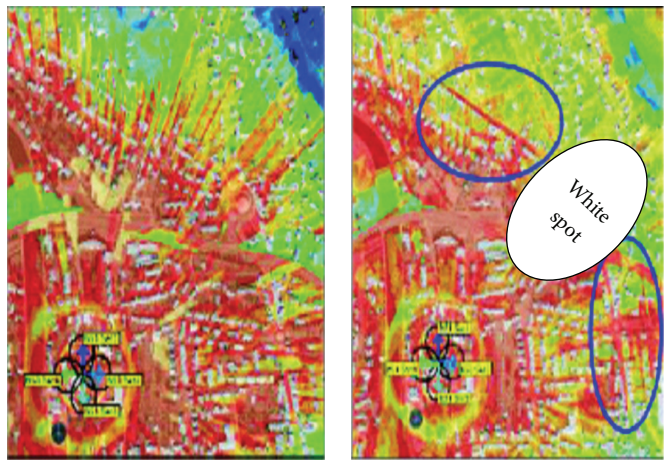
"White spots" are basically the operation situated in the WiMAX grid broadcasting that has nil broadcasting from the serving base station(s) due to undefined and unexpected events. In the literature white spots are also named as blind spots [9], black spots [10], and hot spots [11] where grid broadcasting is not available to WiMAX subscribers. WiMAX broadcasting area cannot be marked with exact certainty on the topographical depiction as white spots are not precisely located on the atlas. Two primary ways for surveying the broadcasting region of a WiMAX grid are presented in subsequent sections [3].

2.1. Empirical Model. Empirical model is the first model that supports the principle of trial based installation and is equipped with empirical measurements. Statistical tools used by empirical model are used to perform in-depth analysis of data rate of a particular broadcasting area of a WiMAX grid. The models based on easy principles of empirical modeling fail to detect white spots in the WiMAX broadcasting area. The principle of statistical probability is followed by empirical modeling which fails to provide accuracy for WiMAX base station sites. Hence empirical modeling fails to provide the appropriate link budget plan for the commissioning of WiMAX base stations [3, 12].

2.2. Site Specific Model. The second type of model is known as site specific model and is based on deterministic methods. The broadcasting area of WiMAX grid is analyzed by software tools. This model is able to detect the exact position of the target on the atlas by using tools like Global Positioning System (GPS). Further, site specific model does not follow the theory of probability and uses software. But, the hurdles located in topographical region cannot be traced by this site specific model, which is the potential source of communication obstruction. The interruption is due to the influence of multipath and fading signals received at the destination which is caused by the reflection of signals from the obstacles present in their path. If the transmitted direct signal and the obstructed signal reach the destination at the same time, then both signals will cancel each other due to 180-degree phase shift. Ultimately, due to cancellation of transmitted signal and obstructed signal, the destination will receive no signal. Software maps based on GPS are used to gather information about dry and wet areas of topographical regions. The ray tracing or ray launching algorithms used in deterministic software model are supported by homogeneous geometrical theory of diffraction. The broadcasting spectrum of WiMAX differs from other wireless technologies with respect to its equipped features. In near line of sight intercommunication, the performance and result accuracy of site specific or deterministic software model are much better than the statistical empirical model which is based on calculations derived from Maxwell's theory. Hence under this contemporary condition, the site specific or deterministic software model has the upper hand in terms of more accurate results as compared to empirical or statistical model [3, 12].

Figure 1 has two parts, where part (a) shows a general broadcasting model [12] and part (b) shows a specialized WiMAX broadcasting model [12]. White spots are not present in the simple broadcasting area which is also referred to as a general broadcasting model. White spots are present in WiMAX grid broadcasting area. It was practically proven by telecom hardware equipment vendors through commissioning and measuring the Received Signal Strength Indicator (RSSI) of single base station in WiMAX grid. This illustrates that the WiMAX technology can be used during line of sight (LoS) communication in Metropolitan Area Network (MAN) grids deployed in cities but it is not feasible for near line of sight (NLoS) communication. During testing some noncoverage network area (NLoS WiMAX grid broadcasting area commissioned in thickly inhabited urban surroundings), some white spots were reported due to obstructions and obstacles present in the path between the transmitting and receiving station as shown in Figure 1(b) [3, 12].

During the test an omnidirectional antenna equipped with spectrum analyzer was utilized at destination site. Coverage up to 30 km can be obtained in LoS scenario for reliable communication. Coverage up to 5 km can be obtained in NLoS scenario. In the event of testing NLoS communication mode, not all of the receivers present in the WiMAX base station coverage area were receiving transmissions. White spots in the WiMAX broadcasting area can be identified from these noncoverage areas or empty spots, where base station RSSI has not been received at the WiMAX receivers.



(a) (b)

Figure 1: WiMAX Coverage Model [12].

After recording these practical results it was noted that the RSSI values being calculated on the basis of theoretical deterministic software models differ from the actual obtained results. Hence the outcome of this trial scenario proved that both empirical and deterministic theoretical models actually failed to calculate accurate RSSI values for link budget design used for commissioning WiMAX sites. Different obstructions and hurdles like skyscrapers, advertisement boards and vegetated areas, and so forth are the major sources of reflection which result in weaker or no signal availability at some spaces in the WiMAX broadcasting region. The obstructions between the base station and the destination result in poor broadcasting due to lack of signal strength. Even though the user is using enhanced modulation schemes for data transmission still a minor movement between the user and the serving base station can result in signal degradation. Due to the hurdles like white spots present in the WiMAX broadcasting spectrum, it is quite complex to design a network based on measurements obtained from software or deterministic models. The crosses in Figure 2 represent white spots in the broadcasting area of the WiMAX spectrum [3, 12].

In order to have uninterrupted communication between the base station and the subscriber station and to avoid white spots in WiMAX grid, the WiMAX operator has to increase capital expenditure for network deployment. White spot can be a potential problem and can even lead to call drop in the event of handoff procedure when the WiMAX subscriber is roaming and is shifting from one base station to another to have an established and uninterrupted connectivity with WiMAX base station. The fast movement of WiMAX mobile user at vehicular speed is supported by the WiMAX 802.16e standard. When the WiMAX mobile user is moving at vehicular speed, there is a potential risk of disconnectivity from WiMAX base station. It is very challenging to reveal the white spots or no signal areas present in the WiMAX broadcasting spectrum. Also, in order to prevent the white spots, it is quite

Figure 2: WiMAX coverage [12].

complex to calculate the exact number of required base stations to cover a specific WiMAX broadcasting area through using some theoretical network planning tools [3, 12].

The propagation of WiMAX broadcasting range exists differently for indoor, urban, and rural environments. The broadcasting spectrum encounters different types of obstacles which are the main cause of the signal loss or fading in WiMAX broadcasting area. The propagation models being used have limited information regarding the objects present on the map. For example, in case of an urban environment, the propagation model only displays the 3D image of the buildings. The height of the building is set by default, which could differ from the original height of the building. This propagation model does not specify the obstacles or obstructions like trees, billboards, and towers, which can affect the WiMAX broadcasting area. For designing an indoor environment, it is essential to plan the number and specific

size of windows and doors present in the building. Further, it is very difficult to combine the two databases of urban and indoor environment because of the limited memory of standard PC systems [3, 13].

In an indoor environment, the main causes of signal loss and diffraction are the horizontal walls and the vertical roofs. The empirical models based on statistical formulas are also used for WiMAX grid planning. This mainly depends upon the direct LoS between the transceiver equipment pieces. In metropolitan environment, the direct beam consists of high rooftop scattering whereas in the indoor environment the direct ray includes direct line between the transmitter and the receiver. In both cases, direct ray cannot be achieved between the transmitter and the receiver due to the obstructions, obstacles, and attenuation. Another WiMAX grid planning model is ray optical model. This requires high computation and processing time. In most of the cases, 98% of energy is acquired by two to three rays only for communication between transmitter and receiver. The rest of the rays are lost due to obstacles, obstructions, and attenuation between the transmitter and receiver. Hence the nonsignal broadcasting area (i.e., also referred to as white spots in the WiMAX grid broadcasting area) cannot be detected by ray optical model [3, 13, 14].

3. Related Work

Here, we describe the related work on white spots in WiMAX grids in detail. We have explained earlier the limitations of two popular types of WiMAX propagation (i.e., empirical and site specific) models. The empirical models which basically rely on statistical tools are used to analyze the broadcasting area for WiMAX grid. But, unfortunately, the white spots present in the WiMAX grid broadcasting area cannot be detected by empirical models. For tracing locations on the global map, site specific models, which mainly depend upon specialized propagation tools like Global Positioning System (GPS), are used. But the site specific model can only find or trace the location on the map. It is unable to provide the evidence regarding obstacles or obstructions present on the earth surface, which results in white spots and further stimulates the call drop or communication disruption for an active call [3, 12].

The problem of white spots in WiMAX grid has been addressed by few researchers. One proposed solution to avoid white spots has been presented by Gierlowski et al. [12]. They have proposed that the problem of white spots can be avoided by using a homogeneous approach based on mesh topology and commissioning large number of base stations. But by doing so, the capital expenditure and operational expenses will increase hysterically, consequently making this solution nonpractical.

The blind spots also referred to as white spots can be removed by multipath cellular architecture as discussed by Al Yusuf and Murshed in [9]. In the suggested idea, a mobile station while communicating with the wireless network base station via a single base station should also be connected to two other base stations. Hence, in the event of appearing blind spot, the mobile station can initiate handoff to the remaining base station present in its range and avoid disconnectivity from wireless network. Under these circumstances, a large number of base stations should be commissioned in a multipath cellular architecture, which is not very suitable to be used in WiMAX grid.

Zaggoulos et al. [15] have proposed that modulation algorithms which basically rely on adaptive channel modulation schemes can be used to eliminate the problem of white spots. In their analysis, the authors have compared the simulation and experimental results obtained for downlink Packet Error Rate (PER). It has been reported that results of soft and hard implementation of the proposed idea significantly differ from each other. According to authors, it is not possible to eliminate the white spots present in wireless broadcasting areas even with the use of well enhanced modulation algorithms in real time scenario. Chadwani et al. have argued about the use of relay stations [10] in wireless grids to eliminate the problem of black/white spots present in the broadcasting area of wireless grids for performance enhancement. For the transmission of data from source to destination, the authors have proposed a solution in the form of relay stations for relaying information from source to destination. The use of fixed relay stations is necessary for increasing cell capacity and improving the base station coverage region. Through the usage of fixed relay station, the battery life of the subscriber mobile station and power required by the mobile user during transmission can be saved. Relay stations cannot be used to eliminate the white spots in wireless network broadcasting area because they can only function as a booster or amplifier. It has been noted that relay station can only be used to increase the capability of the system but fails to increase the broadcasting area of the wireless grid.

The utilization of relay stations and repeaters has been discussed by Davydov et al. [16]. Again through using relay stations, the broadcasting area of the wireless grid cannot be increased; it can only be used to increase the capability of the system in real time scenario. The problem of noise and interference has been observed when a large number of repeaters are being commissioned in the broadcasting area of a single base station. Consequently, this results in decreasing the broadcasting area of serving base station.

The commissioning of distributed wireless sensor nodes with fixed battery source to eliminate white spots has been proposed by Jaichandran et al. [11]. By deploying wireless sensor grid nodes, we can reduce the white spots and can extend the life of the grid but for the time being only. Consequently, the commissioning of wireless sensor grid nodes is not a feasible and long term solution as it still fails to trace the white spots in the wireless grid broadcasting region. Another important point in general to be noted is that wireless sensor nodes are battery operated and are having very small data rate due to which the proposed solution is not very feasible.

Kan et al. in [17] have presented a solution according to which the deployment of dual channel scheme could be feasible to avoid white spots in WiMAX grid. The proposed idea is as follows. Due to the obstructions in grid broadcasting area, when blind/white spots appear, GPS technology will be used to provide location estimation services to the mobile

users through utilizing dual channel positioning systems. It can help to avoid blind spots in grid broadcasting area. But the main drawback of GPS technology is that it is not available to mobile users when they are present in an underground road, underpass, railway tunnel, or the basement of a building. Hence again, the deployment of dual channel scheme, which is basically founded in GSM grids and GPS technology, cannot eliminate blind spots in a wireless broadcasting area. Further, since GPS is based on satellite technology, which is often affected by weather storms and can cause complete telecommunication blackout on earth, the idea of using dual channel positioning system for avoiding blind spots in GSM grid broadcasting area is also not technically very efficient.

Lu and Wu have proposed the idea of using two-hop relay stations [18] in cellular grids to improve the performance and to remove the dilemma of dead spots or white spots present in the broadcasting area of cellular grids. The usage of two-hop relay stations might be useful for increasing capacity of cell and for the enhancement of broadcasting area of cell. Two-hop relay stations can be used to eliminate the white spots located in cellular grid broadcasting region but this will increase the capital cost of the communication system. Another point is that relay station will be acting as an amplifier or booster. Two-hop relay stations may increase the capability of cellular grid but fail to increase broadcasting region of cellular grid. Thus the method proposed here is not cost effective [18, 19].

The problem of blind spots in wireless sensor grids has also been addressed by Jin et al. [20]. Their proposed solution is based on Voronoi graph which basically relies on statistical formulas. Voronoi graph is used to locate blind spots, which are calculated by different wireless sensor nodes present in the wireless grid broadcasting area. The Voronoi graph analyzes the data of the wireless sensor grid broadcasting area by using statistical tools. The statistical techniques like Voronoi graph when used in real time scenario actually fail to detect blind spots in the wireless sensor grid broadcasting area due to their probabilistic approach.

4. Problem Statement

In WiMAX grid broadcasting region, call drop occurs due to obstacles, obstructions, and hurdles present in the path of WiMAX mobile user and WiMAX base station. These regions of nonsignal coverage areas are also referred to as white spots which is the major reason behind weak broadcasting of WiMAX base station. The primary motive for conducting this research is to propose a practical, efficient, and cost effective solution to avoid nonsignal coverage regions, that is, white spots in WiMAX grid broadcasting region. The proposed solution will be able to provide uninterruptible connectivity from mobile subscriber to the base station.

5. Proposed Model

White spots present in WiMAX broadcasting region cannot be eliminated through the utilization of WiMAX mesh architecture, so to counter this dilemma, an innovative scheme based on cognitive heterogeneous wireless grid is proposed. Cognitive heterogeneous approach has been adopted in this work to utilize the existing GSM services. To continuously access the real time applications such as voice, video, and data in a mobility scenario, the WiMAX mobile subscriber holding a cognitive mobile device will be exploiting a heterogeneous scheme. Through utilizing the heterogeneous approach, vertical handoff and switching towards GSM grid will be automatically triggered by cognitive mobile subscriber without any intervention or delay in the scenario of white spots, where the WiMAX grid is not available to accommodate the continuity of the call. Hence the dilemma of disconnection or call cutoff can be avoided by using heterogeneous approach. The signal loss will occur differentially with respect to time when the user will enter in the domain of white spots. Hence, we can say that when the WiMAX subscriber is entering in the nonsignal coverage regions, that is, white spots, there will be no sudden transform in the subscriber's Received Signal Strength (RSS). The Mobile Unit (MU) continuously stores the information about RSS collected at Forward Channel (FC) in its buffer during the event of decrease in RSS with respect to time. For certain optimized interval of time it will find the first derivative $d(\mathrm{RSS})/dt$. The decision for horizontal or vertical handoff is taken on the basis of calculated result of second derivate $d^2(\mathrm{RSS})/dt^2$. In the proposed design, priority is given to horizontal handoff. When the problem of white spot occurs, the WiMAX MU hunt for alternative WiMAX base station present in the nearby vicinity is initiated. If WiMAX MU succeeds in its hunt, then horizontal handoff procedure is executed and MU is handed over to WiMAX BS present in the nearby vicinity. If RSS is approaching critical value, then MU switches to GSM grid and vertical handoff is initiated. The following subsection contains the detail about it.

5.1. Grid Selection Criteria. The grid selection is performed on the basis of RSSI levels and required Quality of Service (QoS). When the RSSI levels at the WiMAX mobile subscriber drop below the desired threshold limit for maintaining connectivity to base station, then the selection of handoff from one type of grid to another type of grid is initiated. In this study the authors have tried to utilize the existing GSM services [9, 12] instead of commissioning the new base station. Hence the projected strategy is more effectual in terms of cost than its predecessors. Figure 3 shows the incorporation between GSM/WiMAX grid and 802.21 information server backhaul connectivity [21].

The main improvement of projected approach is that it does not rely on the principles of point to multipoint technique as in case of homogenous method [22] where all operations of the grid activity are managed and apprehended by a single base station. We can utilize homogenous method efficiently to eliminate white spots, when there are a large number of base stations commissioned in a specific broadcasting area. The capital expenditure of the WiMAX mesh grid will significantly multiply when a large number of base stations are commissioned in a specific broadcasting area. Thus WiMAX mesh architecture is not viable to be commissioned in a metropolitan area grid [12, 21].

The technologies such as WiMAX and GSM being utilized in the proposed method support the mobile subscriber

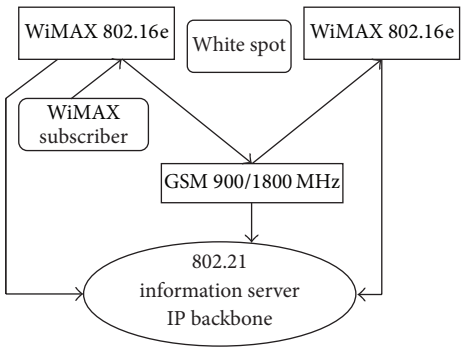


Figure 3: Cognitive heterogeneous approach of proposed model.

when present in a vehicle moving at 140 km/hr. During the scenario of horizontal handoff when a neighboring WiMAX base station is not present in the nearby vicinity of WiMAX grid, then vertical handoff is initiated to another wireless grid present in the area, which avoids the odds of call drop [21].

To disentangle the dilemma of white spots in WiMAX grid broadcasting area, a projected model based on cognitive heterogeneous wireless grid approach is depicted in Figure 3. The vertical handoff scenario of mobile subscriber from WiMAX base station to GSM base station and vice versa during the development of white spot has been elaborated with the help of flowchart in Figure 4 [3].

Figure 5 shows the vertical handoff scheme from WiMAX to GSM grids and vice versa along with customized edition of IEEE 802.21 protocols which is most suitable to our research approach. The handoff procedure of mobile WiMAX when used in heterogeneous grid is depicted in Figure 5. Initially, the WiMAX mobile user is connected to the WiMAX grid and is accessing voice and data services from the WiMAX grid. The RSSI value starts to fall below the desire threshold value required for maintaining connectivity between the WiMAX mobile station and WiMAX base station as the white spot crops up. Now handoff is required for avoiding communication disconnectivity, so the 802.21 information server will search for the availability of the neighboring Public Land Mobile Grid (PMLN) by sending MIH_GET_Information.request to 802.21 information server. Now when the PLMN is located (which in this case is GSM grid), the 802.21 information server will send MIH_Link_Detected message to mobile station. On receiving MIH_Link_Detected message, the mobile station will now send MIH_Candidate_Query.request to 802.21 information server for handoff to GSM grid base station. Now 802.21 information server will send MIH_Candidate_Query.confirm to mobile station which confirms that resources are available for handoff to GSM base station. Now mobile station will initiate handoff to GSM base station by sending MIH_Handover_Commit command. When handoff to GSM base station is completed, GSM base station will be acknowledged by sending MIH_Link_Handover_Complete command to mobile station [21, 23].

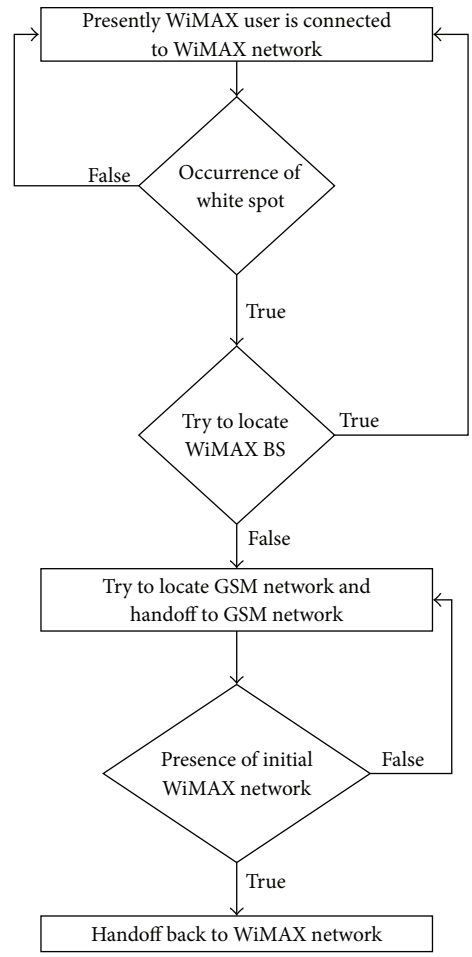


Figure 4: Flowchart of network selection.

During this alternative connectivity between the WiMAX mobile station and GSM grid, the mobile station also searches for the availability of the previous initial WiMAX grid by listening to MIH_Link_PDU_Transmit_Status, which is repeatedly broadcast by 802.21 information server. If previous (initial) WiMAX grid becomes available, then the WiMAX mobile station will first check the RSS value by MIH_Link_Detected. If the RSS value is above the minimum threshold level required for maintaining connectivity between the mobile station and WiMAX base station, then the mobile station will follow the same previous step of sending MIH_Candidate_Query.request to IEEE 802.21 information server [7, 21, 23]. Now 802.21 information server will send MIH_Candidate_Query.confirm to mobile station which confirms that resources are available for handoff back to initial WiMAX base station. Now mobile station will initiate handoff to WiMAX base station by sending MIH_Handover_Commit command. When handoff to WiMAX base station is completed, then WiMAX base station will be acknowledged by sending MIH_Link_Handover_Complete command to mobile station [21, 23].

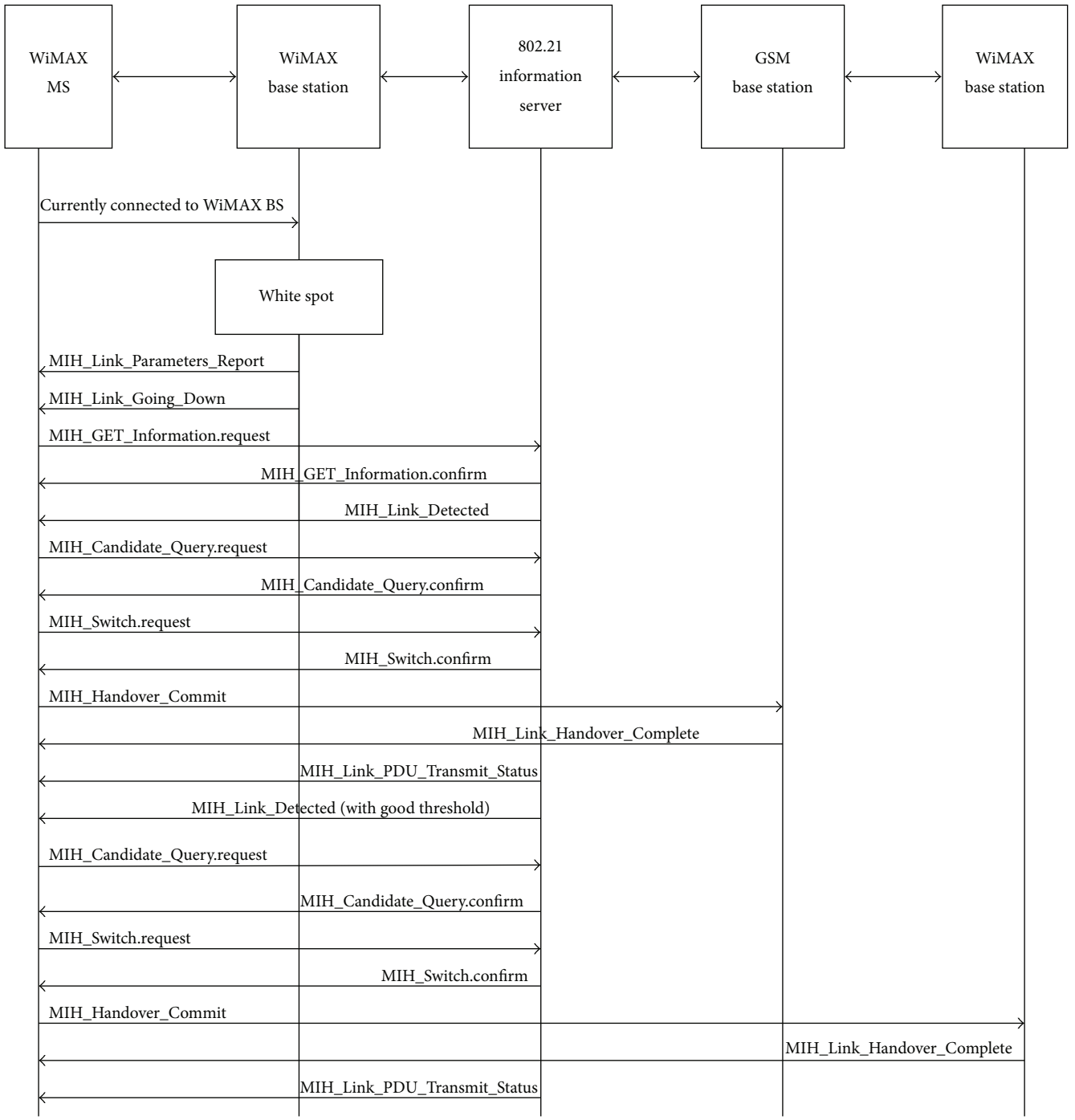


Figure 5: Cognitive mobile handoff procedure.

5.2. Handoff between WiMAX and GSM Grids. Whenever a handoff is triggered from WiMAX to GSM grid, the GSM handover message is encapsulated in the container which is the element of WiMAX grid. The GSM handover message encapsulated in the WiMAX grid container looks exactly the same as it has been sent to GSM radio interface. On the other hand the mobile station consists of two separate and specific WiMAX and GSM parts. The GSM handover message encapsulated in WiMAX container will be received at the WiMAX specific part of the mobile station. Then this GSM handover message will be extracted here and passed to the GSM specific part of the mobile station as shown in Figure 6. As we know that WiMAX and GSM are different technologies, it is difficult to compare the results between these technologies. Hence, to overcome this problem, we calculate the results on the basis of threshold. Same procedure

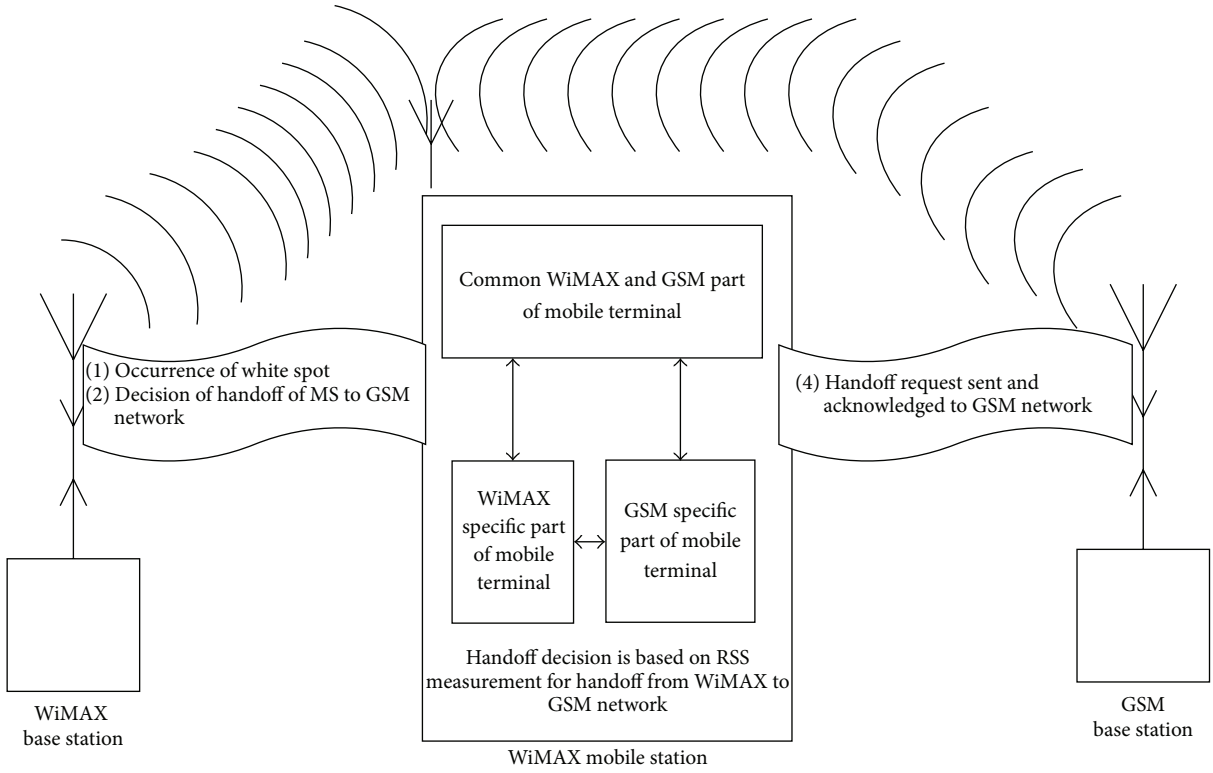


Figure 6: WiMAX to GSM handover.

of handoff is repeated when handoff is performed from GSM to WiMAX grid [24, 25].

In case of handoff from GSM to WiMAX grid, WiMAX handover message is encapsulated in the container which is the element of GSM grid. The WiMAX handover message encapsulated in GSM container will be received at the GSM specific part of the mobile station. Then this WiMAX handover message will be extracted here and passed to the WiMAX specific part of the mobile station as shown in Figure 7. The common part of WiMAX and GSM mobile terminal is a user interface LCD, which shows Received Signal Strength parameters of WiMAX and GSM grids [24, 25].

The cognitive mobile user is handed over to GSM base station present in the GSM grid, when white spot crops up and RSSI level declines below a certain threshold, that is, the minimum level required for retaining connectivity to WiMAX base station present in the WiMAX grid. The cognitive mobile user remains connected to base station of GSM grid unless the WiMAX grid base station is once more accessible for connectivity and is offering again good RSSI level for communication to WiMAX grid. In the scenario where there is no WiMAX base station present in nearby vicinity, the cognitive mobile user present in the GSM grid can even initiate horizontal handoff between GSM base stations unless WiMAX base station with good RSSI level is available within reach of the cognitive mobile user. When the desired appropriate threshold value in terms of RSSI level is sent to cognitive mobile user, then vertical handoff can be initiated back to WiMAX grid. WiMAX cognitive mobile user can initiate horizontal handoff between different WiMAX base stations present in the WiMAX grid on the foundation of good threshold value and RSSI levels [26].

5.3. Distance Formula. During the vertical and horizontal handoff, the MS moves along x- and y-axis and changes its position. The distance formula is utilized to compute the gap between starting and ending point. If the specified points are $(x1, y1)$ and $(x2, y2)$, the distance between these points is specified by the distance formula as below [27]

$$d = \sqrt{\left((x2 - x1)^2 + \left(y2 - y1\right)^2\right)}. \quad (1)$$

5.4. Received Signal Strength Formula. The Received Signal Strength is an essential parameter which helps to decide when to initiate handoff in order to avoid call drop and maintain connectivity [28]. It can be calculated as

$$\mathrm{RSS} = 10\log(e) \times \log\left(\frac{1}{d^2}\right) = 4.343 \times \log\left(\frac{1}{d^2}\right). \quad (2)$$

5.5. Radio Access Selection Formula. The Radio Access Selection (RAS) parameter decides to initiate handoff on the basis of throughput based on its Signal to Interference Noise Ratio (SINR). According to Shannon's formula, RAS is given as [29]

$$\mathrm{RAS} = B \cdot \log 2\left(1 + \mathrm{SINR}\left(j\right)\right), \quad (3)$$

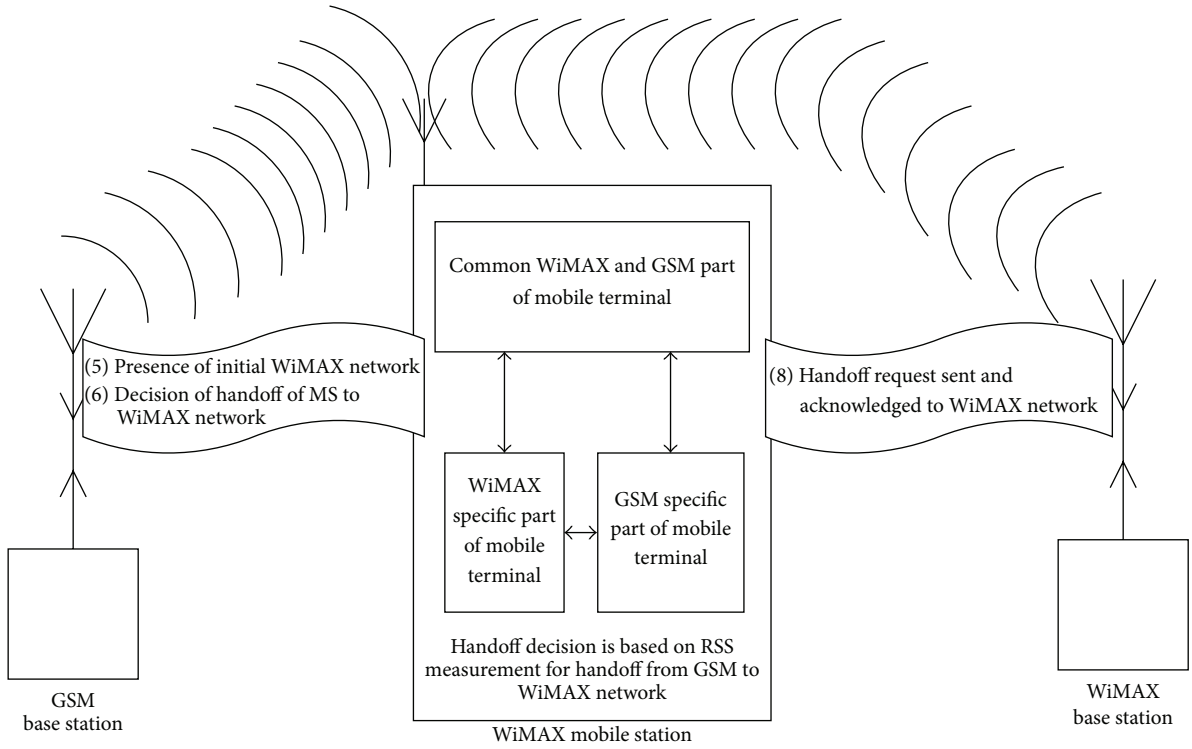


Figure 7: GSM to WiMAX handover.

where B is the system bandwidth per radio access and SINR(j) is the signal to interference plus noise ratio practiced by mobile user in a particular radio access domain [29].

5.6. Channel Capacity Formula. We can also calculate WiMAX and GSM channel capacity, which tells us about the current condition of the channel and how much data rate can be supported by the channel [30]. It can be calculated as

$$C = B \cdot \log 2 \left(1 + \text{SNR}\right), \tag{4}$$

where B is channel bandwidth which is different for WiMAX and GSM channel and SNR is the ratio of communication signal to the Gaussian noise interference [30, 31].

5.7. Path Loss Formula. γ known as path loss exponent relies on the transmission surroundings in which the signal is propagated. Propagation environment consists of the objects or obstacles present in the transmitter and receiver path [31–33]. Path loss exponent can be calculated as

$$\gamma = a - b \cdot h\left(\text{BTS}\right) + \frac{c}{h\left(\text{BTS}\right)}, \tag{5}$$

where a, b, and c are terrain coefficients which depend on different terrain types A, B, and C. h(BTS) represents the base station antenna height [32]. Similarly path loss for terrain type C can be calculated as

$$\text{PL} = A + 10\gamma \log 10\left(\frac{d}{d0}\right) + S, \tag{6}$$

where A is terrain coefficient for terrain C, γ is known as path loss exponent, d is the gap between transmitter and receiver equipment, and S is the shadowing effect [31, 34–36].

5.8. Free-Space Path Loss Formula. According to Free Space Path Loss (FSPL), the more far a receiving antenna from the transmitting antenna is, the less signal strength it will receive [31, 37]. FSPL can be calculated as

$$\text{FSPL} = 10 \log\left(\frac{\text{Pt}}{\text{Pr}}\right), \tag{7}$$

where the transmitted power is Pt and received power is Pr [37].

5.9. Two-Ray Path Loss Model Formula. By using two-ray path loss model, the MS can easily estimate its distance from the neighboring base station to serving base station by measuring RSSI value. Two-ray path loss model can be calculated as [38]

$$\text{Pr} = \frac{(\text{Pt} \cdot \text{Gt} \cdot \text{Gr} \cdot \text{ht} \cdot \text{hr})}{d^4}, \tag{8}$$

where Pr represents the received power which is to be measured from Pt (BS transmitting power). Gt is the antenna gain of transmitter, Gr is the antenna gain of receiver, ht is the transmitting antenna height, hr is the receiving antenna height, and d is gap between the MS and neighboring base station [38, 39].

Table 1: Range of input parameters of Hata model [38].

Carrier frequency	1.5 GHz to 2 GHz
Base antenna height	30 m to 300 m
Mobile antenna height	1 m to 10 m
Distance d	1 km to 20 km

Table 2: Range of input parameter for WiMAX cell radius [40].

WiMAX cell radius	Full featured
LOS	30 to 50 km
NLOS	4 to 9 km
Indoor CPE	1 to 2 km

5.10. Channel Modelling and WiMAX Cell Radius. In channel modelling, it is quite difficult to model the actual real time environment. Most of the simulation models are developed on the basis of empirical models. Empirical models use statistical tools to measure the broadcasting area of grid. By using empirical models, we cannot efficiently perform network planning of grid broadcasting areas, because empirical models are based on statistical probability. COST 231 extension to Hata model is the most suitable empirical model which can be utilized for handoff algorithm. According to IEEE 802.16e, Hata model operating frequency band is 2 to 6 GHz within a cell radius of 1 to 3 miles and has mobility up to 120 km/h. For macrocell network planning environment, Hata path loss model is widely used for coverage calculation, where the base station antenna height must be above the rooftop level of the buildings adjacent to the base station [40, 41].

According to European Cooperative for Scientific and Technical (COST) research, the Hata model for 2 GHz is given as [40]

$$\begin{aligned} &\text{PLurban}\,(d)\ \text{dB} \\ &\quad = 46.3 + 33.9\log 10\,(\text{fc}) - 13.82\log 10\,(\text{ht}) - a\,(\text{hr}) \\ &\qquad + \left(44.9 - 6.55\log 10\,(\text{ht})\right)\log 10\,(d) + \text{CM}, \end{aligned} \quad (9)$$

where CM is 0 dB for medium size cities and suburban areas and 3 dB for metropolitan areas [40, 42]. The range of input parameters of Hata model is shown in Table 1. There are three categories of WiMAX cell such as interior self-installed Customer Premises Equipment (CPE), near line of sight (NLOS), and line of sight (LOS). The input parameters range of WiMAX cells with full featured configuration is shown in Table 2 [40].

6. Simulation Results

For implementation of our proposed model, we have used Microsoft Visual Studio, which provides a complete set of development tools for building XML web services, desktop applications, mobile applications, and ASP.NET Web applications [43]. The simulated topology for wireless grids based on cognitive heterogeneous method is equipped with 802.21 information server which has backhaul IP connectivity with two GSM and three WiMAX base stations. In the event of occurrence of white spots when RSSI value falls below the required threshold level for maintaining connectivity to WiMAX base station, the cognitive mobile user first initiates a vertical handoff to the GSM base station in order to avoid call drop. This vertical handoff of cognitive mobile user from WiMAX network base station to GSM network base station is imminent since there is no other WiMAX base station present in the nearby vicinity of cognitive mobile user. Once entered in the GSM network, the cognitive mobile user has to perform horizontal handoff from one GSM base station to another while remaining inside the GSM grid. When WiMAX grid base station is once again present in the vicinity of cognitive mobile user with the required RSSI threshold level for maintaining connectivity for an ongoing call, then vertical handoff will be executed by cognitive mobile user backward to the preliminary WiMAX grid. Once a cognitive mobile user is back in the preliminary WiMAX grid, it can now execute a horizontal handoff on the basis of required RSSI threshold level within the WiMAX grid from one WiMAX base station to another to maintain connectivity for an ongoing call. If a WiMAX mobile station user is currently connected to a WiMAX base station and the RSS value starts to fall below the threshold limit, then WiMAX handover process will trigger. The WiMAX mobile station will scan the neighboring WiMAX base stations and will decide to stay or to switch to another cell on the basis of RSS threshold limit. We have implemented the heterogeneous model for WiMAX 802.16e mobile user. The topology comprises three WiMAX base stations and two GSM base stations having backbone IP connectivity with 802.21 information server as shown in Figure 8. IEEE 802.21 standard provides media independent handover services and interoperability between different wireless and fixed grids. The simulation parameters being used for vertical and horizontal handoff have been shown in Figure 9. The input parameters based on WiMAX cell channel modeling being used for simulation are shown in Figure 10. For medium size cities and suburbs, the CM is 0 dB and, for metropolitan areas, it is 3 dB [40]. The Hata model is restricted to use the range of parameters being shown in Figure 10 [44]. WiMAX cell radius is categorized into three respective categories: line of sight (LoS), near line of sight (NLoS), and indoor self-installed Customer Premises Equipment (CPE) [45].

6.1. Mobile Assisted IEEE 802.21 Handoff Protocols. IEEE 802.21 is a media independent handover algorithm. IEEE 802.21 has both terminal initiated and grid initiated handover. In terminal initiated handover, the cognitive user terminal exchanges different protocols with IEEE 802.21 information server via the serving base station, whenever the problem of white spots occurs because handoff to neighboring grid base station becomes necessary. During the handoff process, the cognitive user terminal and IEEE 802.21 information server exchange different kinds of link and MIH events as shown in Figure 11.

6.2. Grid Assisted IEEE 802.21 Handoff Protocols. The grid assisted handoffs are initiated by grid and grid mainly controls the handoff procedure. Here the grid uses MIH events,

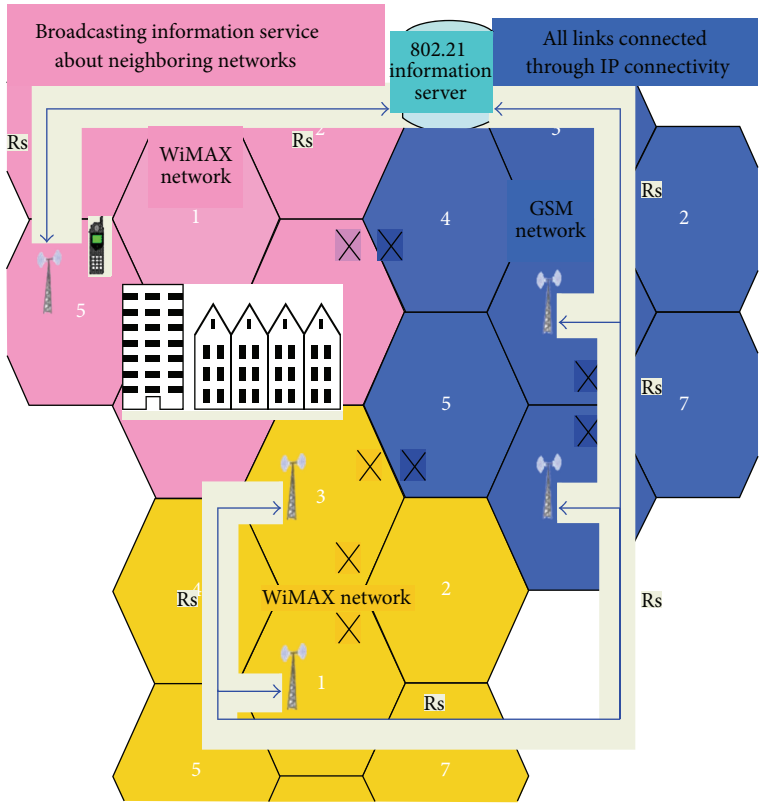


Figure 8: Simulation topology.

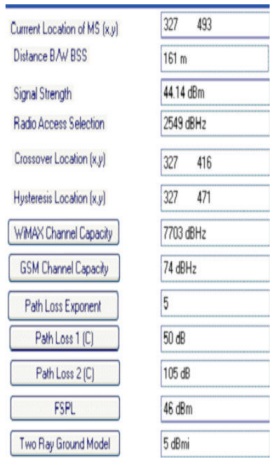


Figure 9: Simulation parameters.

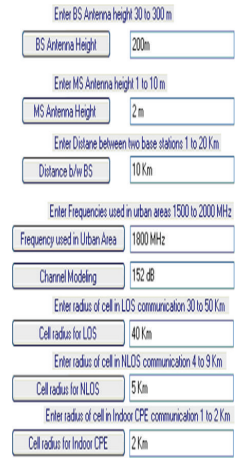


Figure 10: WiMAX system evaluation parameters.

set of commands, and information from the information server to decide that handoff is required or not. If a vertical handoff is required, then the grid finds the target base station of an alternative grid and sends the command to cognitive mobile user to do the handoff to the desired base station. In grid assisted handoff, the grid controls all the steps of disconnection from the previous grid point of attachment to the new grid point of attachment. During the grid assisted handoff process, different MIH events which are exchanged between grid and IEEE 802.21 information server are shown in Figure 12 [26, 46].

6.3. Grid Access Security Steps. Whenever a cognitive mobile station has to connect to point of attachment in a grid, it has to pass through different grid access security steps, in order to authenticate itself. In step 1, the grid access authentication is

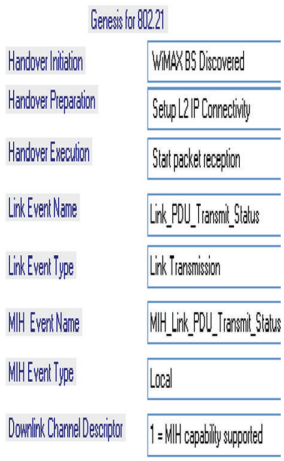


Figure 11: Mobile assisted IEEE 802.21 handoff protocols.

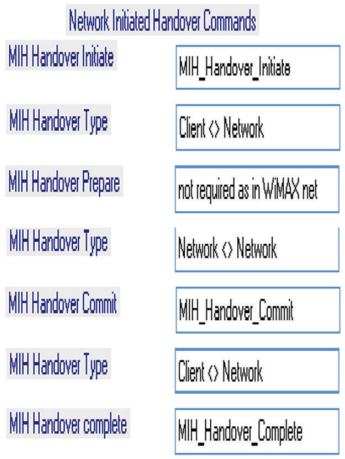


Figure 12: Network assisted IEEE 802.21 handoff protocols.

Figure 13: IEEE 802.21 network access security steps.

used to authenticate the cognitive mobile station to authenticate itself with the authentication server, before connecting to the point of attachment in the grid. In step 2, during secure association session, ciphering keys are used between the cognitive mobile station and authentication server. These keys are used to secure the link layer connectivity between the cognitive mobile station and authentication server. After successful grid access authentication and secure association, in step 3 (which is access control and ciphering), link layer data frames are exchanged between mobile node and point of attachment. Figure 13 shows the grid access security steps [47–49].

The decaying signal graphs due to white spot are shown in Figures 14(a) and 14(b). When Received Signal Strength starts decreasing and the connectivity of the mobile station with the current WiMAX base station starts to fail, it can result in the call drop if an alternative handoff is not initiated on time. The occurrence of the white spots is caused by the change in channel conditions which are sudden and cannot be predicted before time in real time wireless grid architecture.

Figures 15(a) and 15(b) show the graphs of the growing signal of GSM grid, where a cognitive mobile user performs a vertical handoff to neighboring GSM grid base station because there is no alternative WiMAX base station available in the vicinity and horizontal handoff is impossible to perform. The other reason for this vertical handoff is the good Received Signal Strength of the GSM grid base station. Due to the occurrence of white spots, the signal starts to decay and MS interhandoff from WiMAX to GSM base station takes place in order to maintain connectivity during the call, on the basis of good RSSI values from the GSM base station. It has been depicted in Figure 16(a). Figure 16(b) shows an empirical model of signal strength of GSM grid with respect to the distance covered by mobile station towards the GSM base station, where a cognitive mobile user performs a vertical handoff to neighboring GSM grid base station as there is no alternative WiMAX base station available in the vicinity so horizontal handoff is impossible to perform. Another reason for this vertical handoff is the good Received Signal Strength of the GSM grid base station.

Figures 17(a) and 17(b) show the graphs of growing signal of WiMAX grid, where a cognitive mobile user performs a vertical handoff back to the initial WiMAX grid base station as initial WiMAX grid becomes available in the vicinity so no horizontal handoff is required within the GSM grid. Another reason for this vertical handoff back to initial WiMAX grid is the good Received Signal Strength of the WiMAX grid base station. In context to support the real time and killer multimedia applications, the cognitive mobile user will always prefer initial WiMAX grid over alternative GSM grid due to high data rate and greater broadcasting in near LoS communication.

Figure 18 shows the hysteresis margin, when a call drops due to white spot because the Received Signal Strength from serving base station falls below the required limit to main connectivity and therefore a handoff is required to the neighboring WiMAX base station for maintaining connectivity to the WiMAX grid.

Table 3: RSS of WiMAX mobile station user with respect to WiMAX and GSM grids.

Optimized time	Current received signal strength (RSS)	Radio access selection	Current handoff status
$T0$ = above threshold limit	−29 dBm > −31 dBm	1251 dBHz	Currently in WiMAX grid
$T1$ = below threshold limit; white spot occurred $T2$ = searching neighboring WiMAX base station till 20 s and failed	−31 dBm < −30 dBm	600 dBHz	White spot has occurred and efforts to locate neighboring WiMAX base station for horizontal handoff are performed, but handoff fails as neighboring WiMAX base station is not present
$T3$ = vertical handoff to GSM grid	51.31 dBm	4500 dBHz	Neighboring GSM grid located with good RSS; now vertical handoff to GSM base station is initiated
$T4$ = horizontal handoff in GSM grid	44.46 dBm	4500 dBHz	As WiMAX grid is not located yet the mobile user can initiate horizontal handoff within GSM grid to neighboring GSM base station
$T5$ = vertical handoff to WiMAX grid	47.9 dBm	1274 dBHz	Initial WiMAX grid located with good RSS; now vertical handoff to initial WiMAX grid base station is initiated
$T6$ = horizontal handoff in WiMAX grid	44.14 dBm	2549 dBHz	A new WiMAX base station is present so mobile user can initiate horizontal handoff within WiMAX grid to neighboring WiMAX base station.

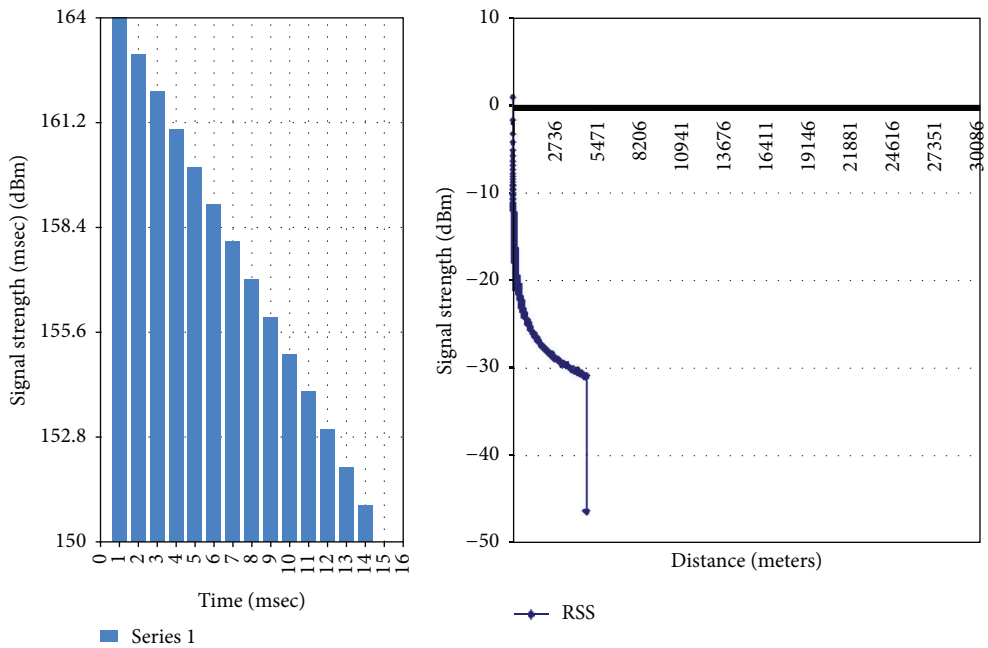


Figure 14: Occurrence of white spot.

Table 3 shows the Received Signal Strength and optimized time of current handoff status of the mobile station user with respect to WiMAX and GSM base stations and scenarios such as horizontal and vertical handoffs within same grid or different wireless GSM and WiMAX grids.

7. WiMAX and Its Relation to Internet of Things (IoT)

First of all, we briefly talk about sensor network because it is the most essential component of Internet of Things (IoT). Sensor network is basically a network of one or more sensor nodes, which communicate with each other through wired and wireless technologies [50, 51]. There are three main architectures of sensor networks found in the literature: flat architecture (in a multihop style, data is transferred from static sensor nodes to the sink node), two-layer architecture (to collect data from sensor nodes, more static and mobile sink nodes are deployed), and three-layer architecture (multiple sensor networks are connected together on the Internet). Therefore, IoT follows three-layer architecture [52]. We refer the readers to [53] where authors have listed 54 application

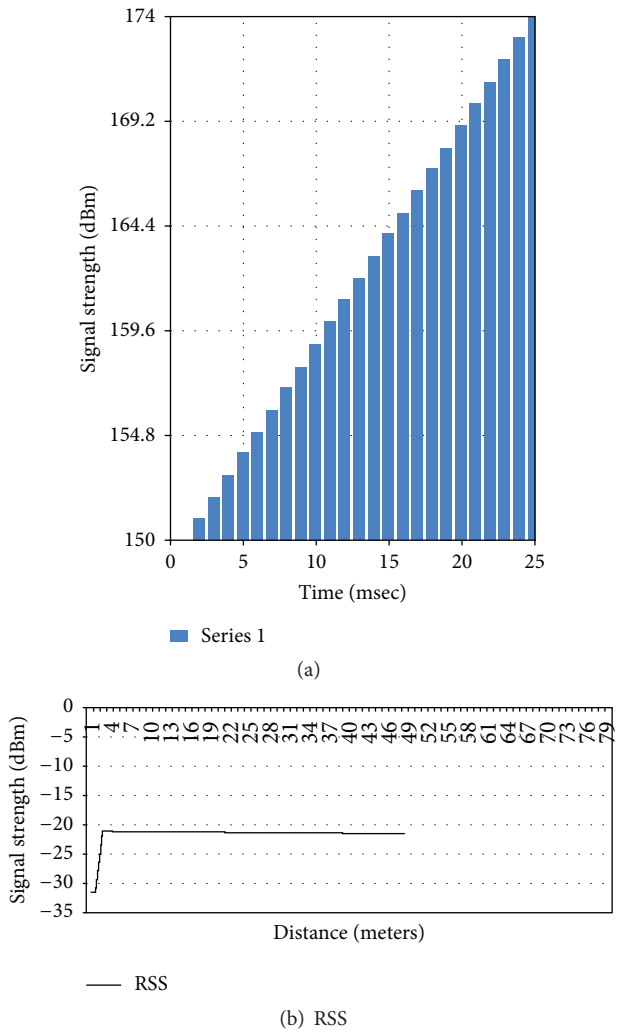


Figure 15: Handover to GSM network.

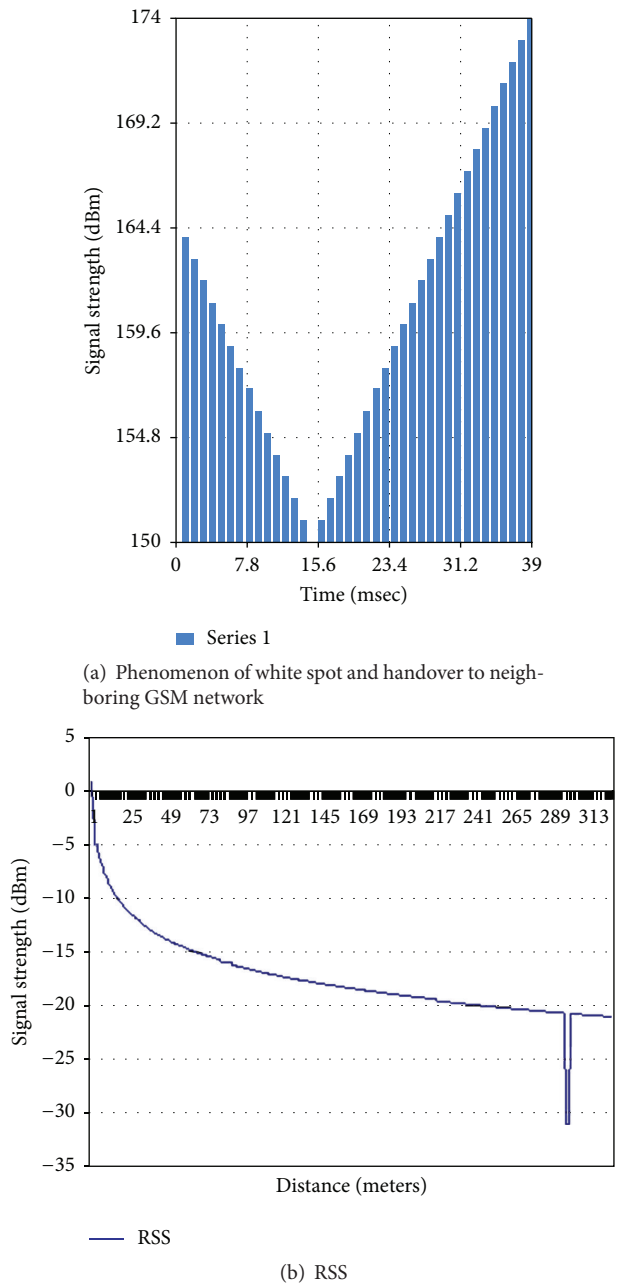


Figure 16: Occurrence of white spot and handoff to GSM.

domains of sensor networks under twelve categories: smart cities, smart environment, smart water, smart metering, security and emergencies, retail, logistics, industrial control, smart agriculture, smart animal farming, domestic and home automation, and eHealth.

The concept of sensor network is existing long time before IoT was introduced. Hence, we can say that sensor network is not a concept that emerged with the IoT. However, the usage of sensor networks was very limited to achieve some specific purposes, such as environment monitoring [54], agriculture [55], medical care [56], event detection [57], and structural health monitoring [58]. Additionally, there are three categories of sensor networks that comprise the IoT [59]: Body Sensor Networks (BSN), Object Sensor Networks (OSN), and Environment Sensor Networks (ESN).

There are other technologies such as ad hoc networks that can also complement the sensing and communication infrastructure in IoT paradigm but these are clearly different from sensor networks and have many weaknesses. The differences are comprehensively discussed in [50]. Most of the sensors deployed today are wireless. There are several major wireless technologies used to build wireless sensor networks: wireless personal area network (WPAN) (e.g., Bluetooth), wireless local area network (WLAN) (e.g., Wi-Fi), Wireless Metropolitan Area Network (WMAN) (e.g., WiMAX), wireless wide area network (WWAN) (e.g., 2G and 3G networks), and satellite network (e.g., GPS) [52].

Most of the sensors are deployed in a wireless fashion and WiMAX is a major wireless platform to provide cost effective multimedia services to the users. On the other side, IoT follows three-layer architecture of sensor networks. The research work carried out in this paper is vital to provide consistent and efficient service to a mobile user in the context of IoT paradigm.

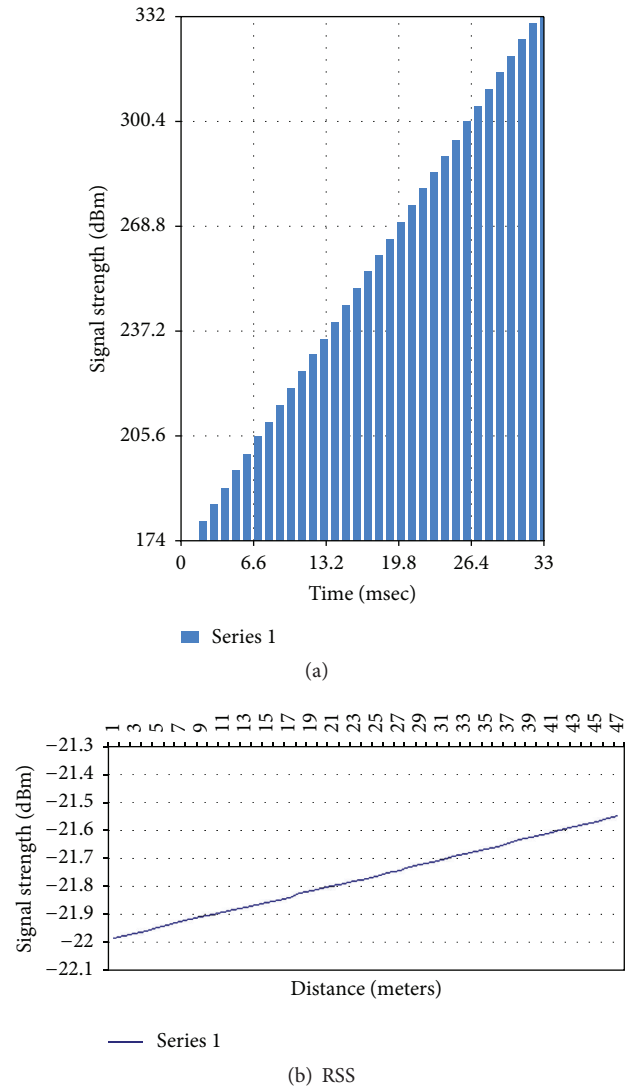


Figure 17: Handoff to WiMAX network.

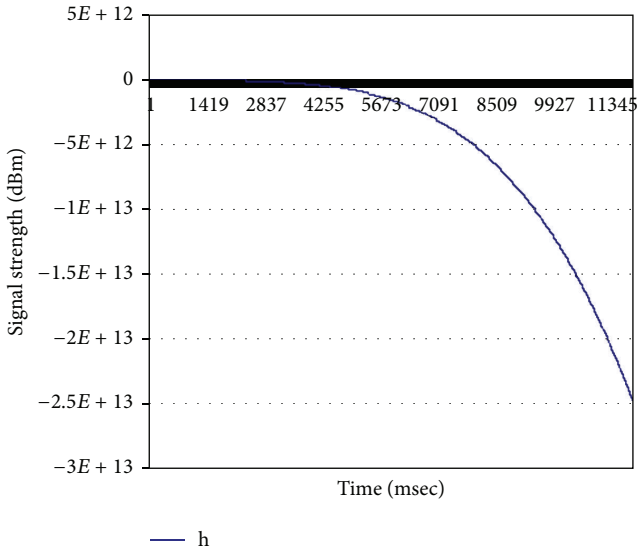


Figure 18: Hysteresis margin.

8. Conclusion and Future Work

The invention of wireless technologies in telecommunications industry has changed our lives and WiMAX is one of the leading technologies in the wireless domain, which has enormous prospects in future. In this paper, we have highlighted how white spot problem in broadcasting operation of WiMAX grids can be eliminated by utilizing horizontal and vertical handoff mechanisms in wireless grids based on the foundation of cognitive heterogeneous approaches. The proposed model based on horizontal and vertical handoff mechanisms is very cost effective in terms of providing economical real time services of voice, data, and streaming video to cognitive mobile subscribers.

Through detailed and comprehensive simulation experiments, we have shown how the problem of white spots can be avoided in different scenarios, that is, when a mobile subscriber switches from WiMAX base station to GSM and vice versa. We have clearly shown through simulation results how the signal grows from WiMAX/GSM grid, when a user performs a handoff in an efficient and timely manner.

The event of interhandoff state between two different technologies from WiMAX to satellite will be simulated in forthcoming research. The main intention for initiation of interhandoff between two different technologies from WiMAX to satellite by mobile subscriber is to retain connectivity in the event of occurrence of white spot, when there is no accessible opportunity for handoff to terrestrial wireless grid technologies.

Abbreviations

OFDMA:	Orthogonal Frequency Division Multiplexing
SOFDMA:	Scalable Orthogonal Frequency Division Multiplexing
MAN:	Metropolitan Area Grid
GPS:	Global Positioning System
GSM:	Global System for Mobile
WLAN:	Wireless Local Area Grid
RTPS:	Real Time Polling Service
UGS:	Unsolicited Grant Service
PSTN:	Public Switched Telephone Network
BPSK:	Binary Phase Shift Keying
QAM:	Quadrature Amplitude Modulation
VoIP:	Voice over IP
SLA:	Service Level Agreement
NRTS:	Nonreal Time Polling Service
FTP:	File Transfer Protocol
AES:	Advanced Encryption Standard
VLAN:	Virtual Local Area Grid
3G:	Third-Generation Mobile Technologies
QoS:	Quality of Service
RSS:	Received Signal Strength
WiMAX:	Worldwide Interoperability for Microwave Access
IEEE:	Institute of Electrical and Electronics Engineers.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

References

[1] A. Conti, O. Andrisano, B. Masini, and A. Bazzi, "Heterogeneous Wireless Communications for Vehicular Grids," Vehicular Grids TechniquesStandards and Applications, 2009.

[2] Westech Communications Inc, *Can WiMAX Address Your Applications?* WiMAX Forum, 2005.

[3] A. A. Minhas, A. Qaddus, and N. Rehmani, "Removal of white spots in WiMAX using cognitive heterogeneous wireless networks," in *Proceedings of the 8th International Conference on High-Capacity Optical Networks and Emerging Technologies (HONET '11)*, pp. 99–105, Riyadh, Saudi Arabia, December 2011.

[4] D. Gray, "Mobile WiMAX—part I: a technical overview and performance evaluation," in *Proceedings of the WiMAX Forum*, August 2006.

[5] S. Bhunia, I. S. Misra, S. K. Sanyal, and A. Kundu, "Performance study of mobile WiMAX network with changing scenarios under different modulation and coding," *International Journal of Communication Systems*, vol. 24, no. 8, pp. 1087–1104, 2011.

[6] M. Riegel, D. Kroeseelberg, A. Chindapol, and D. Premec, *Deploying Mobile WiMAX*, John Wiley & Sons, New York, NY, USA, 2011.

[7] C. Dou, "The maximum available radio resource of a WCDMA downlink," *IEICE Transactions on Communications*, vol. 88, no. 11, pp. 4309–4316, 2005.

[8] S. J. Vaughan-Nichols, "Achieving wireless broadband with WiMax," *Computer*, vol. 37, no. 6, pp. 10–13, 2004.

[9] A. Al Yusuf and M. Murshed, "Performance evaluation of multipath cellular networks in obstacle mobility model for downlink packet video communication," in *Proceedings of the 68th Semi-Annual IEEE Vehicular Technology Conference (VTC '08)*, September 2008.

[10] G. Chadwani, S. N. Datta, and S. Chakrabarti, "Relay assisted cellular system for energy minimization," in *Proceedings of the Annual IEEE India Conference (INDICON '10)*, October 2010.

[11] R. Jaichandran, A. Anthony Irudhayaraj, and J. Emerson Raja, "Effective strategies and optimal solutions for hot spot problem in wireless sensor grids (WSN)," in *Proceedings of the 10th International Conference on Information Science*, June 2010.

[12] K. Gierlowski, J. Woźniak, and K. Nowicki, "A hybrid-mesh solution for coverage issues in WiMAX metropolitan area networks," *Journal of Telecommunications and Information Technology*, vol. 2008, no. 1, pp. 1–24, 2008.

[13] R. Wahl, O. Stäbler, and G. Wölfle, "Propagation model and network simulator for stationary and nomadic WiMAX networks," in *Proceedings of the IEEE 66th Vehicular Technology Conference (VTC-Fall '07)*, pp. 941–945, Baltimore, Md, USA, October 2007.

[14] G. Wolfle, "Propagation model and grid simulator for stationary and nomadic WiMAX grids," in *Proceedings of the IEEE 66th Vehicular Technology Conference*, October 2007.

[15] G. Zaggoulos, M. Tran, and A. Nix, "Mobile WiMAX System Performance—Simulated versus Experimental Results," 2008.

[16] A. Davydov, A. Papathanassiou, and A. Maltsev, "System level comparison of relay and RF repeater based technologies in WiMAX systems," in *Proceedings of the IEEE Mobile WiMAX Symposium (MWS '09)*, pp. 205–208, Napa Valley, Calif, USA, July 2009.

[17] K. K. H. Kan, S. K. C. Chan, and J. K.-Y. Ng, "A dual-channel location estimation system for providing location services based on the GPS and GSM networks," in *Proceedings of the 17th International Conference on Advanced Information Networking and Applications (AINA '03)*, IEEE Computer Society, Crans-Montana, Switzerland, July 2003.

[18] W.-F. Lu and M. Wu, "Analysis of communication traffic characteristics of two-hop-relay cellular system in the dead spots," in *Proceedings of the 8th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD '07)*, vol. 2, pp. 540–545, IEEE, Qingdao, China, July-August 2007.

[19] N. Sultana, M. A. J. B. Bakkre, M. I. Islam, and M. R. Amin, "Performance evaluation of a mobile cellular network with two hop ad-hoc relaying," in *Proceedings of the 12th International Conference on Computer and Information Technology (ICCIT '09)*, pp. 685–690, Dhaka, Bangladesh, December 2009.

[20] L. Jin, G. Chang, D. Jiang, and X. Wang, "Research on the discovery of blind spots in wireless sensor networks based on voronoi diagram," in *Proceedings of the 9th International Conference on Hybrid Intelligent Systems (HIS '09)*, pp. 209–214, Shenyang, China, August 2009.

[21] S. McCann, "Next generation multimode terminals," 2008.

[22] International Engineering Consortium Conference, "Global system for mobile communication," in *Proceedings of the International Engineering Consortium Conference*, 2006.

[23] S. Ahmadi, "An overview of next-generation mobile WiMAX technology," *IEEE Communications Magazine*, vol. 47, no. 6, pp. 84–98, 2009.

[24] G. Alsenmyr, J. Bergström, M. Hagberg et al., "Handover between WCDMA and GSM," *Ericsson Review*, no. 1, pp. 6–11, 2003.

[25] A. Olsson, *Understanding Changing Telecommunications: Building a Successful Telecom Business*, John Wiley & Sons, 2005.

[26] http://www.ieee802.org/21/Tutorials/802%2021-IEEE-Tutorial.ppt.

[27] http://www.teacherschoice.com.au/maths_library/analytical%20geometry/alg_15.htm.

[28] R. K. Martin, A. S. King, R. W. Thomas, and J. Pennington, "Practical limits in RSS-based positioning," in *Proceedings of the 36th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '11)*, pp. 2488–2491, Prague, Czech Republic, May 2011.

[29] L. Jorguseski, R. Litjens, C. Zhiyi, and H. Nikookar, "Radio access selection in multi-radio access systems," in *Proceedings of the 14th IEEE Symposium on Communications and Vehicular Technology in the Benelux*, pp. 1–5, Delft, Netherlands, November 2007.

[30] http://www.vmsk.org/Shannon.pdf.

[31] L. Jorguseski, R. Litjens, C. Zhiyi, and H. Nikookar, "Radio access selection in multi-radio access systems," in *Proceedings of the IEEE Symposium on Communications and Vehicular Technology in the Benelux*, Delft, The Netherlands, November 2007.

[32] W. Afric, B. Zovko-Cihlar, and S. Grgic, "Methodology of path loss calculation using measurement results," in *Proceedings of the 14th International Workshop on Systems, Signals and Image Processing and 6th EURASIP Conference Focused on Speech and Image Processing, Multimedia Communications and Services*, pp. 257–260, IEEE, Maribor, Slovenia, June 2007.

[33] S. Grgic, "Methodology of path loss calculation using measurement results," in *Proceedings of the 14th International Workshop on Systems Signals and Image Processing and 6th EURASIP Conference focused on Speech and Image Processing Multimedia Communications and Services*, June 2007.

[34] "Channel Models for Fixed Wireless Applications," Version 1.0, in IEEE, Croatia, July 2001.

[35] J. Zhou and J. K.-Y. Ng, "Using LDA method to provide mobile location estimation services within a cellular radio network," *Journal of Computers*, vol. 1, no. 7, pp. 41–50, 2006.

[36] J. Zhou, K. M.-K. Chu, and J. K.-Y. Ng, "A probabilistic approach to mobile location estimation within cellular networks," in *Proceedings of the 15th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA '09)*, pp. 341–348, IEEE, Beijing, China, August 2009.

[37] http://www.mike-willis.com/Tutorial/PF4.htm.

[38] S. K. Ray, S. K. Ray, K. Pawlikowski1, A. McInnes, and H. Sirisena, "Self-tracking mobile station controls its fast handover in mobile WiMAX," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '10)*, pp. 1–6, Sydney, Australia, April 2010.

[39] Y.-C. Chen, "An MAC protocol for wireless ad-hoc grids using smart antennas," in *Proceedings of the 11th International Conference on Parallel and Distributed Systems (ICPADS '05)*, Fukuoka, Japan, July 2005.

[40] WiMAX Forum, *WiMAX System Evaluation Methodology*, Version 2.0, 2007.

[41] N. F. Rehmani, A. A. Minhas, and M. M. Alam, "Time domain efficient handoff scheme for mobile WiMAX," *Australian Journal of Basic & Applied Sciences*, vol. 5, no. 8, pp. 1135–1150, 2011.

[42] A. Mourad and I. Gutierrez, "System level evaluation for WiMAX IEEE 802.16m," in *Proceedings of the IEEE 28th International Performance Computing and Communications Conference (IPCCC '09)*, pp. 418–424, IEEE, Scottsdale, Ariz, USA, December 2009.

[43] http://www.msdn.microsoft.com.

[44] J. Cao, J. Liu, S. Zhu, and Y. Zhang, "A wide broadcasting wireless communication scheme for the intelligent distribution grid," in *Proceedings of the World Congress on Information and Communication Technologies*, Trivandrum, India, October-November 2012.

[45] E. Crozier and A. Klein, "WiMAX's technology for LOS and NLOS environments," in *Proceedings of the WiMAX Forum*, San Diego, Calif, USA, June 2013.

[46] S.-J. Yoo, "Predictive handover mechanism based on required time estimation in heterogeneous wireless grids," in *Proceedings of the IEEE Military Communications Conference (MILCOM '08)*, San Diego, Calif, USA, November 2008.

[47] http://ieee802.org/21/Tutorials/802%2021-IEEE-Security_Tutorial.ppt.

[48] C. Makaya, "Mobile virtual private grids: issues and challenges," in *Emerging Wireless Grids Concepts Techniques and Applications*, 2011.

[49] E. Qi, S. Bangolae, K. Sood, and J. Walker, "BSS transition optimizations and analysis for VoIP over WLAN," *Wireless Personal Communications*, vol. 43, no. 3, pp. 907–918, 2007.

[50] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102–114, 2002.

[51] M. Alnuaimi, F. Sallabi, and K. Shuaib, "A survey of wireless multimedia sensor networks challenges and solutions," in *Proceedings of the International Conference on Innovations in Information Technology (IIT '11)*, pp. 191–196, IEEE, Abu Dhabi, The United Arab Emirates, April 2011.

[52] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos, "Context aware computing for the internet of things: a survey," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 414–454, 2014.

[53] A. Asin and D. Gascon, "50 Sensor applications for a smarter world," Tech. Rep., Libelium Comunicaciones Distribuidas, Zaragoza, Spain, 2012, http://www.libelium.com/top_50_iot_sensor_applications_ranking/pdf.

[54] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications (WSNA '02)*, pp. 88–97, ACM, New York, NY, USA, September 2002.

[55] J. Burrell, T. Brooke, and R. Beckwith, "Vineyard computing: sensor networks in agricultural production," *IEEE Pervasive Computing*, vol. 3, no. 1, pp. 38–45, 2004.

[56] D. Malan, T. Fulford-Jones, M. Welsh, and S. Moulton, "CodeBlue: an ad hoc sensor network infrastructure for emergency medical care," in *Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks*, 2004.

[57] S. Rooney, D. Bauer, and P. Scotton, "Techniques for integrating sensors into the enterprise network," *IEEE Transactions on Network and Service Management*, vol. 3, no. 1, pp. 43–52, 2006.

[58] A. R. Da Rocha, F. C. Delicato, J. N. de Souza, D. G. Gomes, and L. Pirmez, "A semantic middleware for autonomic wireless sensor networks," in *Proceedings of the Workshop on Middleware for Ubiquitous and Pervasive Systems (WMUPS '09)*, pp. 19–25, ACM, Dublin, Ireland, June 2009.

[59] A. Gluhak and W. Schott, "A WSN system architecture to capture context information for beyond 3G communication systems," in *Proceedings of the 3rd International Conference on Intelligent Sensors, Sensor Networks and Information (ISSNIP '07)*, pp. 49–54, IEEE, Melbourne, Australia, December 2007.

Hindawi Publishing Corporation
Mobile Information Systems
Volume 2016, Article ID 3038515, 20 pages
http://dx.doi.org/10.1155/2016/3038515

Research Article

Wearable Device Control Platform Technology for Network Application Development

Heejung Kim,[1,2] Misun Ahn,[2] Seunghyun Hong,[2] SeungGwan Lee,[3] and Sungwon Lee[2]

[1]*Korea Telecom, 206 Jungja-dong, Bundang-gu, Seongnam-si, Gyeonggi-do 463-711, Republic of Korea*
[2]*Department of Computer Engineering, Kyung Hee University, 1 Seocheon-dong, Giheung-gu, Yongin-si, Gyeonggi-do 446-701, Republic of Korea*
[3]*Humanitas College, Kyung Hee University, 1 Seocheon-dong, Giheung-gu, Yongin-si, Gyeonggi-do 446-701, Republic of Korea*

Correspondence should be addressed to Sungwon Lee; drsungwon@khu.ac.kr

Received 4 August 2015; Revised 24 November 2015; Accepted 6 January 2016

Academic Editor: Yassine Hadjadj-Aoul

Application development platform is the most important environment in IT industry. There are a variety of platforms. Although the native development enables application to optimize, various languages and software development kits need to be acquired according to the device. The coexistence of smart devices and platforms has rendered the native development approach time and cost consuming. Cross-platform development emerged as a response to these issues. These platforms generate applications for multiple devices based on web languages. Nevertheless, development requires additional implementation based on a native language because of the coverage and functions of supported application programming interfaces (APIs). Wearable devices have recently attracted considerable attention. These devices only support Bluetooth-based interdevice communication, thereby making communication and device control impossible beyond a certain range. We propose Network Application Agent (NetApp-Agent) in order to overcome issues. NetApp-Agent based on the Cordova is a wearable device control platform for the development of network applications, controls input/output functions of smartphones and wearable/IoT through the Cordova and Native API, and enables device control and information exchange by external users by offering a self-defined API. We confirmed the efficiency of the proposed platform through experiments and a qualitative assessment of its implementation.

1. Introduction

The market for smartphones has experienced explosive growth since the development of Apple's iPhone and Samsung's Omnia2 to the extent that smart devices have now become vital to daily life. Furthermore, wearable devices have attracted considerable attention as the next generation of mobile technology that will replace smartphones. The research service BI Intelligence has predicted that the market for wearable devices will grow to approximately $12 billion by 2018 [1].

A "wearable device" refers to a small electronic device that can be worn on the body so that a user can freely use it even when moving. Google Glass, Samsung Galaxy Gear, and Sony SmartWatch belong to this category. Furthermore, even sports equipment companies such as Nike and Adidas have lately begun introducing innovative products and services in the wearable device market [2].

The Internet of Things (IoT) is a recent technology for collecting data and transferring data through sensors and adding a communication function to every object. The market size of IoT was estimated to be $203.1 in 2013 and is expected to reach $1 trillion, with an average annual growth of 21.8%, by 2022 [3]. In December 2013, subscribers of smartphones exceeded 37.5 million in Korea [4]. Along with the popularity of smartphones and wearable devices, the market for mobile applications is steadily growing as well [5].

Cross-Platform Development. Due to the emergence of smartphones and the existence of multiple platforms, developers in the stage of application development have to build platform-specific environments, use multiple programming languages,

and learn the relevant supporting application programming interfaces (APIs). These constraints lead to wasted time and effort and increase the cost of application development. In order to solve such problems, mobile programming is in the process of standardization. Cross-platform development frameworks, such as Cordova [6] and Titanium [7], have garnered considerable attention as a solution. The Association for Computing Machinery (ACM), the Institute of Electrical and Electronics Engineers (IEEE), and related organizations are publishing an increasing number of studies on platform research that analyze the characteristics of cross-platform development frameworks. This suggests that application development on a single, unified platform will become possible in the near future [8, 9].

Research Challenges and Contributions. Cross-platform development is an attribute that assists the development of applications using web languages such as HyperText Markup Language 5 (HTML5), Cascading Style Sheets (CSS), and JavaScript. Its greatest benefit is that it provides an application with a single source that is feasible on multiple mobile platforms, which increases development productivity. However, the range of APIs in Cordova and Titanium is not sufficiently wide, especially with regard to supporting wearable devices and IoT devices, since they are still in their developmental stages. Developers are thus inevitably required to learn the native language supported by the platform of the relevant device and initiate follow-up development because cross-platform development alone is insufficient for application development.

Furthermore, wearable/IoT devices support Bluetooth technology for communication among them. Bluetooth is a communication technology over short distances, due to which it is impossible to communicate with wearable/IoT devices beyond a limited range [10].

In this regard, we propose a "Network Application Agent" (NetApp-Agent) platform that integrates a development environment for wearable devices and supports Internet Protocol-(IP-) based communication. NetApp-Agent is a smart device platform for network application development that allows outside users to control input/output (I/O) functions and exchange data. This is because it is based on the Apache Cordova platform and uses Cordova API and a Native API that enable the I/O function in smartphones and wearable devices as well as the I/O function control of IoT devices by providing a self-defined API.

The potential benefits of our proposed platform are as follows. First, it eases application development by supporting integrated development environments that supply essential APIs or facilitate device development. Second, it enables wearable devices that use Bluetooth to communicate with the outside by supporting IP communication based on WebSocket.

The outline of the remainder of this paper is as follows. Section 2 examines the benefits and drawbacks of the existing development platforms. In Section 3, we introduce NetApp-Agent, our proposed integrated development platform, together with its structure and features. Section 4 presents the results of our experiments involving NetApp-Agent as well as a qualitative assessment of its implementation. Finally, we offer our conclusions and recommendations for further research in Section 5.

2. Review of Existing Approaches and Issues

Approaches to mobile application development can be divided into three major categories: native applications (native apps), web applications (web apps), and hybrid applications (hybrid apps). Native apps involve application development on the platform of each device, whereas web apps utilize HTML, JavaScript, and CSS. Hybrid apps assume the form of native apps, but all or part of its internal configuration is developed in a web app environment [11–13].

In this section, we examine the development of native and hybrid apps together with the features and challenges of a network application development system that provides an IoT development platform and cloud service.

2.1. Native Development Platform. A native development platform involves developing applications on the platform of each device, such as the iPhone, the Android phone, and the Windows Phone, which operate in machine language code. It ensures optimized application performance. However, it has a few major disadvantages: it builds a different development environment for each platform, and the developer needs to learn the relevant development language and software development kit (SDK). We examine the native development environments for both the Sony and the Pebble SmartWatch, which are representative of wearable devices.

2.1.1. Sony SmartWatch. Sony SmartWatch is based on Android 4.0 operating system, and Java is used as its development language. A developer cannot check his/her developed screen on the actual device, but on a computer with a separate emulator. The procedure for developing applications for the Sony SmartWatch is as follows [14].

A development environment must first be constructed. Sony SmartWatch provides the Sony Add-on SDK as an additional installation to the existing Android SDK. Therefore, the Android's development environment needs to be built in advance for SmartWatch development. The Java Development Kit (JDK) and the integrated development tool Eclipse are installed in order to create the Java Runtime Environment.

Following this, the Android Development Tools (ADT) plugin is installed on Eclipse to support the Android system together with the Android SDK. The Android development environment is then constructed [15]. The development environment for the SmartWatch is finally constructed after downloading the Add-on SDK from Sony's developers' website and installing it on Eclipse.

Furthermore, the system structure and the API supporting the development need to be learned. The system architecture of Sony SmartWatch is shown in Figure 1 and can be divided into three major components: Smart Extension, Host Application, and Accessory. Smart Extension refers to the application to perform in wearable device.

Table 1: Smart Extension API of Sony SmartWatch.

API	Description
Registration and Capabilities API	Provide API data of SmartWatch/Smart Extension
Notification API	Notify the event that occurred in smartphone to the device
Control API	Control the display of device
Sensor API	Transmit sensor data
Widget API	Preview contents

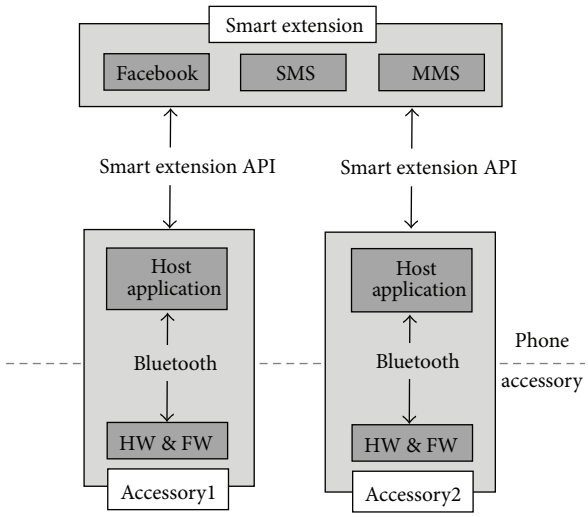


Figure 1: System architecture of Sony SmartWatch.

This application utilizes the Smart Extension API to communicate with the Host Application of the wearable device. The Host Application is installed in smartphones and connected to the wearable device using Bluetooth technology. The Smart Extension API for the development of Sony SmartWatch must be learned in sequence.

Table 1 lists various supporting APIs. The Registration API and the Capabilities API are employed when the Host Application provides data for the API of the SmartWatch or Smart Extension provides its data on the API to the Host Application. The Notification API is used when the Host Application notifies the event occurred in smartphones to device, whereas the Control API enables Smart Extension application to the control display of the device. Control API in particular is crucial to controlling the display or the light-emitting diode (LED) of the device and to processing key events or touching events that require close attention. The Sensor API transmits data from the accelerometer and the illumination sensor of the device to the Smart Extension application. The Widget API affords content preview.

The lengthy and complex procedure described above concludes the preparation for the development of an operating application for Sony SmartWatch. However, information regarding implemented classes, functions, and variables still needs to be checked, and extensive research needs to be conducted on the API implementation code and the sample code, along with code analysis, by using API reference documents for application development.

2.1.2. Pebble SmartWatch. The Pebble SmartWatch functions on an independent Pebble operating system (OS) that is compatible with both Android and iOS. Pebble supports different types of languages, such as JavaScript and Objective-C, where the latter is primarily used for application development. Moreover, a Pebble SDK is provided for development. The procedure for developing applications for Pebble SmartWatch can be divided into two parts.

First, the application development environment must be constructed. Pebble SDK can be installed on Mac OS X and Linux. We assume the construction of a development environment conducted on Mac OS X. The Pebble SDK is downloaded from the Pebble developers' website and the Pebble ARM Toolchain is installed after the installation of Xcode Command Line Tools [16], which is a developers' command line tool. Development environment construction is completed after building in Python library because Pebble SDK is based on Python. Pebble is compatible with smartphones that use Android and iOS platforms. Therefore, the development environment of a smartphone application should be built and an SDK called PebbleKit should then be installed to create an application that is in sync with the smartphone [17].

Second, a development support API must be learned. The Pebble API consists of Pebble Watch App SDK (for SmartWatch applications), PebbleKit Android (for Android), and PebbleKit iOS (for iOS). In order to develop an application that is compatible with smartphones that use Android OS, one needs to learn Watch App SDK and PebbleKit for Android. The scope of the supporting SDK is presented in Table 2.

When developing a Pebble application, two issues need careful scrutiny in addition to the construction of the development environment and the examination of the supporting APIs. First, the Pebble SmartWatch does not support Korean characters. Hence, expressions in Korean need to be considered when developing an application that sends text messages or notifications for Social Network Services (SNS) to Pebble devices in Korean. Second, the image format of Pebble is problematic. Pebble uses its own image format, called Pebble Binary Image (PBI), when displaying an image on the screen. PBI represents each pixel using one bit that contains image information in the header file. Thus, a developer must create a tool for image conversion to enable images in Pebble Watch.

The relevant Pebble Watch application is then installed on a computer terminal by the way of inputting build and install command. Pebble Watch should be connected to a smartphone through Bluetooth for application installation, following which the computer and the smartphone that have already progressed in development should be connected to the same Wi-Fi network. Therefore, developers should pay particular attention to network configuration when installing the application.

Table 2: Watch App SDK of Pebble SmartWatch.

Watch App SDK Pebble SmartWatch	
Foundation	(i) App (ii) Media Utilities (iii) Timer (iv) Wall Time (v) Math (vi) Dictionary (vii) AppMessage (viii) Resources (ix) AppSync (x) Logging (xi) App Communication
Graphics	(i) Graphics Context (ii) Drawing Primitives (iii) Graphics Types (iv) Drawing Text (v) Fonts
User interface	(i) Layers (ii) Animation (iii) Window (iv) Vibes (v) Light
Standard C	(i) Math (ii) Memory (iii) Format (iv) String (v) Time

2.1.3. Problems with Native Development Platforms. In the current mobile market, several mobile platforms exist, for example, Android, iOS, and Windows Mobile. Thus, manufacturers produce smartphones on a variety of platforms. The coexistence of different platforms has led developers to establish suitable development environments for each platform when creating applications and learn the relevant development languages, the SDK, and the API.

Table 3 shows diverse development environments according to types of wearable device. As shown in the table, a developer needs to learn ten programming languages and seven APIs for an application adaptable to four devices. With the growing trend of wearable devices, more and more devices are expected to be introduced. Moreover, time spent on application development will increase in proportion to the number of devices. To solve these kinds of problems, an integrated development platform is required.

2.2. Cross-Platform Mobile Development Framework. Table 4 shows that the development of mobile applications can be divided into three types: native app, web app, and hybrid app.

As shown in Section 2.1, native apps guarantee optimized application performance. However, they have a few constraints given that they require building a different development environment for each platform and that the developer needs to learn the relevant development language and the SDK.

Application development in web apps is based on widespread Internet technologies, such as HTML, JavaScript, and CSS. The advantage of web apps is that they can attract and train developers, since learning a development language is relatively easy. On the other hand, difficulty in hardware control, slow speed of applications, and vulnerability on networks are the major weaknesses of web apps. Applications developed by hybrid apps assume the form of native apps, but all or part of their internal configuration is developed in web app. The final form of the application is a binary file, which has the same file extension but is developed using web languages, such as HTML5, CSS, and JavaScript.

Hybrid apps improve development productivity because they can be operated in various mobile platform using a single source. Hardware control is also possible. The appropriate incorporation of two apps' advantages brings forth a new development strategy. Cross-platform is applied as a development tool for hybrid apps. In this section, we discuss Cordova and Titanium, two typical instances of cross-platform development.

2.2.1. Cordova. Cordova is an open-source framework that enables hybrid application development. It was first developed as "Phone Gap" by Nitobi and was subsequently taken over by Adobe in October 2011. Following the takeover, it reinforced the open-source policy with the development of the Apache license, and then he changed the name of the application to Cordova from version 1.4 onward. Additional functions for Cordova are developed as plug-ins that are shared in open-source communities. Moreover, Cordova supports seven smartphone platforms, Android, BlackBerry, Firefox OS, iOS, Windows Phone, Windows 8, and Tizen, with high product quality that renders it the most competitive among cross-platforms [18, 19].

The structure of Cordova's applications is shown in Figure 2. The developer creates applications using HTML5, CSS, and JavaScript. The completed codes are then packaged through the Cordova library. The application in packaging is distributed to the device in which web-kit provided browser is equipped.

Cordova provides APIs shown in Table 5. The APIs are called by JavaScript, whereas the JavaScript engine exchanges data with the native engine using string type. However, Cordova has a limited range of APIs because of its incomplete platform. With regard to network APIs, it only checks the status of the connection to Wi-Fi or cellular data. APIs related to Bluetooth are not yet available. As a consequence, implementation through the native language is inevitable in order to call particular functions, even though the application is developed using the Cordova platform [6].

2.2.2. Titanium. Titanium is a cross-platform development framework created by Appcelerator. Unlike Cordova which concentrates on the mobile application development, it is possible to develop desktop applications using Titanium Studio in addition to mobile applications. We focus on cross-platform mobile application development in this paper.

Figure 3 shows the mobile application development process using Titanium Studio. Titanium delivers a development tool called Titanium Studio. Therefore, application

Table 3: Development environments for each wearable device.

Types of device	Operating system	Supporting languages	SDK/API
Google Glass	Android 4.0.4	Go, Java, .NET, PHP, Python, Ruby	Google Mirror API
Sony SmartWatch	Android 4.0	Java	Sony Add-on SDK/Smart Extension API
Pebble SmartWatch	Pebble OS	C, JavaScript (Objective-C/Java)	Pebble SDK PebbleKit
Samsung Gear	Tizen OS	Java, HTML, JavaScript	Tizen SDK Samsung Mobile SDK
Total	**4**	**10**	**7**

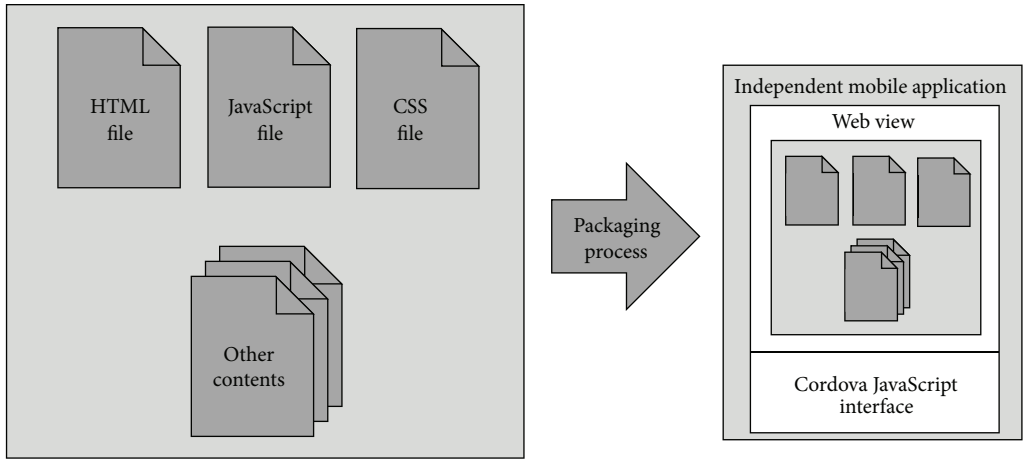


Figure 2: Structure of Cordova.

Table 4: Features of each mobile application development method.

Feature	Native App	Hybrid App	Web App
Development language	Native only	Native and web or web only	Web only
Code portability and optimization	None	High	High
Access device-specific features	High	Medium	Low
Leverage existing knowledge	Low	High	High
Advanced graphics	High	Medium	Medium
Upgrade flexibility	Low	Medium	High
Installation experience	High	High	Medium

development begins after the construction of the development environment by installing Titanium Studio. The application development code is written in HTML and JavaScript and utilizes the interpreter method in case of code translation. In short, Titanium Bridge, which is embodied in Titanium SDK, substitutes Titanium API to a native development language when the developer calls a Titanium API in an application written in JavaScript. Thus, it creates an almost identical product with the application written in native code. Nonetheless, the spectrum of the supporting platform is quite narrow because it is difficult to format a JavaScript engine for the substitution of the development language into each platform [7, 20].

Table 6 presents the list of Titanium mobile APIs. Titanium has a limited domain of supporting APIs, like Cordova. From a communication aspect, it only supplies APIs for socket and HTTP client production and communication. The range of the hardware module control is restricted to camera, audio, and video control [21].

2.3. Internet of Things (IoT) Development Platform. IoT refers to a technology that collects and transfers data by installing sensors and adding network connectivity to every object. To satisfy rising user demand, it generates massive amounts of information through smart sensors installed in smart devices such as smartphones, tablet PCs, and smart TVs. The development of mobile devices of every kind together with built-in sensors has ushered in the age of IoT [22].

Smart devices, such as smart TVs, smartphones, and wearable devices, communicate by forming a network structure in IoT environment. Bluetooth, Wi-Fi, ZigBee [23], and Near Field Communication (NFC) [24] are typical close-range wireless communication technologies used for communication among IoT devices. Bluetooth Smart, widely known as Bluetooth Low Energy (BLE), is particularly widely used.

Table 5: Cordova plug-in API.

API	Description
Battery Status	Check battery status of device
Camera	Take picture and browse gallery
Contacts	Search contacts and add/edit contacts
Device	Provide information about device
Device Motion (Accelerometer)	Provide information on accelerometer sensor
Device Orientation (Compass)	Provide information on compass sensor
Dialogs	Show notification of device
FileSystem	Access file system of device
File Transfer	Receive and transfer file
Geolocation	Provide information on location
Globalization	Provide international expressions
InAppBrowser	Run new application browser
Media	Record and play voice file
Media Capture	Capture media files
Network Information (Connection)	Provide information on network status and connectivity to cellular data and Wi-Fi
Splashscreen	Show/hide start screen of application program
Vibration	Generate vibration on device
StatusBar	Hide/configure status bar background
Whitelist	Whitelist network requests
Legacy Whitelist	Use the old style of whitelist

Table 6: Titanium mobile API.

API	Description
App	Provide information on application and system event
Calendar	Access native calendar
Cloud	Access ACS (Appcelerator Cloud Services)
Contacts	Search contacts and add/edit contacts
Database	Access to SQLite DB within application and produce DB
Facebook	Support application connection to Facebook
Filesystem	Access device's file and folder
Geolocation	Provide information on device's location
Map	Produce native map
Media	Call out media-related function of device (Audio, Video, ImageView, Camera, and Photo Gallery API)
Network	Produce Socket, HTTPClient, TCPSocket, and support communication
Platform	Access function per device's platform (check battery status)
UI	Form UI of application
XML	XML-based content parsing

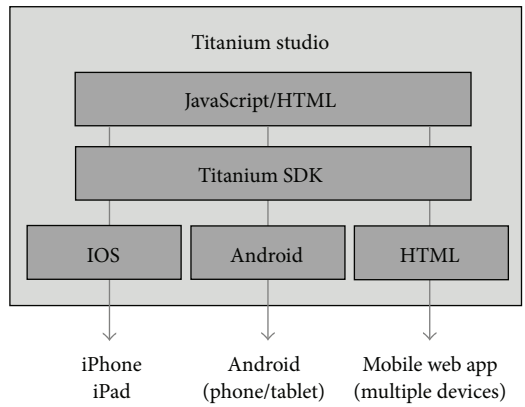


Figure 3: Mobile application development process using Titanium Studio.

The structure of Bluetooth Smart consists of traditional Bluetooth, Bluetooth Smart Ready, and Bluetooth Smart, as shown in Figure 4. Bluetooth Smart Ready is in the form of a hub device that can be connected to both traditional Bluetooth and Bluetooth Smart devices. Yet, Bluetooth communicates through Mac addresses when a device does not have an IP. This leads to a problem where deviating from the given range renders communication among devices impossible.

2.4. Network Application Development. Cloud computing refers to a structure of computing systems where I/O operations are carried out through a user's device, but information analysis, process, storage, management, and distribution are accomplished in another space called a "cloud." Figure 5 shows an outline of this structure [25].

Cloud technology can be divided into three parts, Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS), from the perspective of service provision. A representative service of IaaS is Amazon's EC2, which provides server computation and hardware storage online [26]. PaaS represents platform supply for host and enterprise services. PaaS makes use of resources of IaaS to support certain platforms, including Microsoft's Azure and Google's App Engine [27, 28]. SaaS is a basic technology that supplies user-driven service. It can be divided into applied software service, web-based service, and component-based service. The sole function of several SaaS companies that research maps, images, videos, documents, mail, and so forth is service, and they distribute representational state transfer- (REST-) based Open APIs, numerous platforms, and development languages supporting SDK. IT companies, such as Google, Facebook, Baidu, Kakao, and NAVER, cater to many types of Open APIs for developers [29, 30].

For cloud service development, Amazon offers smartphone platforms for Android and SDKs for different programming languages (Java, .NET, PHP, Ruby, etc.) at the same time [31]. The Azure platform also provides program APIs in the structure of .NET, Node.js, Java, and PHP [32]. SaaS companies distribute REST-based Open APIs and development languages supporting SDK. However, most SDKs and

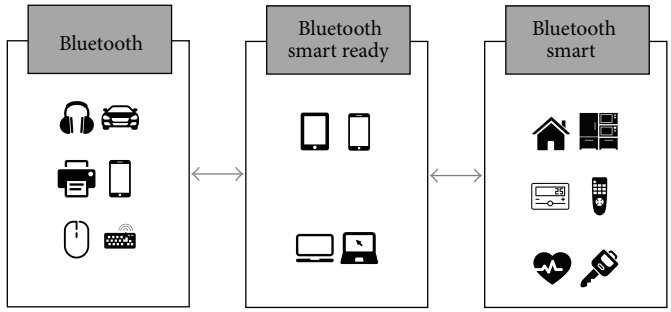


Figure 4: Structure of Bluetooth Smart.

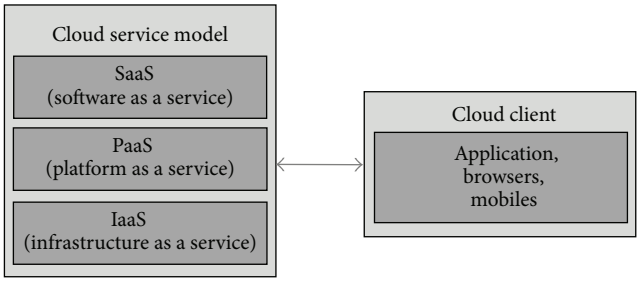


Figure 5: Structure of cloud service model.

APIs focus on establishing cloud servers. Thus, the client application needs to be separately implemented on each platform of a service-providing device where a native app is used. This implies that a network application that provides cloud service must construct a development environment for each platform and learn the development language to create an application, in a similar manner to the native app development process.

3. Proposal

Sections 1 and 2 examined the characteristics of and challenges faced by the existing development platforms. In Section 3, we suggest a solution for the foregoing problems and discuss the detailed structure of our proposed platform, followed by a discussion of scenarios to which the technique is applicable.

3.1. Wearable Device Control Platform

3.1.1. Network Application Agent (NetApp-Agent). In this paper, we propose a wearable device control platform called Network Application Agent (NetApp-Agent) for network application development. Figure 6 shows our proposed platform diagram. NetApp-Agent, in the form of a smartphone application, controls the I/O function of the smartphones, its connected wearable devices, and IoT devices by following commands from a remote controller. In order to do this, a self-defined JavaScript Object Notation- (JSON-) based NetApp-Agent API equipped with WebSocket is provided to the user for remote control.

3.1.2. Network Application Agent API (NetApp-Agent API). The user connects to NetApp-Agent through an IP. NetApp-Agent controls the I/O functions of smart devices or linked wearable devices and requests the necessary information through supporting APIs. Therefore, the proposed platform allows the developer to easily create a device without background knowledge of wearable or IoT devices. Outside users can be any programmable device that wants to communicate through IP-based cloud server, tablet PC, or smartphone using a supporting API.

3.1.3. Structure of Proposed Platform. Figure 7 shows the detailed structure of our proposed platform. NetApp-Agent aims to be a cross-platform development framework that can run applications regardless of the type of mobile platform in question. However, as mentioned before, Cordova's API is limited in its range of support. Our proposed platform amplifies the range of APIs by the binary use of Cordova and Native device wrappers. Functions such as battery check, vibration, and acquisition of global positioning system (GPS) information are developed with the API, whereas I/O functions such as sensors and display are developed with the Native API. For network connection control, Cordova provides an API for receiving cellular and Wi-Fi connection information. Thus, cellular communication and Wi-Fi connection information is implemented using Cordova API, and connection control, such as obtaining Bluetooth connection information and simple On/Off functions of the network, is implemented by using Native API. Wearable devices and IoT devices do not support Cordova API, and thus related functions are established through the Native API. The established functions are controllable by a self-defined JSON-based NetApp-Agent API, where NetApp-Agent provides its implemented API to the outside user.

The WebSocket in NetApp-Agent allows IP-based communication between the device and the user. The outside user thus has remote access to control I/O functions of the device. The inventory database saves the device's connectivity and specification information related to the smartphone. If remotely requested to connect, NetApp-Agent provides its stored information to the outside user. The user then confirms the desired device to control. An intelligence-processing module supports the establishment of intelligible operations in the device. Pebble's SmartWatch does not

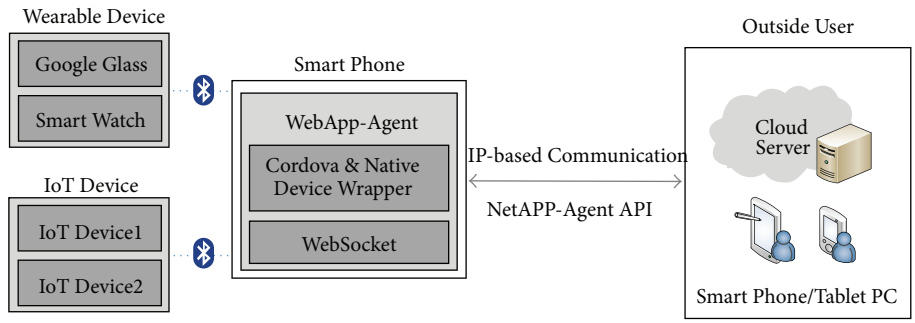


Figure 6: Proposed platform diagram.

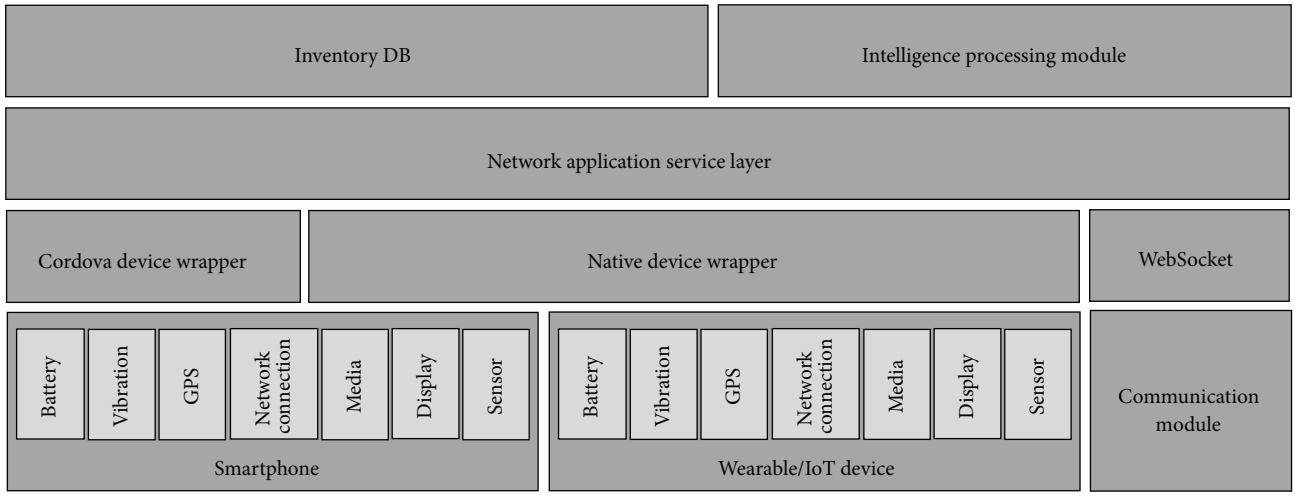


Figure 7: Detailed structure of proposed platform.

support Korean characters and uses a self-defined Perceptual-Backdrop Image (PBI) format for images. As a consequence, if an outside user sends a message containing Korean characters, NetApp-Agent automatically transmits it after translating it into an image. NetApp-Agent does not simply read or write the information of the device connected to the smartphone, but it provides intelligent service through the intelligence-processing module.

3.2. Cloud Service Application Scenario for Proposed Platform. Figure 8 shows the interface of interaction between NetApp-Agent and the cloud server. There is a network application in the cloud server for communication with NetApp-Agent.

The network application uses the NetApp-Agent API for the cloud server to supply the service demanded by the user without modification. We examine here the application of the proposed platform to a cloud service scenario. Possible scenarios are divided into three parts according to their function.

3.2.1. Acquisition of QR Code Information Using Cloud Server. Figure 9 shows a scenario concerning the analysis of Quick Response (QR) code spotted in the camera of a wearable device. The code is processed in the cloud server, and the extracted information is sent back to the wearable device for display.

The video containing the QR code is sent to the cloud server. Images are extracted from the video and form the basis of QR code recognition. The cloud server simultaneously runs various code extraction methods to improve QR code recognition rate. There are three methods to recognize QR code: the traditional method, recognition using the image of the object itself, and an analysis of the similarity of images. In the analysis of similarity, information related to previously stored QR images and location information on the product by beacon are used together.

The information extracted from the QR code goes through a series of processes to be displayed on the screen of the wearable device. In this scenario, the wearable device is only used as a tool for I/O, while the extraction and processing of information from the QR code are carried out in the cloud server. The cloud server can implement an application that displays the information on the wearable device using the NetApp-Agent API without having to develop a native app.

3.2.2. Voice and Video-Sharing Scenario Involving Smart Device and Cloud Server. Figure 10 shows a scenario for voice and video sharing between a smart device and the cloud

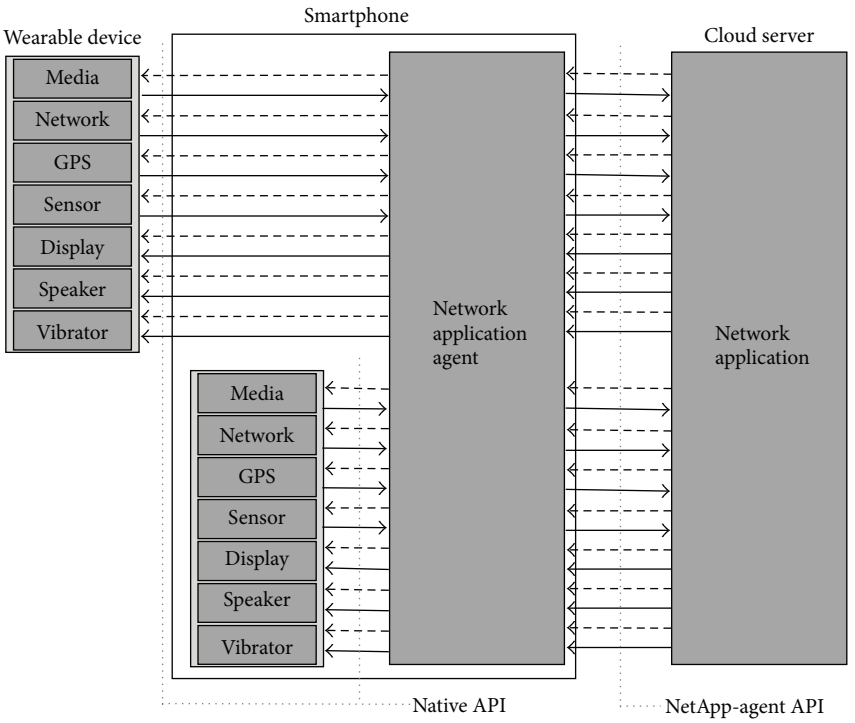


Figure 8: The NetApp-Agent interface.

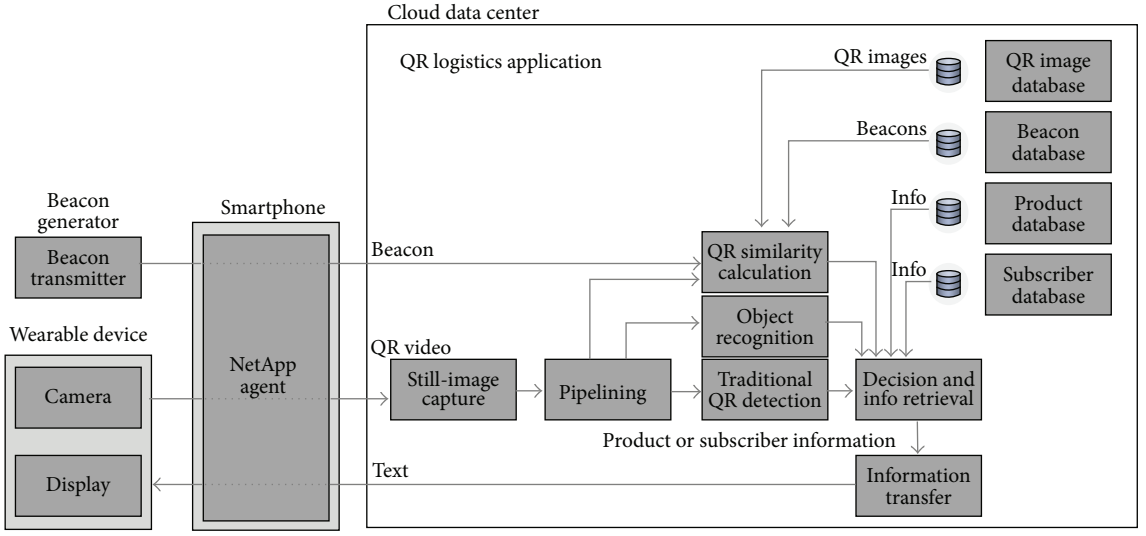


Figure 9: QR code analysis scenario using cloud server.

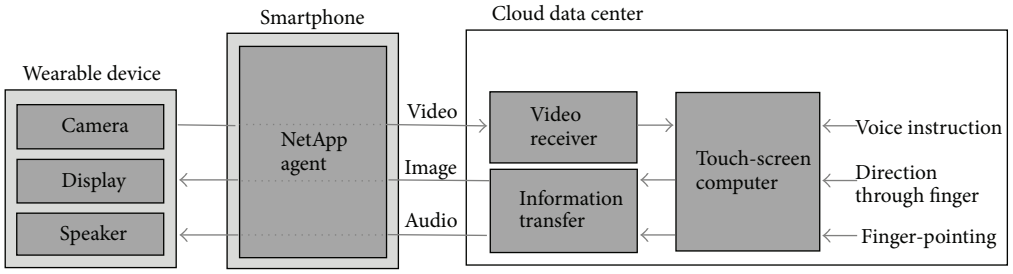


Figure 10: Voice and video-sharing scenario involving smart device and cloud server.

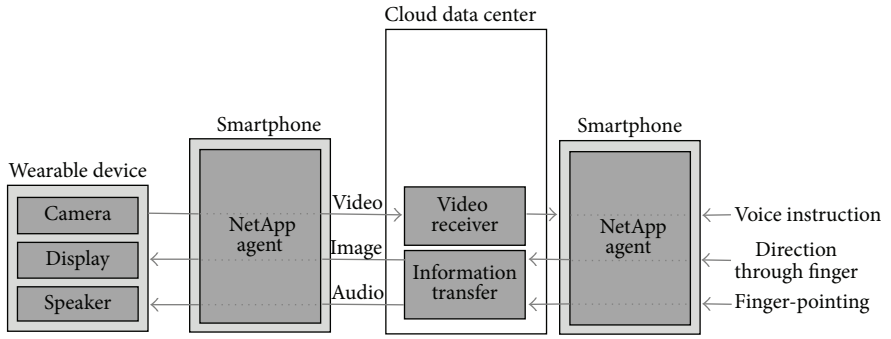


Figure 11: Voice and video data-sharing scenario involving smart devices.

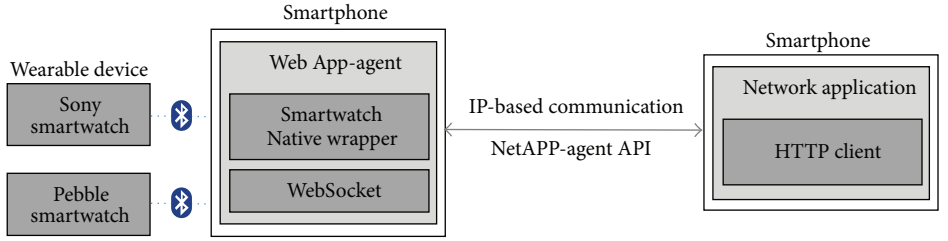


Figure 12: Implemented NetApp-Agent.

server. A user wearing Google Glass transmits information in real time to the cloud server, which shares the transferred video with the user again and issues an order. An order can be issued by pointing at the transferred video. The cloud server once again utilizes the NetApp-Agent API to implement an application to display information on the wearable device. Thus, modification of the wearable device is unnecessary.

3.2.3. Voice and Video Data Sharing among Smart Devices. Figure 11 shows a more detailed scenario than the previous one. Two smart devices capable of IP-based communication share video, voice, and image through a camera in real time with the cloud server acting as mediator. The realization of this scenario can lead to innovative services in the market. For instance, suppose User A shares his/her location with User B through an image as User B guides User A through vocal instructions or ostensive guides using the shared image.

4. Experiment and Qualitative Assessment

In this section, we report an experiment to show the feasibility of the suggested platform and discuss the results. The efficiency of the proposed platform is also confirmed through a comparative analysis with the existing development platforms.

4.1. Empirical Research of the Proposed Platform

4.1.1. NetApp-Agent Implementation by Utilizing Native Development Method. The proposed platform was implemented in order to demonstrate how proposal works as the prototype. The setup consisted of a smartphone equipped with NetApp-Agent, another smartphone with a network application communicating with NetApp-Agent, and controllable wearable devices, as shown in Figure 12. NetApp-Agent was implemented in Android 4.1.2, and a web server function was added to support Hypertext Transfer Protocol (HTTP) communication based on IP. The two wearable devices used were a Sony SmartWatch and a Pebble SmartWatch. Sony Add-on SDK 2.1 and Pebble SDK 2.0 were installed to develop the SmartWatch application. The user application was implemented in Android 4.1.0 version and the HTTPClient class in Android was used to support HTTP communication. Both NetApp-Agent and the user application exchanged JSON data using a self-defined NetApp-Agent API and sending HTTP communication based on IP.

Table 7 lists the NetApp-Agent APIs. NetApp-Agent enabled the transmission of texts and images to wearable devices. Accordingly, the four former APIs were used, Connected Device List (which offered the list of connected wearable devices and detailed information regarding each device), Select Device (which selected the messaging terminals), Send Text (to send texts), and Send Image (to send images to devices). The four latter APIs will soon be added to NetApp-Agent APIs.

The detailed working structure of the implemented platform is shown through a sequence diagram. Figure 13 shows the process of acquiring information related to the detailed specifications and connections of wearable devices using the network application. The network applications transmit an HTTP GET method to set connections with NetApp-Agent. Upon receiving requests, the NetApp-Agent sends messages containing information regarding connected wearable devices and detailed specifications of each device to the network application, which follows a JSON type. On receiving

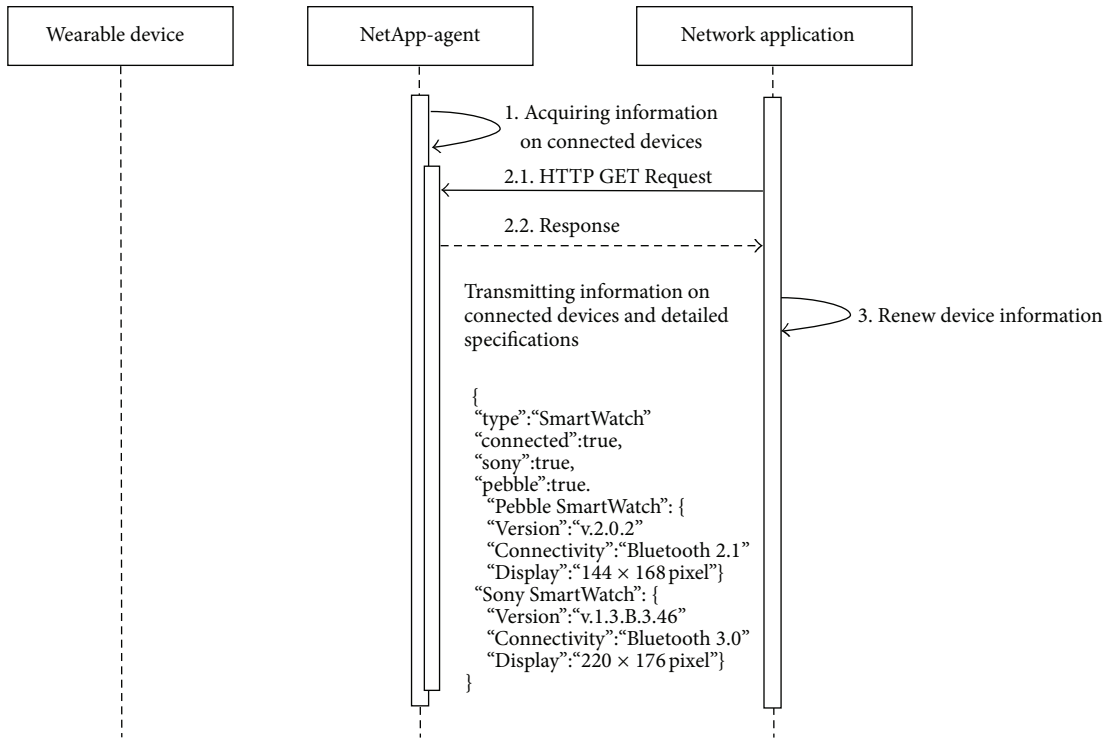


Figure 13: Acquiring information on wearable devices using network application.

Table 7: List of NetApp-Agent APIs.

API	Description
Connected Device List	Provide detailed information on each device and the list of wearable devices connected to smartphone through Bluetooth
Select Device	Select the kinds of wearable devices to control
Send Text	Make the text appear on the screen by sending it to wearable devices
Send Image	Make the image appear on the screen by sending it to wearable devices
Send Voice	Make the voice and sound appear at speaker by sending it to wearable devices
Cellular	Monitor cellular state and make cellular state enabled or disabled
Wi-Fi	Monitor Wi-Fi state and make Wi-Fi enabled or disabled
Bluetooth	Monitor Bluetooth state and make Bluetooth enabled or disabled

information about wearable devices, the network application adds the device information to the List View so that users can select devices after checking related information.

Figure 14 shows the process of sending texts to wearable devices via NetApp-Agent, and Figure 15 shows the implementation of the platform. Information on devices for submission and text values selected by users are transmitted in JSON messages through HTTP POST. On receiving the messages, NetApp-Agent sends texts to the selected wearable devices and prints it on the screen. Depending on the context, NetApp-Agent supports intelligent tasks.

In case of the Pebble SmartWatch, it used to show the issue of incompatibility with Korean language. Thus, if text has to be transmitted through Pebble SmartWatch, NetApp-Agent senses the Korean characters in the messages before sending messages to wearable devices. While messages containing no Korean characters are transmitted as text, those containing Korean characters are transformed into PBI images and then transmitted to Pebble SmartWatch and vice versa.

Figures 16 and 17 show the process of sending images to wearable devices via NetApp-Agent and the results obtained on the screen. As in the case of sending texts, device information regarding the data and the images is sent to NetApp-Agent using HTTP POST. The images may need to be converted depending on the types of targeted devices when sending images. Sony SmartWatch displays images by temporarily saving them on the smartphone and sending uniform resource identifier (URI) information. By utilizing its self-defined image format, the Pebble SmartWatch converts the received bitmap images to PBI format and then sends it.

4.1.2. NetApp-Agent Implementation Utilizing Cordova Platform. NetApp-Agent aspires to be a cross-platform development framework that implements applications regardless of the type of mobile platform at hand and expands the coverage of API support by adding Native API functions to the Cordova API.

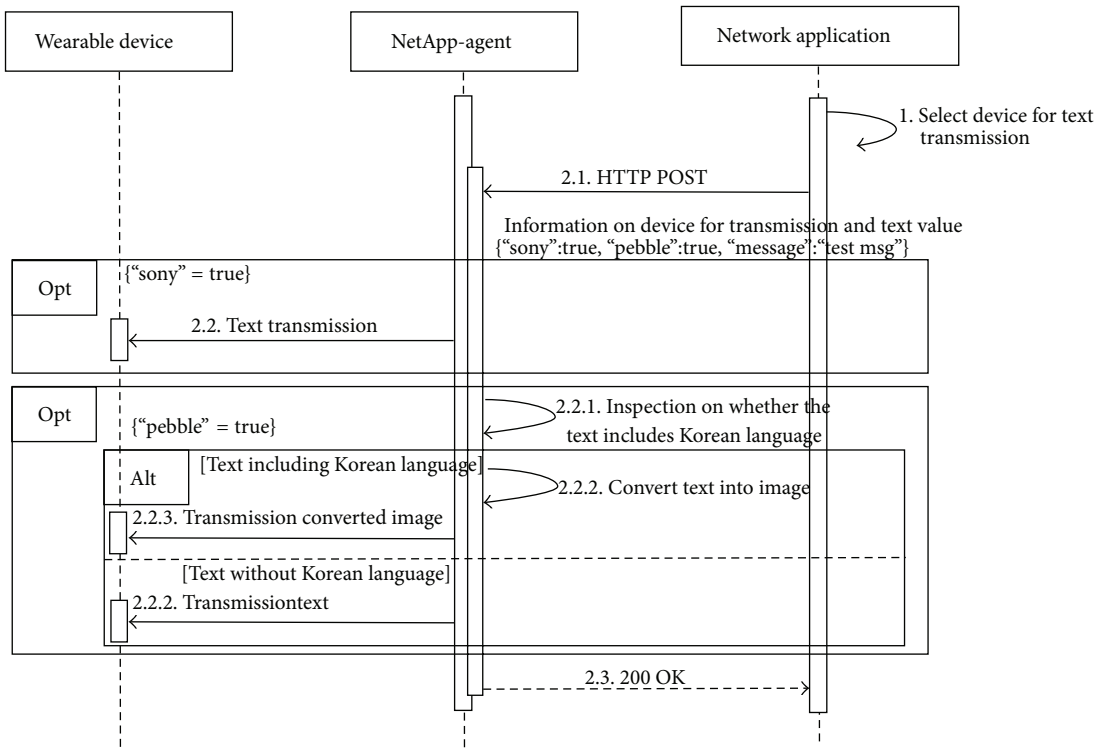


Figure 14: Sending text via NetApp-Agent.

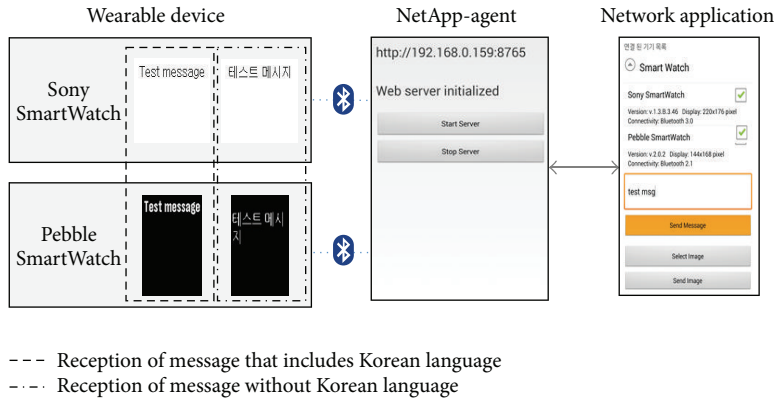


Figure 15: Result of sending text via NetApp-Agent.

Cordova supports the acquisition of the location information of devices and an API that provides status information regarding cellular connected to devices or to Wi-Fi networks. Figure 18 shows sequence of Acquiring GPS and network information. Consequently, when the location information of a smartphone and a network is requested to NetApp-Agent by network application, NetApp- Agent responds to a call in a way of acquiring information by calling Cordova API.

4.2. Qualitative Assessment of the Proposed Method. In this subsection, we conduct a functional analysis of the existing development platforms as well as our proposed NetApp-Agent and use this qualitative assessment to identify the pros and cons of the proposed platform. The efficiency of the proposed platform is tested throughout this process.

4.2.1. Provision of Integrated Development Environment for Devices. The current mobile market has several mobile platforms, such as Android, iOS, and Windows Mobile. Owing to the diversity of platforms, developers need to construct development environments according to platform and learn several programming languages, which slows down the pace of development. Mobile programming is gradually becoming standardized to cope with such issues, and cross-platforms such as Cordova and Titanium are welcome as a result. Nevertheless, cross-platforms are limited in their scope of

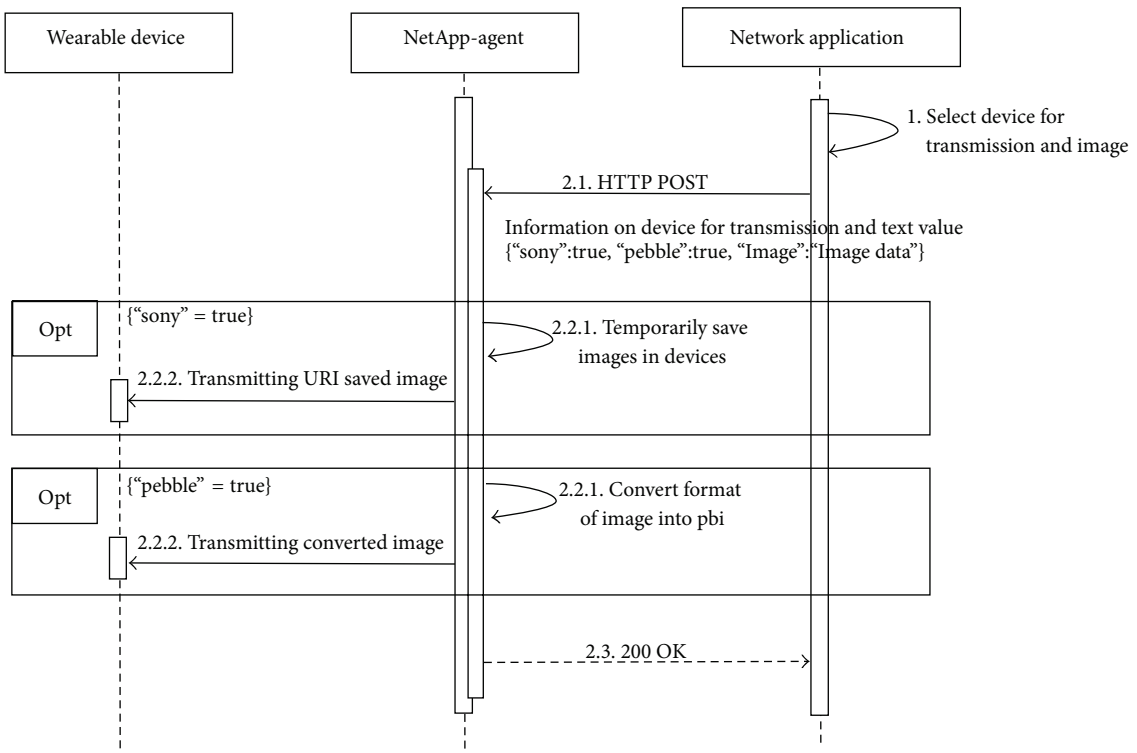


Figure 16: Sending image via NetApp-Agent.

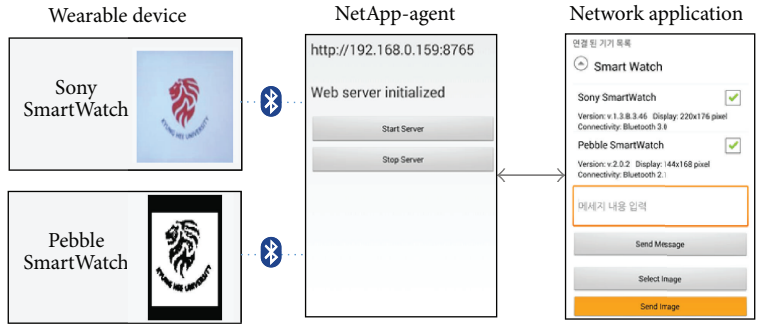


Figure 17: Result of sending image via NetApp-Agent.

API coverage and are still in the process of developing APIs to control wearable devices. All of this underscores the necessity of the acquisition and development of native languages. In other words, cross-platform by itself is inadequate for developing applications for wearable devices. A comprehension and development of applications that take into account unfamiliar development environments and API assistance for wearable devices would be extremely time-consuming.

Table 8 shows the development environment of the Cordova development platform. When developing wearable devices outside Cordova's stipulated coverage, the native development approach is preferred. Hence, the development of applications necessitates the construction of development environments depending on target devices as well as knowledge of compatible APIs. In the case of applications for Google Glass, for instance, developers need to learn both the development language, such as Python and Java, and the Google Mirror API after installing the relevant development tools.

Developing applications for Pebble SmartWatch requires that developers construct suitable development environments and learn specific languages such as Pebble SDK and PebbleKit SDK, which are built-in functions of a smartphone. Even when developing different applications for the same purpose, a considerable amount of time needs to be spent on creating different development environments and learning unfamiliar languages. The table shows development environments for four different devices, where developers had to learn 10 programming languages and 7 APIs. As the number of devices supplying applications increases, the workload and time spent on development increase proportionately.

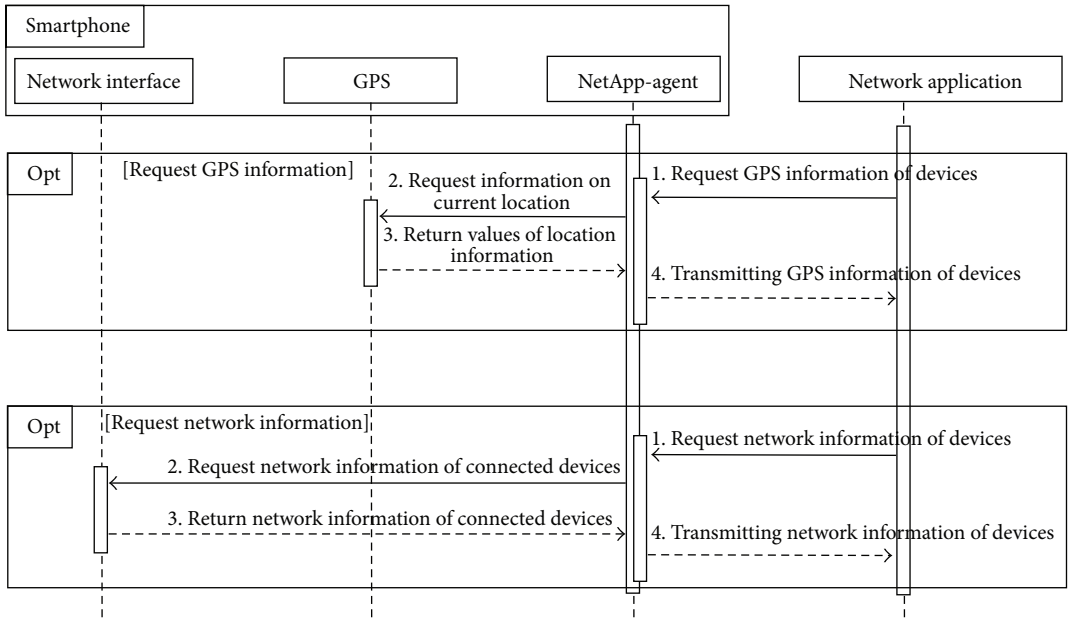


Figure 18: Acquiring GPS/network information of a device via Cordova API.

Table 8: Development environment of Cordova platform.

	Under Cordova platform		
Types of devices	Operating system	Supporting languages	SDK/API
Total	**4**	**10**	**7**
Google Glass	Android 4.0.4	Go, Java, .NET, PHP, Python, and Ruby	Google Mirror API
Sony SmartWatch	Android 4.0	Java	Sony Add-on SDK/Smart Extension API
Pebble SmartWatch	Pebble OS	C, JavaScript, and so forth (Objective-C/Java)	Pebble SDK PebbleKit
Samsung Gear	Tizen OS	Java, HTML, JavaScript, and so forth	Tizen SDK Samsung Mobile SDK

Table 9: Development environment of NetApp-Agent.

	Under NetApp-Agent		
Types of device	Operating system	Supporting languages	SDK/API
Total	**1**	**1**	**1**
NetApp-Agent supporting devices (extendable)	No limitations	No limitations	NetApp-Agent API (WebSocket/JSON-based)

Table 9 shows the application development environment based on NetApp-Agent. If developers are aware of the specific features of JSON-based NetApp-Agent APIs and the implementation of WebSocket for communication with a platform, other applications can be easily developed using programming languages with which developers are, presumably, already familiar. Thus, procedural redundancies, such as constructing a separate development environment for each platform and learning both the relevant languages and the API, can be effectively avoided. This approach can downsize the time and cost needed to develop applications, and developers can easily expand the selection of devices compatible with the applications.

4.2.2. Expansion of API Supply Coverage of Platform. Cross-platform development frameworks such as Titanium and Cordova are still in the process of development and hitherto have failed to guarantee wide API coverage. The proposed platform widens the coverage of serviceable APIs by adding a native device wrapper to the Cordova wrapper. Table 10 lists the kinds of APIs per development platform. We see that NetApp-Agent supports wider API coverage than the

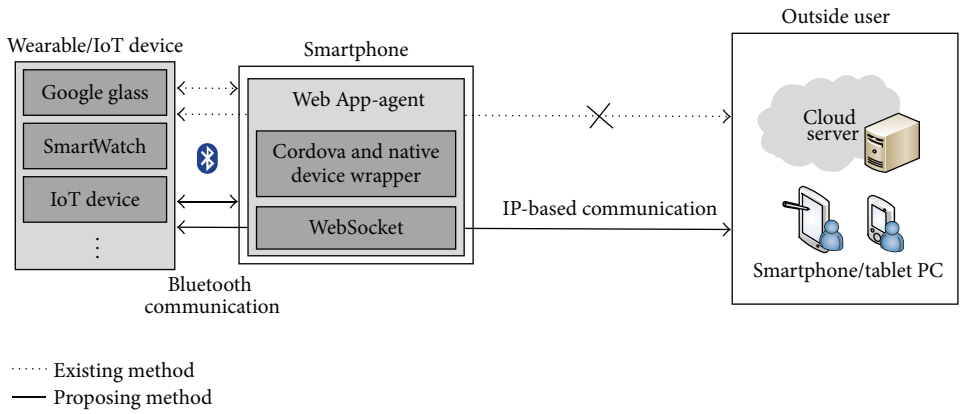


Figure 19: Comparison of communication methods between existing and proposed platforms.

Table 10: Supporting APIs per development platform.

API		Titanium	Cordova	NetApp-Agent
Battery		O	O	O
Vibration		X	O	O
Camera		O	O	O
GPS		X	O	O
Network	Cell_State	X	O	O
	Cell_ON	X	X	O
	Cell_OFF	X	X	O
	WiFi_State	X	O	O
	WiFi_ON	X	X	O
	WiFi_OFF	X	X	O
	Bluetooth_State	X	X	O
	Bluetooth_ON	X	X	O
	Bluetooth_OFF	X	X	O
	Device List	X	X	O
Media	Send Audio	X	X	O
	Recv Audio	X	X	O
	Play Audio	O	O	O
	Send Video	X	X	O
	Recv Video	X	X	O
	Play Video	O	X	O
Display	Send Text	X	X	O
	Recv Text	X	X	O
	Send Image	X	X	O
	Recv Image	X	X	O
Sensor	Compass	X	O	O
	Accelerometer	X	O	O
	Light Sensor	X	X	O

existing platforms. Consequently, the use of NetApp-Agent API allows in-depth control of hardware/software modules as well as communication without further development of the native language.

4.2.3. IP-Based Communication Service for Bluetooth Communication Devices. It is rare to find stand-alone functions in wearable devices as most functions are usually dependent on smartphones. Connections between smartphones and wearable devices are needed to guarantee the practical use of devices, and Bluetooth-based interdevice communication is utilized at this stage. Bluetooth as a communication system for short distances is implemented by referring to the Mac addresses of the paired devices. Accordingly, it is virtually limited for a wearable device that only supports Bluetooth communication to communicate with external devices or users as it is the same issue of IoT devices.

Our proposed platform forms Bluetooth connections with wearable/IoT devices and supports WebSocket-based IP communication. It creates a network by grouping together Bluetooth-based devices and enables IP communication with the outside. Thus, outside users can communicate with and control wearable/IoT devices using IP without having to form Bluetooth connections.

Figure 19 shows the differences between the existing platform and the proposed one. In case of the existing method, wearable/IoT devices only support Bluetooth communication. To use and control devices, smartphones connected to them need to be carried to within the coverage range of Bluetooth communication. The proposed platform supports remote access to devices without IP. Accordingly, devices can be controlled from outside using IP communication.

4.2.4. Provision of Network Application Development Environment. SDKs and APIs being offered to implement cloud services mostly focus on building the cloud server. Consequently, client applications need independent implementation for each platform in order to service devices. The proposed NetApp-Agent platform supports the development of network applications for cloud services.

Network applications carry out IP-based communication with NetApp-Agent and use NetApp-Agent APIs to provide services using cloud servers to users without the need for further implementation or revision of smart devices. Considering that smart devices are only implemented through the API offered, the revision of smart device programs is unnecessary even if the function of the native application is extended or a brand-new network application is developed.

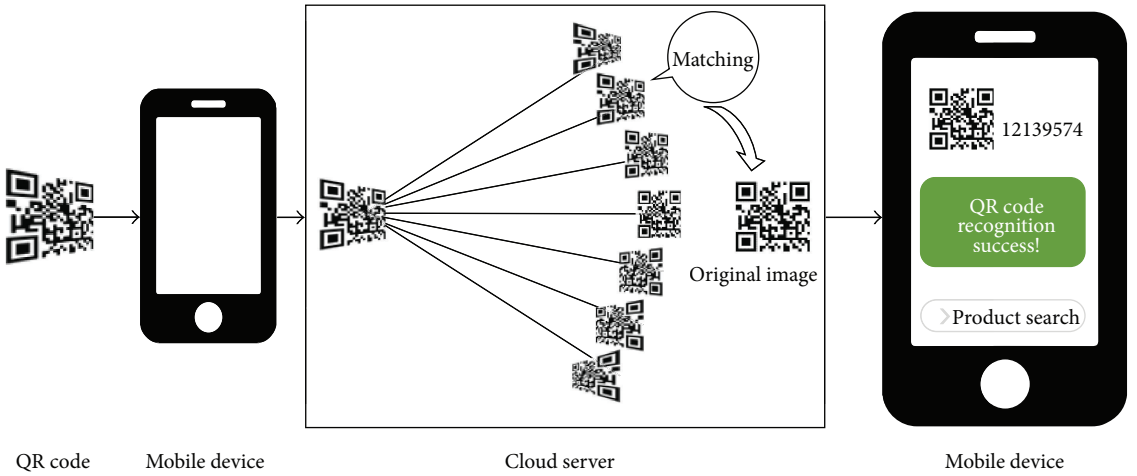


Figure 20: Operating process of the pregenerated Image Matching Method.

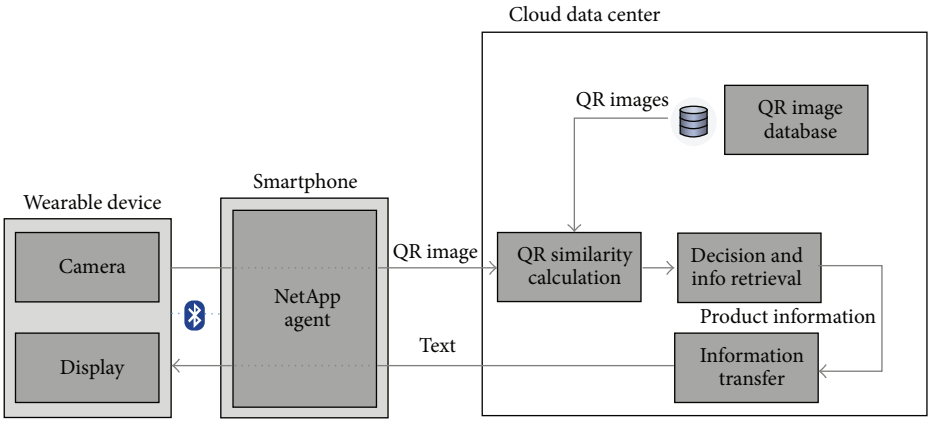


Figure 21: Expandable Cross-Platform for QR code recognizing scenario based NetApp-Agent.

Furthermore, smart devices can support intelligent functions by connecting to the cloud server, which was hitherto unavailable for smart devices. This enables the provision of advanced services to users as a result.

4.3. Experimental Implementation of the Proposed Method

4.3.1. Expandable Cross-Platform for QR Code Recognition Using the Pregenerated Image Matching Method. QR code is two-dimensional code developed in 1994, and the amount of a QR code is rapidly increasing [33]. It can handle more information than barcode about several hundreds of times and anyone can make and use it easily. Also, it enables the user to access website through recognizing code without entering URL [34]. The important key of QR code recognition is caught "Finder Pattern" of QR code. QR code can be recognized only if camera catches these three patterns. When angle bends largely, camera cannot catch three patterns so that it becomes impossible to recognize QR code [35].

The QR code recognition method is "pregenerated image matching" that finds original QR code through similarity test of shooting photograph and database's images in the server. We suggest the composition of server-device environment and operational process.

Figure 20 shows the entire operating process of QR code recognition method. At this method, wireless device sends shooting QR code image without any processing to the cloud server. After receiving QR code image, cloud server performs whole operating process. It is a contrast to traditional technology where shooting device performs whole operating process.

Before the recognition, cloud server must store QR code images in the database. This process includes transform QR code image in four directions with every single angle.

Once wireless device takes a picture of QR code image, device sends image to server. Server performs similarity test among input image and images stored into the database. Server finds an original image with the largest similarity coefficient. Finally, it sends detecting QR code image and decoding information to the wireless device.

Figures 21 and 22 show the process of QR code recognition based NetApp-Agent. We create a NetApp-Agent application using Cordova wrapper, which supports "pregenerated image matching" on various devices. A device made of web

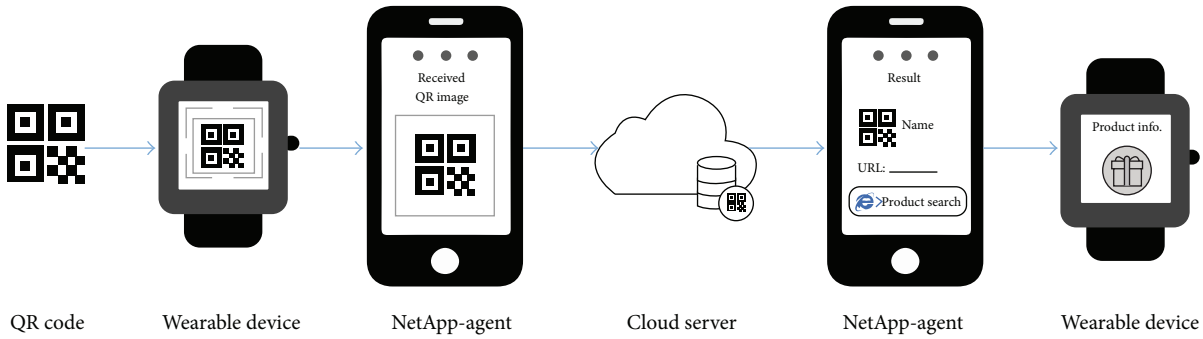


Figure 22: The process of QR code recognition using the pregenerated Image Matching Method based NetApp-Agent.

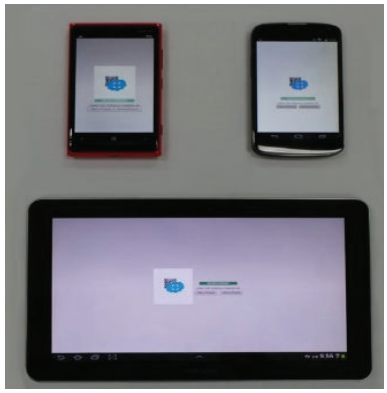


Figure 23: Operating screen of the application.

Table 11: Recognition angle using NAVER application.

Distance	Model			
	Galaxy Note 3	Galaxy S4	Nexus 4	iPhone 5
15 cm	35°	40°	45°	0°
30 cm	45°	43°	50°	X

Table 12: Recognition angle using SCANY application.

Distance	Model			
	Galaxy Note 3	Galaxy S4	Nexus 4	iPhone 5
15 cm	50°	45°	55°	45°
30 cm	55°	55°	55°	45°

application communicates with the server made of Node.js using WebSocket. Because different operating system can use the same web application, the service can be provided to various devices with separate operating system [36].

We used Google's reference phone, which is Nexus 4 installed Android 4.4 Kitkat. In addition, we had a test in Lumia 920 installed Windows Phone 8.0 and Samsung Galaxy Tab 10.1 installed Android 4.0 Ice Cream Sandwich. Figure 23 shows that application of the QR code recognition is portable to other operating systems.

4.3.2. Performance Evaluation of the Pregenerated Image Matching Method. This paper's experiment is verifying the recognition of QR code in the angle where it cannot be recognized in traditional QR code applications.

First, we implement an experiment to find the maximum recognition angle of traditional applications. We measure the maximum recognition angle with four smartphones using NAVER and SCANY application. Tables 11 and 12 are the test result of using 1.8 cm ∗ 1.8 cm QR code.

Those two applications display the maximum recognition angle at 55° so we experiment the possibility whether our proposed method can recognize the image at 60°.

Figure 24 is 1.8 cm ∗ 1.8 cm 10 QR code images for recognition. These images of the database were taken at a distance of 20 cm by using the phone camera directly. Setting the maximum angle of the QR code to 70 degrees, the images were taken 5 degrees from 0 degrees by incrementing. The images are stored in binary. We saved 15 images per QR code, and the 150 images were set as the comparison of the experiment. We examined by comparing the similarity between the 150 QR code images and we saved the 10 QR code images inclined at 60 degrees. After checking the similarity, the similarity of each QR code was output in descending rank from first to third rankings.

Figure 25 is a graph using a QR code 01 taken at 60-degree angles, representing the similarity of the 150 QR code images. It is expressed by a line of a different color for each QR code, the horizontal axis represents the value of the angle, and the vertical axis represents the degree of similarity (%).

The most significant similarity is the image at 60 degrees of QR code 01, because the QR code image has high similarity with itself. Other QR codes can be matched with target of experiment, having relatively low similarity. These experimental results show that the proposed method is superior to the traditional method, which is used by NAVER and SCANY application; this paper presents recognition of QR code over other applications' maximum limit angle.

5. Conclusion and Future Research

In this paper, we proposed a wearable device control platform to develop network applications. We also carried out an experiment to confirm the feasibility and efficiency of the proposed platform. The advantages of the proposed platform

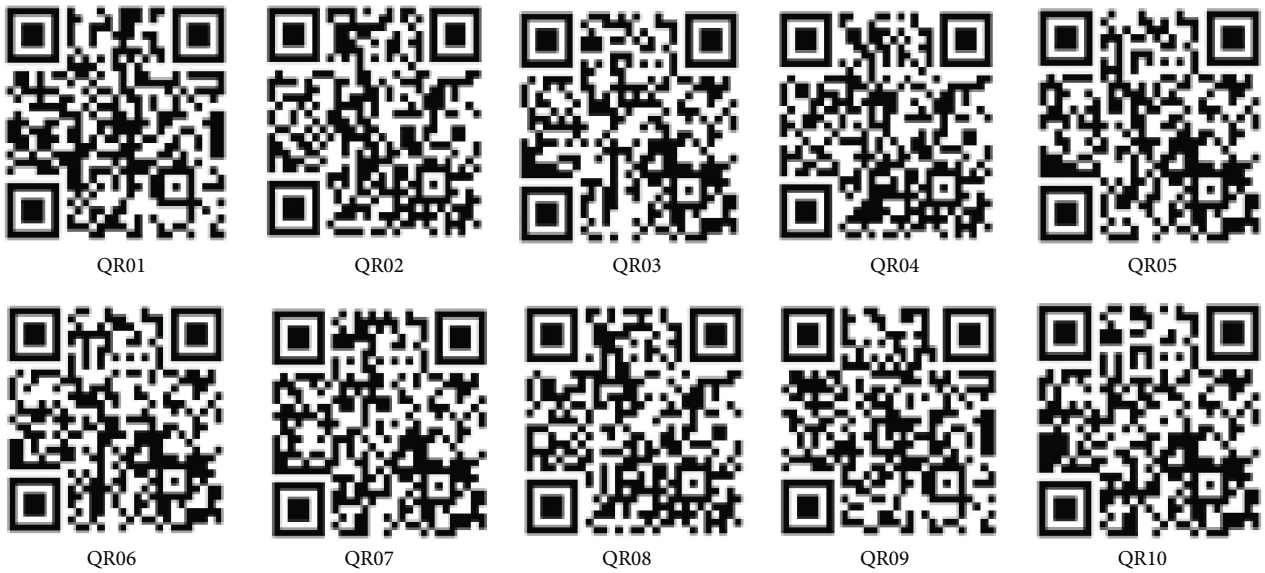


Figure 24: 10 QR code images for experiment.

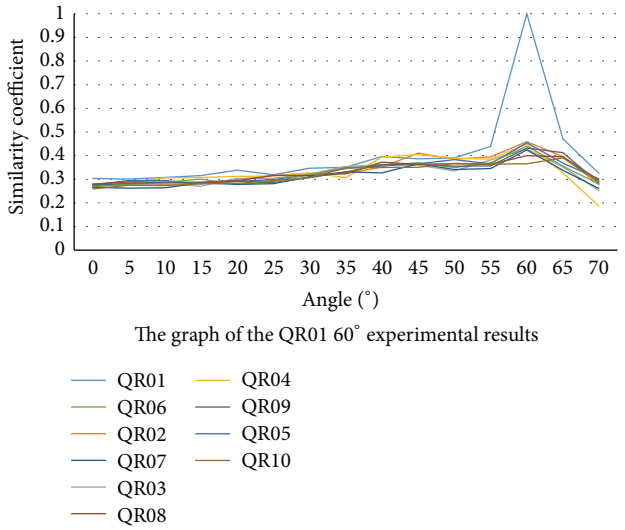


Figure 25: The graph of the experimental results.

were tested through a comparative analysis with the existing platform.

There are two main advantages of our proposed platform. First, it reduces the time and cost needed to develop applications by providing a single API to developers. In past development approaches, developers needed to construct development environments for each platform and had to learn several programming languages and APIs when developing applications. For the development of applications working on a Sony SmartWatch, an Android-based development environment had to be constructed, and, in order to install the Sony Add-on SDK, learning about Java and API had to be gone through. Furthermore, the development of applications with the same features for Pebble SmartWatch required that developers acquire relevant knowledge of the programming languages, such as C and Pebble SDK. The increase in the number of platforms led to a rise in the time and resources required to develop applications, which are falling behind the current trends in which a variety of wearable devices are consistently launched. By integrating development environments and offering a JSON-based API, our proposed platform eases the burden of application development regardless of environment and type of platform. Moreover, the issue of limited API coverage in the Cordova platform was addressed by using a Cordova device wrapper and a native wrapper together. The proposed platform enables the development of a wide variety of applications by extending API coverage to software/hardware module control of devices.

Second, wearable devices that support Bluetooth communication can communicate with the external environment in our platform using IP. Existing wearable devices support Bluetooth communication and are heavily dependent on smartphones. However, in the existing platform, a smartphone needs to be carried to control and utilize wearable devices, which need to be located within the range of Bluetooth coverage. The proposed platform allows the control of devices through remote access as well as the exchange of relevant information. Due to rising demand for IoT, the era of wider communication even with blub, TV, remote control, and vehicles is coming. The application of the proposed platform to IoT will make it possible to form an interdevice network where outside users can control devices.

At this point, in technological development, when the need for a standardized mobile programming method is pressing on account of the growing number of smart devices on a variety of platforms, our proposed platform shows advantages by unveiling its unique features which are incorporating the development methods of applications and supporting IP-based communications to external devices. Thus, since our platform facilitates the development of service applications without additional implementation or revision,

the widespread adoption of this approach is expected in developing network applications that service smart devices.

An issue to consider is that consistent attention and updates to the platform will be required for new smart devices and platforms in order to guarantee the continual use and development of NetApp-Agent. The experiment described here shows the feasibility of NetApp-Agent, which adds some limitations on the coverage of serviceable API. Hence, the proposed NetApp-Agent is in need of further implementations.

Conflict of Interests

The researchers claim no conflict of interests.

Acknowledgments

This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Republic of Korea, under the ITRC (Information Technology Research Center) support program (IITP-2015-(H8501-15-1015)) supervised by the IITP (Institute for Information & Communications Technology Promotion); the Institute for Information & Communications Technology Promotion (IITP) Grant funded by the Korea government (MSIP) (B0190-15-2013, Development of Access Technology Agnostic Next-Generation Networking Technology for Wired-Wireless Converged Networks); and the ICT R&D Program of MSIP/IITP, Republic of Korea (B0101-15-1366, Development of Core Technology for Autonomous Network Control and Management).

References

[1] M. Ballve, *Wearable Computing: From Fitness Bands to Smart Eyewear. A New Mobile Market Takes Shape*, Business Insider, 2013.

[2] K. Daegun, "Trends and implications of wearable device," *Policy of Broadcasting and Telecommunication*, vol. 25, no. 21, 2013.

[3] Korea Association for ICT Promotion, *A Monthly ICT Statistics*, Monthly ICT Item Trend Investigation, ICT Business Survey Index (BSI), 2013.

[4] Machina Research, *Strategy Report—M2M Communication Service Provider Benchmarking*, 2012.

[5] Gartner, *Forecast: Mobile App Stores, Worldwide, 2013 Update*, Gartner, 2013.

[6] A. Zibula and T. A. Majchrzak, "Cross-platform development using HTML5, jQuery mobile, and phonegap: realizing a smart meter application," in *Web Information Systems and Technologies*, vol. 140 of *Lecture Notes in Business Information Processing*, pp. 16–33, Springer, Berlin, Germany, 2013.

[7] Titanium, http://www.appcelerator.com/titanium/.

[8] B. Zhang, T.-G. Xu, W. Wang, and X. Jia, "Research and implementation of cross-platform development of mobile widget," in *Proceedings of the 3rd International Conference on Communication Software and Networks (ICCSN '11)*, pp. 146–150, IEEE, Xi'an, China, May 2011.

[9] S. Xanthopoulos and S. Xinogalos, "A comparative analysis of cross-platform development approaches for mobile applications," in *Proceedings of the 6th Balkan Conference in Informatics (BCI '13)*, pp. 213–220, Thessaloniki, Greece, September 2013.

[10] N. Fernando, S. W. Loke, and W. Rahayu, "Mobile cloud computing: a survey," *Future Generation Computer Systems*, vol. 29, no. 1, pp. 84–106, 2013.

[11] J. Donggeun, *Web Application Hybrid Application Programming*, Answerbook, 2013.

[12] C. Jaekyu, *PhoneGap for Development of Hybrid Mobile Application*, WannaBooks, 2012.

[13] IBM, *Native, Web or Hybrid Mobile-App Development*, Software Thought Leadership White Paper, 2012.

[14] Sony SmartWatch Developer World web site, 2013, http://developer.sonymobile.com/tag/smartwatch/.

[15] K. Sanghyung, *Android Programming Complete Guide*, Hanbit Media, 2010.

[16] Xcode web site, https://developer.apple.com/xcode/.

[17] Pebble, http://developer.getpebble.com.

[18] L. Tian, H. Du, L. Tang, and Y. Xu, "The discussion of cross-platform mobile application based on Phonegap," in *Proceedings of the 4th IEEE International Conference on Software Engineering and Service Science (ICSESS '13)*, pp. 652–655, IEEE, Beijing, China, May 2013.

[19] J. M. Wargo, *Phonegap Programming*, Acorn, 2013.

[20] B. Pollentine, *Titanium Mobile Application Programming*, Acorn, 2012.

[21] H. Heitkötter, S. Hanschke, and T. A. Majchrzak, "Evaluating cross-platform development approaches for mobile applications," in *Web Information Systems and Technologies: 8th International Conference, WEBIST 2012, Porto, Portugal, April 18–21, 2012, Revised Selected Papers*, vol. 140 of *Lecture Notes in Business Information Processing*, pp. 120–138, Springer, Berlin, Germany, 2013.

[22] J. Daeyoung and K. Jongki, *Creative Convergence Visualization Method of IoT*, KIET Industrial Research, 2014.

[23] M. Younis, I. F. Senturk, K. Akkaya, S. Lee, and F. Senel, "Topology management techniques for tolerating node failures in wireless sensor networks: a survey," *Computer Networks*, vol. 58, no. 1, pp. 254–283, 2014.

[24] K. Curran, A. Millar, and C. Mc Garvey, "Near Field Communication," *International Journal of Electrical and Computer Engineering*, vol. 2, no. 3, pp. 371–382, 2012.

[25] A. Jula, E. Sundararajan, and Z. Othman, "Cloud computing service composition: a systematic literature review," *Expert Systems with Applications*, vol. 41, no. 8, pp. 3809–3824, 2014.

[26] A. Marathe, R. Harris, D. K. Lowenthal, B. R. de Supinski, B. Rountree, and M. Schulz, "Exploiting redundancy for cost-effective, time-constrained execution of HPC applications on amazon EC2," in *Proceedings of the 23rd International Symposium on High-Performance Parallel and Distributed Computing (HPDC '14)*, pp. 279–290, ACM, Vancouver, Canada, June 2014.

[27] Z. Sanaei, S. Abolfazli, A. Gani, and R. Buyya, "Heterogeneity in mobile cloud computing: taxonomy and open challenges," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 369–392, 2014.

[28] A. Bedra, "Getting started with Google app engine and Clojure," *IEEE Internet Computing*, vol. 14, no. 4, pp. 85–88, 2010.

[29] DNA Developer Network web site, http://developers.daum.net/services.

[30] C. Jaekyu, *Understanding of Cloud Technology and Development Platform*, MicroSoftware Article, 2012.

[31] Amazon Webservice SDK, http://aws.amazon.com/ko/tools/.

[32] Q. Zhang, L. Cheng, and R. Boutaba, "Cloud computing: state-of-the-art and research challenges," *Journal of Internet Services and Applications*, vol. 1, no. 1, pp. 7–18, 2010.

[33] Y. Liu, J. Yang, and M. Liu, "Recognition of QR Code with mobile phones," in *Proceedings of the Chinese Control and Decision Conference (CCDC '08)*, pp. 203–206, IEEE, Yantai, China, July 2008.

[34] L. F. F. Belussi and N. S. T. Hirata, "Fast QR code detection in arbitrarily acquired images," in *Proceedings of the 24th SIB-GRAPI Conference on Graphics, Patterns and Images (Sibgrapi '01)*, pp. 281–288, Maceió, Brazil, August 2011.

[35] M. Ahn and S. Lee, "A research on QR Code recognition enhancement using pre-constructed image matching scheme," in *Proceedings of the International Conference on Information and Communication Technology Convergence (ICTC '14)*, pp. 82–83, IEEE, Busan, South Korea, October 2014.

[36] M. Ahn, S. Hong, and S. Lee, "A research on the QR Code recognition improvement using the cloud-based pre-generated image matching scheme," in *Proceedings of the International Conference on Information Networking (ICOIN '15)*, pp. 356–357, IEEE, January 2015.

Hindawi Publishing Corporation
Mobile Information Systems
Volume 2016, Article ID 1340973, 10 pages
http://dx.doi.org/10.1155/2016/1340973

Research Article

Mining Sequential Update Summarization with Hierarchical Text Analysis

Chunyun Zhang,[1] Zhongwei Si,[2] Zhanyu Ma,[2] Xiaoming Xi,[1] and Yilong Yin[3]

[1]*School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan, Shandong 250014, China*
[2]*School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China*
[3]*School of Computer Science and Technology, Shandong University, Jinan, Shandong 250014, China*

Correspondence should be addressed to Zhanyu Ma; mazhanyu@bupt.edu.cn

Received 1 October 2015; Revised 21 December 2015; Accepted 5 January 2016

Academic Editor: Yassine Hadjadj-Aoul

The outbreak of unexpected news events such as large human accident or natural disaster brings about a new information access problem where traditional approaches fail. Mostly, news of these events shows characteristics that are early sparse and later redundant. Hence, it is very important to get updates and provide individuals with timely and important information of these incidents during their development, especially when being applied in wireless and mobile Internet of Things (IoT). In this paper, we define the problem of sequential update summarization extraction and present a new hierarchical update mining system which can broadcast with useful, new, and timely sentence-length updates about a developing event. The new system proposes a novel method, which incorporates techniques from topic-level and sentence-level summarization. To evaluate the performance of the proposed system, we apply it to the task of sequential update summarization of temporal summarization (TS) track at Text Retrieval Conference (TREC) 2013 to compute four measurements of the update mining system: the expected gain, expected latency gain, comprehensiveness, and latency comprehensiveness. Experimental results show that our proposed method has good performance.

1. Introduction

Internet of Things (IoT) is a new type of the Internet. It is the network of physical objects or "things" embedded with electronics, software, sensors, and network connectivity, which enables these objects to collect and exchange data [1]. Many high-tech companies over the world have already started developing IoT products and services and promoting their early stage of IoT products and services in a number of market domains. Among the most notable challenges, wireless and mobile technologies are the underlying technologies for realizing the IoT [2, 3]. Resource constrained devices are required to communicate with other devices in wireless networks. The devices are also required to communicate on the move. In addition to these requirements, various technical and scientific research considerations are also required. One of the key techniques is to develop semantic and intelligent web for IoT [4]. The core of this technique is the combination of the traditional Internet technologies and the wireless and mobile technologies. For example, when an unexpected news event occurs, such as natural disaster (e.g., earthquake) or human accidents (e.g., air crash), some event data can be collected by IoT devices, and they submit these event data to the Internet. And these data in the Internet will form some real-time news. Based on effective sequential update summarization system, the IoT system can send individuals useful, new, and timely updates by mobile devices. Hence, developing effective sequential update summarization techniques is very important for the IoT.

However, due to the special characteristic of unexpected news event, it is a big challenge to construct an effective sequential update summarization system. Mostly, the information about unexpected news events is rapidly developing [5]. For instance, immediately after the outbreak of an unexpected event, the corpus may be sparsely populated with relevant news. Even when, after a few hours, relevant

news is available, it is often inaccurate or highly redundant. That is because news of the event is widely spread through multilevel news channels around the world. However, based on the diversity of journalistic sources, details reported about the event are redundant, dynamic, and sometimes mistaken. Furthermore, it becomes much harder to gather authoritative news, when facing major events which involve extensive damage to life or crippling of infrastructure. This may cause rumors and unsubstantiated information to propagate [6]. Meanwhile, the sudden events are also very important topics to individuals. People want to get timely information, especially for these people who are relative to these sudden events; they even cannot afford waiting for comprehensive reports to materialize [7].

Unfortunately, existing solutions cannot satisfy people's demands in getting useful, new, and timely sequential update summarizations about these events. That is because the problem of sequential update summarization extraction refers to techniques intercrossed among text summarization, topic detection and tracking, and time-based summarization. However, most current summarization systems can either use static summarization methods [8–13] or use topic detection and tracking (TDT) methods [14–18]. These methods only provide sentences extracted with particular properties based on traditional techniques of natural language processing (NLP) [19] or only provide topic-level summaries. In most ways, the sequential update summarization is an event- and sentence-level analogue of "first topic detection" problem [20]. In all, there is no support for only presenting people with novel content (i.e., updates to the user) and updates can suffer from poor coverage and unreliable information.

In this paper, we define the problem of sequential update summarization extraction for unexpected news events. This task can be considered as a variation of topic detection and tracking and time-based document summarization. Hence, the problem definition, the evaluation, and the used method are based on these techniques. With significant extension of the abovementioned techniques, we present a new hierarchical summarization system, which focuses on extracting sequential update summarization on unexpected news events. The system tries to broadcast with useful, new, and timely sentence-length updates about a developing event by incorporating the technologies of time-based topic-level and sentence-level summarization. With the application to the sequential update summarization (SUS) task of temporal summarization (TS) track [21] at Text Retrieval Conference (TREC) [22], we evaluated the effectiveness of our new method in view of precision, recall, timeliness, and novelty of updates. By computing the expected gain, expected latency gain, comprehensiveness, and latency comprehensiveness (evaluation metric of SUS task) of our extracted updates of 10 topics, we conclude that our proposed method has a good performance.

The contributions of this paper are threefold:

(a) A general definition of problem SUS is proposed.

(b) A novel framework for SUS that incorporates the technologies of time-based topic-level and sentence-level summarization is introduced.

(c) An application of this framework to the sequential update summarization task of temporal summarization (TS) track is implemented.

In the rest of this paper, we firstly review some related work on information retrieval and text summarization in Section 2. Then, we formalize the problem of sequential update summarization extraction in Section 3. In Section 4, we present the novel hierarchical update mining system and we introduce the evaluation criteria in Section 5. We conduct experiment to verify the effectiveness of our proposed method in Section 6 and conclude in Section 7.

2. Related Work

The problem of sequential update summarization has its roots in topic detection and tracking [23], time-based summarization techniques [14, 20], and multidocuments summarization [9, 24, 25].

2.1. Topic Detection and Tracking. Topic detection and tracking (TDT) refers to the document-level tasks which associated with detecting and tracking news events [23]. It is a body of research and an evaluation paradigm that addresses event-based organization of broadcasting news.

Authors of [20] suggested retrospectively selecting novel and relevant sentences from a stream of news articles. However, the TDT is more topic based than sentence based. In most ways, the sequential update summarization is an event- and sentence-level analogue of TDT's "first topic detection" problem [20].

Referring to the time-based summarization as the task of temporal summarization, most of these systems focus on temporal expression extraction from text normalizing references to dates, times, and elapsed times [14]. The system in [26] generated the meaningful temporal summarization of event-related updates and automatically annotates the identified events in a timeline. Methods proposed in [27] retrieved sequential versions of a single web page during predefined time intervals. The paper [28] presented a framework that extracts events relevant to a query from a collection of documents and placed these events along a timeline.

2.2. Multidocuments Summarization. Text summarization techniques leverage a wide range of information retrieval (IR) and natural language processing (NLP) techniques. Some focus primarily on techniques that have been developed in IR [25], while most try to leverage both IR approaches and some aspects of NLP [19]. As one of the subproblems of text summarization, multidocument summarization (MDS), which refers to the task of generating a text summary of a pool of documents on the same topic, includes two broad approaches: extractive summarization and abstractive summarization. The extractive summarization extracts summary which consists of sentences extracted from the pool of documents, while the abstractive summarization extracts summary generated based on the pool of documents.

The core technique of the extractive summarization research is to summarize a body of texts by extracting

sentences that have particular properties. Sentence extraction techniques consider the words in the sentences, look for cue words and phrases [11, 24], consider even more focused features such as sentence length and case of words [29], or compare patterns of relationships between sentences [30–32]. Most of these approaches use statistics from the corpus itself to decide on the importance of sentences, and some leverage existing training sets of summaries to learn the properties of a summary [29, 33]. Other methods computed sentence importance based on the eigenvector of a graph representation of sentences [34].

Methods investigated in this paper are mainly similar to extractive summarization. The goal of our proposed method is to extract time-based sentences which have high confidence.

3. Problem Definition

The problem of sequential update summarization has been investigated in many literatures. However, until now, there is still no clear definition on it. In this section, we will give a general definition on the problem of sequential update summarization as follows.

An unexpected event, e, is a temporally acute topic with a clear onset time, $[t_s, t_e]$. An event query, Q_e, is the representation of the event description expressed by a user during the event. The set of keywords associated with the event, $\mathcal{K}(e)$, represents the important information that should be included in updates to deliver to users (e.g., the location where the event happened, the death number caused by this event). The system observes a temporally ordered stream of documents, $[d_1, d_2, \ldots]$. On the observation of d_t, the system makes a decision to emit zero or more updates. The pool of candidate updates consists of sentences in documents comprised of the most recent k documents in the event timeframe. Figure 1 illustrates a schematic diagram of the sequential update summarization system. Based on the schematic diagram of Figure 1, we present a general framework of sequential update summarization in Algorithm 1. According to Algorithm 1, an effective sequential update summarization system should be supported by time-sensitive information retrieval technique, accurate keywords mining method, and effective updates scoring algorithm.

Require:
SequentialUpdateSummarization $\{S, C, Q_e, t_s, t_e\}$:
S = the SUS system;
C = time-ordered corpus;
Q_e = keyword query of a sudden event;
t_s = start time of a sudden event;
t_e = end time of a sudden event;
Ensure: updates set U
(1) $U \leftarrow \{\}$;
(2) S: *Retrieval*(Q_e)
(3) **for** $d \in C$ **do**
(4) *do*
(5) S : *Process*(d);
(6) $t \leftarrow d.Time()$;
(7) **if** $t \in \{t_s, t_e\}$ **then**
(8) *then*
(9) $U_t \leftarrow S.Decide$;
(10) **for** $u \in U_t$ **do**
(11) *do*
(12) $U.Append(u; t)$;
(13) **end for**
(14) **end if**
(15) **end for**

Algorithm 1: Sequential Update Summarization System.

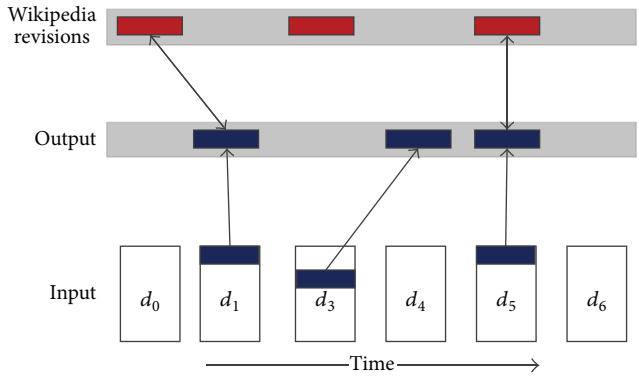


Figure 1: The sequential update summarization system observes a buffered stream of documents and makes decisions based on the contents of the input buffer and the timestamp of these documents to form the real update, which actually are time-based Wikipedia revisions.

4. Hierarchical Sequential Update Summarization System

To investigate update mining methods on unexpected events, we construct a hierarchical sequential update summarization mining system in this section. The framework of the system is illustrated in Figure 2. The framework contains three main modules: preprocessing and information retrieval module, keywords mining module, and sentence scoring module. The first module makes sure the event-relevant documents are time sensitive. The second module extracts time-based event-relevant keywords by using the hierarchical text analysis techniques. The third module focuses on scoring novel sentence-level updates.

4.1. Preprocessing and Information Retrieval Module. Because the original dataset is processed with some specific technologies, such as encryption, compression, and serialization [37], the system should firstly do some preprocessing on the available data and extract event-relevant document during each timeframe. The overall process of this module is described as follows:

(i) *Decrypt File*. The first step is to decrypt the files using the authorized key from authority. This step converts the GPG file format to SC file format.

(ii) *Deserialization*. We use stream corpus toolbox to parse these SC files to TXT files. The authority of

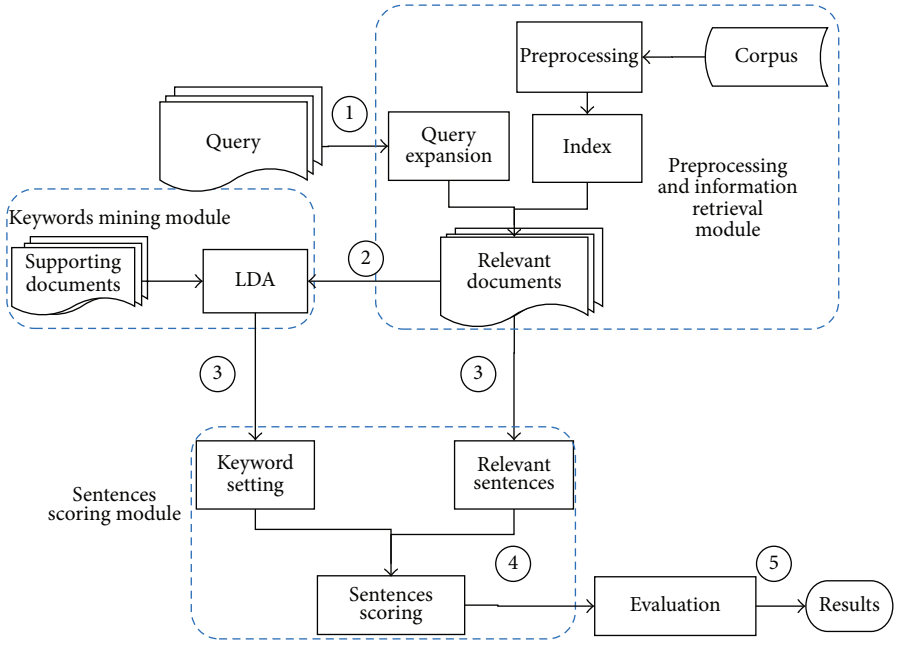


Figure 2: The framework of the hierarchical sequential update summarization system [35].

TREC provided the stream corpus toolbox to parse these SC files. The stream corpus toolbox gives a common data interchange format for document processing pipelines, which applies language processing tools to large streams of text.

(iii) *Build Index*. To obtain topic-relevant documents from large stream documents, the index of these big data should be built. This step is to build index by Indri [38] for query-based information retrieval. Indri is one of the mostly used search engines in information retrieval domain, which combines inference networks with language modeling. The query language of Indri, which is reminiscent of the Inquery query language, allows researchers to experiment with proximity, document structure, text passages, and other document features without writing code.

(iv) *Information Retrieval*. The last step is to use Indri as a tool for information retrieval. Given an event query Q_e, Indri returns all ranked relevant documents according to their responding confidence computed by the criterion of Indri. This step enables users to submit the queries and obtain the most relevant documents in each timeframe.

4.2. Keywords Mining Module. In this module, we utilize hierarchical Latent Dirichlet Allocation to find potential topics and return the most representative words of each topic as keywords.

Latent Dirichlet Allocation (LDA) [36] is a statistical model, specially a topic model, which can be used to identify hidden topic from a large document collection corpus. The basic idea of LDA is that a document can be considered as a mixture of a limited number of topics and each meaningful word in the document can be associated with one of these topics. Given a corpus of documents, LDA attempts to identify a set of topics, associate a set of words with a topic, and define a specific mixture of these topics for each document in the corpus. A thorough and complete description of the LDA model can be found in [36]. The vocabulary for describing the LDA model is as follows:

(i) *Word.* A word is a basic unit defined to be an item from a vocabulary of size W.

(ii) *Document.* A document is a sequence of n words denoted by $d = (w_1, \ldots, w_n)$, where w_n is the nth word in the sequence.

(iii) *Corpus.* A corpus is a collection of M documents denoted by $D = (d_1, \ldots, d_M)$.

In the statistical natural language processing, it is common to model each document d as a multinomial distribution θ_d over T topics and each topic z_j, $j = 1 \cdots T$, as a multinomial distribution $\phi^{(j)}$ over the set of words W. In order to discover the set of topics used and the distribution of these topics in each document in a corpus of documents D, we need to obtain an estimate of ϕ and θ. Blei et al. [36] have shown that the existing techniques of estimating ϕ and θ are slow to converge and propose a new model LDA. The LDA based model assumes a prior Dirichlet distribution on θ, thus allowing the estimation of ϕ without requiring the estimation of θ.

LDA assumes a generative process for creating a document [36] as presented below:

(i) Choose $N \sim Poisson(\xi)$: select the number of words N.

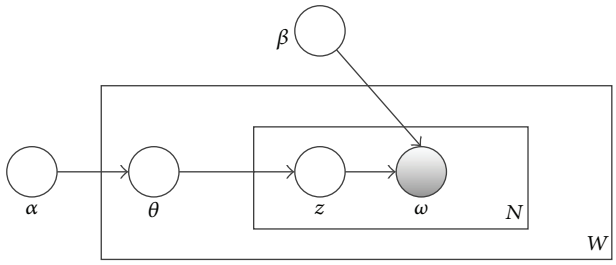


Figure 3: The probabilistic graphical model of Latent Dirichlet Allocation (LDA) [36].

(ii) $\theta \sim Dir(\alpha)$: select θ from the Dirichlet distribution parameterized by α.

(iii) For each $w_n \in w$,

(a) choose topic $z_n \sim Multinomial(\theta)$;

(b) choose a word (w_n) from $p(w_n \mid z_n, \beta)$, a multinomial probability ϕ^{z_n}.

In this model, various distributions, namely, the set of topics, topic distribution for each of the documents, and word probabilities for each of the topics, are in general intractable for exact inference [36]. The probabilistic graphical model of LDA is illustrated in Figure 3. The joint probability distribution of LDA is

$$p(\theta, \mathbf{z}, \mathbf{w} \mid \alpha, \beta) = p(\theta \mid \alpha) \prod_{n=1} p(z_n \mid \theta) \cdot p(w_n \mid z_n, \beta). \quad (1)$$

Hence, a wide variety of approximate algorithms are considered for LDA. These algorithms attempt to maximize likelihood of the corpus given the model. A few algorithms have been proposed for fitting the LDA model to a text corpus such as variational Bayes [36, 39–41], expectation propagation [42], and Gibbs sampling [43].

In this paper, for each event in each hour, we firstly retrieve the most 500 relevant documents and then extract keywords by LDA in current hour. In this module, we use the GibbsLDA++ tool [44] to extract keywords. We firstly use the LDA toolkit to discover two topics and choose the most representative words for each topic; secondly, we discover 5 new topics by the same method under the topic discovered in the last step and choose the most representative words of each topic; lastly, we integrate the two level representative words of each topic to form keywords set $\mathcal{K}(e)$.

4.3. Sentences Scoring Module. We utilize three sentence scoring methods in this module: KLP method, SKD method, and KS method [45].

The first method assumes an update is a long sentence which should shoot many keywords and be placed on the first place in a paragraph. Hence, it considers three important factors: the keywords diversity, the length of a sentence, and the position of the sentence, which we named KLP method. The scoring metric is as follows:

$$\text{Score}(s_i) = \alpha \frac{\sum_{w \in s_i} tf(w) \cdot idf(w)}{\max_{s_j \in d} \left\{ \sum_{w \in s_j} tf(w) \cdot idf(w) \right\}} + \beta \frac{\text{Length}_{s_i}}{\max_{s_j} \left\{ \text{length}_{s_j} \right\}} + \gamma \, \text{position}_{s_i}, \quad (2)$$

where $w \in \mathcal{K}(e)$ is one of the keywords of event e extracted in Section 4.2 and α, β, and γ are weights of the keyword diversity, length, and position, respectively. When computing $idf(w)$, the documents are referred to relevant documents in the current hour. If a sentence is placed on the first place of a paragraph, $\text{position}_{s_i} = 1$, or $\text{position}_{s_i} = 0$.

The second method assumes that an update should be a short length sentence with larger keywords diversity, because a too long sentence is normally a retrospective summary of an event, not an update. We named this metric SKD, whose scoring metric is as follows:

$$\text{Score}(s_i) = \frac{1}{N(N+1) \cdot \text{Length}} \cdot \sum_{j=1}^{k-1} \frac{\text{Score}(w_j) \cdot \text{Score}(w_{j+1})}{\text{distance}(w_j, w_{j+1})}, \quad (3)$$

where N is the number of keywords included in s_i, Score(w) is the confidence of keyword w obtained from Section 4.2, and distance(w_j, w_{j+1}) is the distance between w_j and w_{j+1}.

The third method is a keyword shooting method, which only considers the diversity of keywords included in the sentence. We named it KS method. Its scoring metric is as follows:

$$\text{Score}(s_i) = \frac{\left| V_{\text{keywords}} \cap s_i \right|}{s_i}, \quad (4)$$

where V_{keywords} is the keyword vector of the event e. s_i is the ith related sentences of event e.

After getting high confidence sentences, the postprocessing module will do the duplicate removal to sentences, which first finds the same sentences with different sentences ID and then compares the stream ID of all sentences and chooses the one with the earliest time information as the submission sentence.

5. Evaluation Method

Document summaries are difficult to evaluate, because all results in minor variations, such as rewording portions of the summary, reordering the sentences, and omitting dubiously important information, are still excellent summaries. The most popular summary evaluation method is comparing agreement between sentences selected by experts and sentences selected by computer [9, 46], or comparing agreement in the ranks of sentences that a system generates [47].

However, since comparing based on some sentence variant is difficult, we introduce the concept of gold nugget, which is defined as atomic novel pieces of information relevant to unexpected events. For example, in the task of SUS, gold nuggets are text perceived as relevant and novel for the edit of Wikipedia articles. Each gold nugget $n \in N$ is assigned with an importance grade by annotators: $R : N \rightarrow [0, 1]$. Hence, we can compare extracted updates with these atomic gold nuggets in a more accurate manner.

Traditional IR and text summarization evaluations are concerned with the quality and the quantity of relevant materials. In this paper, the sequential update summarization system focuses on the following properties:

(i) Updates are relevant sentences to the unexpected events.

(ii) Updates should be novel which must match with at least one gold nugget and can be matched with several gold nuggets.

(iii) Updates are sentences which are early extracted from event-relevant news. The first sentence about an event is clearly novel; the earlier the time of the first sentence of an event, the lower the latency of the update.

(iv) Updates are short sentences which should not be too verbose.

That is to say, we want to measure the relevance, latency, verbosity, and matching of the extracted updates. To measure the abovementioned properties, the SUS task of TREC defined the measurement of four parameters: expected gain, latency expected gain, comprehensiveness, and latency comprehensiveness [48].

Before introducing the definition of the four parameters, we firstly explain some fundamental definitions. Given an update set U and a gold nugget n, the matching function between them is

$$M(n, U) = \underset{\{n \in U : n \approx u\}}{\operatorname{argmin}} u \cdot t. \tag{5}$$

Besides the matching function, two discounts are defined to evaluate the timeliness and conciseness of the extracted updates set: latency discount and verbosity discount. Given a nugget whose timestamp is t', the latency discount is a latency penalty $L_d(t', t)$, which is a monotonically decreasing function of $t - t'$. Similarly, verbosity discount is also a penalty function $V_d(u)$, which is defined as a string length penalty function, monotonically decreasing in the number of words of the update string. Based on the abovementioned concepts, the discounted gain between an update u and a matching nugget n is

$$g(u, n) = R(n) * \text{discount factor}, \tag{6}$$

where the discount factor can be latency discount, verbosity discount, or the compound of the two discounts (e.g., $L_d * V_d$).

Hence, the overall expected gain is similar to traditional notions of precision in IR. It is defined as

$$\text{MEG} = \frac{1}{|E|} \sum_{e \in E} \text{EG}\left(U^e\right), \tag{7}$$

where E is the set of evaluation events and U^e is the system submission for event e, and EG(U) is defined as

$$\text{EG}(U) = \frac{1}{\sum_{u \in U} V(u)} \sum_{\{n \in N : M(n,U) \neq \phi\}} g(M(n, U), n). \tag{8}$$

To evaluate the system performance on the time after an event, the latency gain is defined as the time-sensitive expected gain for the first τ seconds as

$$\text{EG}_\tau(U) = \text{EG}\left(U_\tau\right). \tag{9}$$

In addition to good expected gain, the performance of providing a comprehensive set of updates is also very important. That is to say, the more nuggets the extracted updates set covers, the better the system performs. It is similar to traditional notions of recall in information retrieval evaluation. Given a set of system updates, the comprehensiveness is similar to the recall of IR, which evaluates the coverage on gold nuggets as

$$C(U) = \frac{1}{\sum_{n \in N} R(n)} \sum_{\{n \in N : M(n,U) = \phi\}} g(M(n, U), n). \tag{10}$$

Similarly, the latency comprehensiveness is a time-sensitive notion of comprehensiveness as follows:

$$C_\tau(U) = C\left(U_\tau\right). \tag{11}$$

6. Experimental Results and Discussions

6.1. Data and Topics. The data used in the SUS task of TS track is provided by the Organizer of KBA track [49] at TREC, which is hosted by Amazon Public Dataset service. This corpus [50] consists of a set of timestamped documents from a variety of news and social media sources covering the time period October 2011 through January 2013, whose time span is 17 months with 11,248 hours. There are more than 1 billion documents, each with absolute timestamp that places it in the stream, which is mainly composed of news, social (blog, forum), and web (e.g., arxiv, linking events) content. All documents contain a set of sentences, each with a unique identifier.

There are 10 events/topics (listed in Table 1) [51] in the SUS task; each has a single type title, description (URL to Wikipedia entry), beginning and end times, and query keywords. Types are taken from {accident, shooting, storm, earthquake, bombing} and they have a set of attributes, such as location, death, and financial impact. Algorithm 2 illustrates the definition of the event of "2012 Buenos Aires Rail Disaster." For each sudden event query, we chose the top 500 relevant documents returned by Indri as the relevant documents of each sudden event query in one hour.

6.2. Results. We applied our hierarchical update mining system on the overall ten topics. For each topic, to evaluate these extracted updates, we chose the top 60 updates as the assessment data due to their confidences computed by the KLP, SKD, and KS method. The evaluation processes were

Table 1: Queries and titles of 10 topics of temporal summarization track [51].

Query of topics	Title of topics
(1) Buenos aires train crash	2012 Buenos Aires Rail Disaster
(2) Pakistan factory fire	2012 Pakistan garment factory fires
(3) Colorado shooting	2012 Aurora shooting
(4) Sikh temple shooting	Wisconsin Sikh temple shooting
(5) Hurricane isaac	Hurricane Isaac (2012)
(6) Hurricane sandy	Hurricane Sandy
(7) Midwest derecho	June 2012 North American derecho
(8) Typhoon bopha	Typhoon Bopha
(9) Guatemala earthquake	2012 Guatemala earthquake
(10) Tel aviv bus bombing	2012 Tel Aviv bus bombing

```
<event>
  <id>1</id>
  <start>1329910380</start>
  <end>1330774380</end>
  <query>buenos aires train crash</query>
  <type>accident</type>
  <locations/>
  <deaths/>
  <injuries/>
</event>
```

Algorithm 2: An unexpected event definition for "2012 Buenos Aires Rail Disaster" in the SUS task [48].

mainly gold nuggets extraction and update-nugget matching. In this paper, the gold nugget was extracted by assessors by reading all edits of the Wikipedia article for each topic, manually extracting text perceived as relevant and novel for that edit. Additionally, they assigned an importance grade to every text fragment, or nugget, and noted any dependencies in the information. The update-nugget matching refers to matching our extracted updates to these gold nuggets to evaluate their accuracy and coverage of the information. The latency discount function and the verbosity discount function [48] used in this paper are

$$L(n \cdot t, u \cdot t) = 1 - \frac{2}{\pi} \arctan\left(\frac{u \cdot t - n \cdot t}{\alpha}\right),$$
$$V(u) = 1 + \frac{|u| - \left|\bigcup_{n \in M^{-1}(u,U)} M(n,U)\right|}{\text{avg}_{n \in N} |n|}, \quad (12)$$

where $\alpha = 3600 * 6$ is the latency step (6 hours) and $|u|$ and $|n|$ are the length (in number of words) of the update u and nugget n. By applying the abovementioned functions on evaluation metric introduced in Section 5, we computed the four performance parameters.

Table 2: The μ and σ of expected gain and expected latency gain over all events of the multi-level SUS system. (The E [gain] is the expected gain which is similar to traditional notions of precision in information retrieval; E [latency gain] is the time-sensitive expected gain).

Methods	E [Gain]	E [latency gain]
The best reported	***0.149*** (***0.101***)	***0.136*** (***0.090***)
ICTNET-run2	*0.102* (*0.045*)	*0.127* (*0.075*)
ICTNET-run1	*0.101* (*0.045*)	*0.125* (*0.075*)
Mid-value	*0.053* (*0.041*)	*0.067* (*0.057*)
KS	0.149 (0.101)	0.136 (0.090)
SKD	0.103 (0.084)	0.103 (0.050)
KLP (0.6, 0.2, 0.2)	0.071 (0.039)	0.074 (0.031)
KLP (0.5, 0.2, 0.3)	0.065 (0.034)	0.067 (0.026)
KLP (0.5, 0.3, 0.2)	0.065 (0.034)	0.067 (0.026)

Table 3: The μ and σ of comprehensiveness and latency comprehensiveness over all events of the multi-level SUS system (Comprehensiveness is similar to recall in IE, which evaluates coverage of nuggets; latency Comp. is the time-sensitive comprehensiveness).

Methods	Comprehensive	Latency Comp.
The best reported	***0.445*** (***0.191***)	***0.571*** (***0.358***)
UWaterloo-rg2	*0.441* (*0.198*)	*0.562* (*0.349*)
UWaterloo-qlec2t25	*0.433* (*0.170*)	*0.537* (*0.322*)
Mid-value	*0.204* (*0.146*)	*0.260* (*0.217*)
KLP (0.5, 0.3, 0.2)	0.224 (0.178)	0.292 (0.270)
KLP (0.5, 0.2, 0.3)	0.224 (0.178)	0.288 (0.262)
KLP (0.6, 0.2, 0.2)	0.204 (0.146)	0.260 (0.217)
SKD	0.131 (0.138)	0.176 (0.203)
KS	0.099 (0.099)	0.126 (0.164)

In addition to our previously reported results [35], Tables 2 and 3 illustrate some results reported by the SUS task of TREC 2013 and the five results of these three methods. The four parameters are evaluated by comparing generated updates with gold nuggets by using expected gain, expected latency gain, comprehensiveness, and latency comprehensiveness metrics. The expected gain is similar to traditional notions of precision in IR. Expected latency gain is a time-sensitive expected gain. Comprehensiveness is similar to recall in IR, which evaluates the coverage of gold nuggets. The latency comp. is the time-sensitive comprehensiveness [48]. The results in italics are the top 3 and the midvalue results based on corresponding parameters, which are reported in the SUS task in 2013 [48].

Table 2 illustrates the top three and the midvalue results of the expected gain and the latency expected gain in the SUS task in TREC 2013. The ICTNET-run2 and ICTNET-run1 [52] are results submitted by the Institute of Computing Technology, Chinese Academy of Sciences. They firstly chose event-relevant sentences and decided a sentence as an update if it includes words of a handpicked trigger word list, such as kill and death. From Table 2, we can see that the KS method has the best expected gain and expected latency gain, which are equal to the best reported results. That is because

the keywords list of KS method is generated by hierarchical LDA method, which can generate much more accurate keywords lists compared with the man-made keywords list of ICTNET methods. Hence, the KS method is superior to the two ICTNET methods. Table 2 also illustrates that the expected gain and latency gain of the KLP and SKD are all above the midvalue result, which shows that the KLP and SKD methods are effective methods in extracting updates of unexpected news events. By comparing the results of our three investigated methods in Table 2, we can conclude that the KS method is the most effective method in evaluating the metric of expected gain and expected latency gain; for example, it can extract updates in an accurately and timely manner.

Table 3 illustrates the top three and the midvalue results of comprehensiveness and latency comprehensiveness in the SUS task in TREC 2013. The reported top three results are submitted by the University of Waterloo. The three Waterloo methods tried to extracted updates in two aspects: sentence scoring and event-relevant terms' expansion [53]. The term expansion method investigated by the University of Waterloo is based on bootstrap learning from seed terms. The good results of the three Waterloo methods indicate the effectiveness of the term expansion method, which lead to the best comprehensiveness and latency comprehensiveness. Table 3 shows that our three investigated methods are all above the reported midvalue result, which shows the effectiveness of the three methods. From Table 3, we can see that the KLP method has the best comprehensiveness and latency comp., and KS method has the worst comp. and latency comp., while the performance of SKD method is between the KLP method and KS method. That is to say, comparing with the KS and SKD method, the KLP method utilized a more general metric on scoring updates which can cover much more nuggets.

By comparing the different weights of KLP method from Tables 2 and 3, we can see that the weights on sentence length and sentence position have little effect on the update extraction results in the KLP method. It indicates that the keyword diversity is more important than sentence length and sentence position in the KLP method.

In addition, by combining the results of Tables 2 and 3, we can see that the expected gain has reciprocal relationship with comprehensiveness, like the precision and recall in information retrieval. The KLP method utilizes a more comprehensive metric which considers more factors in scoring sentences. But it is threatened to choose long sentences which leads to the worst gain and latency gain. The KS method proposed only the keyword diversity to evaluate sentences, and it has good performance on expected gain and expected latency gain.

In summary, the keywords in our proposed system are extracted in topic level by using hierarchical LDA. The good results of KS and SKD method, whose key criterion is keywords mentioning in sentence level, show that the SUS extraction is an event- and sentence-level analogue of first topic detection problem. Hence, it is effective when extracted by hierarchical text analysis. Experimental results indicate that a good update should not be a too long sentence which covers many keywords. Generally speaking, the KS method is suitable for systems which have demands on high accuracy, while the KLP method is more suitable for systems which demand high recall.

7. Conclusions

This paper defined the problem of sequential update summarization extraction for unexpected events. To extract relevant timely updates, we formalized a hierarchical sequential update summarization system, which incorporates techniques from topic-level and sentence-level summarization. The hierarchical mining system focused attention on the SUS task and tried to broadcast with useful, new, and timely sentence-length updates about developing unexpected events. To verify the effectiveness of our proposed system, we provided a rounded system based on the SUS task of TREC 2013, including query topics, updates extraction system, and evaluation metrics. We applied the hierarchical update mining system to extract updates of ten unexpected events of the SUS task. Experimental results showed that our proposed system has good performance.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is partly supported by the National Natural Science Foundation of China (NSFC) Grant nos. 61402047 and 61511130081, NSFC Joint Fund with Guangdong under Key Project no. U1201258, Scientific Research Foundation for Returned Scholars, Ministry of Education of China, Sweden STINT Initiation Grant Dnr. IB2015-5959, EU FP7 IRSESMobileCloud Project (Grant no. 612212), and Shandong Natural Science Funds for Distinguished Young Scholar under Grant no. JQ201316. Part of the work presented in this paper has been published in [35].

References

[1] V. Madisetti and A. Bahga, "Internet of things," 2014.

[2] L. Atzori, A. Iera, and G. Morabito, "From 'smart objects' to 'social objects': the next evolutionary step of the internet of things," *IEEE Communications Magazine*, vol. 52, no. 1, pp. 97–105, 2014.

[3] J. A. Stankovic, "Research directions for the internet of things," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 3–9, 2014.

[4] C.-W. Tsai, C.-F. Lai, M.-C. Chiang, and L. T. Yang, "Data mining for internet of things: a survey," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 77–97, 2014.

[5] Q. Guo, F. Diaz, and E. Yom-Tov, "Updating users about time critical events," in *Advances in Information Retrieval*, pp. 483–494, Springer, 2013.

[6] M. Mendoza, B. Poblete, and C. Castillo, "Twitter under crisis: can we trust what we RT?" in *Proceedings of the 1st Workshop on Social Media Analytics (SOMA '10)*, pp. 71–79, ACM, Washington, DC, USA, July 2010.

[7] E. Yom-Tov and F. Diaz, "Out of sight, not out of mind: on the effect of social and physical detachment on information need," in *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '11)*, pp. 385–394, ACM, Beijing, China, July 2011.

[8] H. P. Luhn, "The automatic creation of literature abstracts," *IBM Journal of Research and Development*, vol. 2, no. 2, pp. 159–165, 1958.

[9] J. Goldstein, M. Kantrowitz, V. Mittal, and J. Carbonell, "Summarizing text documents: sentence selection and evaluation metrics," in *Proceedings of the 22nd ACM Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '99)*, pp. 121–128, Berkeley, Calif, USA, August 1999.

[10] D. Wang, S. Zhu, T. Li, and Y. Gong, "Comparative document summarization via discriminative sentence selection," *ACM Transactions on Knowledge Discovery from Data*, vol. 6, no. 3, article 12, 2012.

[11] J. J. Pollock and A. Zamora, "Automatic abstracting research at chemical abstracts service," *Journal of Chemical Information and Computer Sciences*, vol. 15, no. 4, pp. 226–232, 1975.

[12] A. A. A. Esmin, R. S. C. Júnior, W. S. Santos, C. O. Botaro, and T. P. Nobre, "Real-time summarization of scheduled soccer games from twitter stream," in *Natural Language Processing and Information Systems*, E. Métais, M. Roche, and M. Teisseire, Eds., vol. 8455 of *Lecture Notes in Computer Science*, pp. 220–223, Springer, 2014.

[13] A. Patil, K. Pharande, D. Nale, and R. Agrawal, "Automatic text summarization," *International Journal of Computer Applications*, vol. 109, no. 17, pp. 18–19, 2015.

[14] I. Mani and G. Wilson, "Robust temporal processing of news," in *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics*, pp. 69–76, Association for Computational Linguistics, Stroudsburg, Pa, USA, October 2000.

[15] R. Swan and J. Allan, "Automatic generation of overview timelines," in *Proceedings of the 23rd ACM Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '00)*, pp. 49–56, Athens, Greece, July 2000.

[16] W. Ding and C. Chen, "Dynamic topic detection and tracking: a comparison of HDP, C-word, and cocitation methods," *Journal of the Association for Information Science and Technology*, vol. 65, no. 10, pp. 2084–2097, 2014.

[17] M. Osborne, S. Moran, R. McCreadie et al., "Real-time detection, tracking, and monitoring of automatically discovered events in social media," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations (ACL '14)*, pp. 37–42, Association for Computational Linguistics, Baltimore, Md, USA, June 2014.

[18] A. Guille and C. Favre, "Mention-anomaly-based event detection and tracking in twitter," in *Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM '13)*, pp. 375–382, IEEE, Beijing, China, August 2014.

[19] E. Hovy and C.-Y. Lin, "Automated text summarization and the summarist system," in *Proceedings of the TIPSTER Text Program*, pp. 197–214, Association for Computational Linguistics, Baltimore, Md, USA, October 1998.

[20] J. Allan, R. Gupta, and V. Khandelwal, "Temporal summaries of new topics," in *Proceedings of the 24th ACM Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '01)*, pp. 10–18, New Orleans, LA, USA, September 2001.

[21] Temporal summarization, 2013, http://www.trec-ts.org/.

[22] Trec, 2013, http://trec.nist.gov/.

[23] J. Allan, "Introduction to topic detection and tracking," in *Topic Detection and Tracking*, pp. 1–16, Springer, 2002.

[24] H. P. Edmundson, "New methods in automatic extracting," *Journal of the ACM*, vol. 16, no. 2, pp. 264–285, 1969.

[25] J. Goldstein, V. Mittal, J. Carbonell, and M. Kantrowitz, "Multi-document summarization by sentence extraction," in *Proceeding of the NAACL-ANLP Workshop on Automatic Summarization (NAACL-ANLP-AutoSum '00)*, vol. 4, pp. 40–48, Association for Computational Linguistics, April 2000.

[26] M. Georgescu, D. D. Pham, N. Kanhabua, S. Zerr, S. Siersdorfer, and W. Nejdl, "Temporal summarization of event-related updates in wikipedia," in *Proceedings of the 22nd International Conference on World Wide Web (WWW '13)*, pp. 281–284, International World Wide Web Conferences Steering Committee, Rio de Janeiro, Brazil, May 2013.

[27] A. Jatowt and M. Ishizuka, "Temporal web page summarization," in *Web Information Systems—WISE 2004: 5th International Conference on Web Information Systems Engineering, Brisbane, Australia, November 22–24, 2004. Proceedings*, vol. 3306 of *Lecture Notes in Computer Science*, pp. 303–312, Springer, Berlin, Germany, 2004.

[28] H. L. Chieu and Y. K. Lee, "Query based event extraction along a timeline," in *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 425–432, ACM, July 2004.

[29] J. Kupiec, J. Pedersen, and F. Chen, "A trainable document summarizer," in *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 68–73, ACM, July 1995.

[30] G. Salton, A. Singhal, M. Mitra, and C. Buckley, "Automatic text structuring and summarization," *Information Processing & Management*, vol. 33, no. 2, pp. 193–207, 1997.

[31] C. Zhang, W. Xu, Z. Ma, S. Gao, Q. Li, and J. Guo, "Construction of semantic bootstrapping models for relation extraction," *Knowledge-Based Systems*, vol. 83, pp. 128–137, 2015.

[32] C. Zhang, Y. Zhang, W. Xu, Z. Ma, Y. Leng, and J. Guo, "Mining activation force defined dependency patterns for relation extraction," *Knowledge-Based Systems*, vol. 86, pp. 278–287, 2015.

[33] A. L. Berger and V. O. Mittal, "OCELOT: a system for summarizing Web pages," in *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '00)*, pp. 144–151, ACM, Athens, Greece, July 2000.

[34] G. Erkan and D. R. Radev, "LexRank: graph-based lexical centrality as salience in text summarization," *Journal of Artificial Intelligence Research*, vol. 22, no. 1, pp. 457–479, 2004.

[35] C. Zhang, Z. Ma, J. Zhang, W. Xu, and J. Guo, "A multi-level system for sequential update summarization," in *Proceedings of the IEEE 11th International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness (QSHINE '15)*, pp. 144–148, Taipei, Taiwan, August 2015.

[36] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *The Journal of Machine Learning Research*, vol. 3, no. 4-5, pp. 993–1022, 2003.

[37] C. Zhang, W. Xu, R. Liu et al., "Pris at trec kba," in *Notebook of the TExt Retrieval Conference*, 2013.

[38] Indri, http://www.lemurproject.org/indri.php.

[39] Z. Ma and A. Leijon, "Bayesian estimation of beta mixture models with variational inference," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 11, pp. 2160–2173, 2011.

[40] Z. Ma, P. K. Rana, J. Taghia, M. Flierl, and A. Leijon, "Bayesian estimation of dirichlet mixture model with variational inference," *Pattern Recognition*, vol. 47, no. 9, pp. 3143–3157, 2014.

[41] Z. Ma, A. E. Teschendorff, A. Leijon, Y. Qiao, H. Zhang, and J. Guo, "Variational bayesian matrix factorization for bounded support data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 4, pp. 876–889, 2015.

[42] T. Minka and J. Lafferty, "Expectation-propagation for the generative aspect model," in *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence (UAI '02)*, pp. 352–359, Morgan Kaufmann Publishers, Alberta, Canada, August 2002.

[43] T. L. Griffiths and M. Steyvers, "Finding scientific topics," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, supplement 1, pp. 5228–5235, 2004.

[44] "Gibbs lda ++," http://sourceforge.net/projects/gibbslda/.

[45] C. Zhang, Z. Ma, J. Zhang, W. Xu, and J. Guo, "A multi-level system for sequential update summarization," in *Proceedings of the 11th International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness (QSHINE '15)*, Taipei, Taiwan, August 2015.

[46] G. J. Rath, A. Resnick, and T. R. Savage, "The formation of abstracts by the selection of sentences. Part I. Sentence selection by men and machines," *American Documentation*, vol. 12, no. 2, pp. 139–141, 1961.

[47] R. L. Donaway, K. W. Drummey, and L. A. Mather, "A comparison of rankings produced by summarization evaluation measures," in *Proceedings of the NAACL-ANLP 2000 Workshop on Automatic summarization*, pp. 69–78, Seattle, Wash, USA, April 2000.

[48] J. Aslam, M. Ekstrand-Abueg, V. Pavlu, F. Diaz, and T. Sakai, "Trec 2013 temporal summarization," in *Proceedings of the 22nd Text Retrieval Conference (TREC '13)*, Gaithersburg, Md, USA, November 2013.

[49] "Knowledge based acceleration," 2013, http://trec-kba.org/.

[50] "Kba data," 2013, http://s3.amazonaws.com/aws-publicdatasets/trec/kba/index.html.

[51] Test topics, 2013, http://trec.nist.gov/data/tempsumm/2013/testTopics.xml.

[52] Q. Liu, Y. Liu, D. Wu, and X. Cheng, "ICTNET at temporal summarization track TREC 2013," in *Proceedings of the 22nd Text Retrieval Conference (TREC '13)*, 2013.

[53] G. Baruah, R. Guttikonda, A. Roegiest, and O. Vechtomova, "University of waterloo at the TREC 2013 temporal summarization track," in *Proceedings of the 22nd Text Retrieval Conference (TREC '13)*, Gaithersburg, Md, USA, November 2013.

Hindawi Publishing Corporation
Mobile Information Systems
Volume 2016, Article ID 7853219, 8 pages
http://dx.doi.org/10.1155/2016/7853219

Research Article

The PMIPv6-Based Group Binding Update for IoT Devices

Jianfeng Guan,[1] Ilsun You,[2] Changqiao Xu,[1] and Hongke Zhang[3]

[1]*State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China*
[2]*Department of Information Security Engineering, Soonchunhyang University, Asan-si 31538, Republic of Korea*
[3]*National Engineering Laboratory for Next Generation Internet Interconnection Devices, Beijing Jiaotong University, Beijing 100044, China*

Correspondence should be addressed to Ilsun You; ilsunu@gmail.com

Received 1 October 2015; Accepted 17 January 2016

Academic Editor: Kamal Deep Singh

Internet of Things (IoT) has been booming with rapid increase of the various wearable devices, vehicle embedded devices, and so on, and providing the effective mobility management for these IoT devices becomes a challenge due to the different application scenarios as well as the limited energy and bandwidth. Recently, lots of researchers have focused on this topic and proposed several solutions based on the combination of IoT features and traditional mobility management protocols, in which most of these schemes take the IoT devices as mobile networks and adopt the NEtwork MObility (NEMO) and its variants to provide the mobility support. However, these solutions are in face of the heavy signaling cost problem. Since IoT devices are generally combined to realize the complex functions, these devices may have similar movement behaviors. Clearly analyzing these characters and using them in the mobility management will reduce the signaling cost and improve the scalability. Motivated by this, we propose a PMIPv6-based group binding update method. In particular, we describe its group creation procedure, analyze its impact on the mobility management, and derive its reduction ratio in terms of signaling cost. The final results show that the introduction of group binding update can remarkably reduce the signaling cost.

1. Introduction

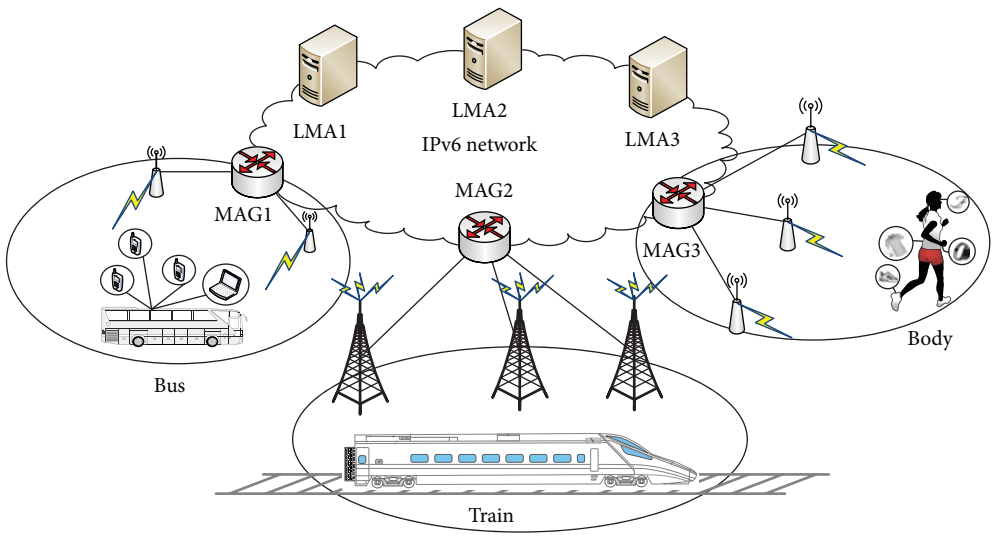
According to the statistics analysis, the number of Internet of Things (IoT) devices is expected to reach up to 50 billion by 2020 [1]. Figure 1 shows several typical application scenarios of IoT, which consist of vehicle networks, Wireless Body Networks, and so on. Due to the large volumes of mobile IoT devices, it becomes a big challenge to provide well mobility support for IoT [2]. IPv6 is believed to be a suitable protocol [3] thanks to its large address space and specific mechanisms to support mobility, such as Mobile IPv6 (MIPv6) [4] and its potential solutions for mobility management.

However, these solutions are designed for portable devices such as cell phone or PC which have different application requirement from the IoT devices; therefore, they need to be improved or enhanced in terms of bandwidth, energy consumption, and scalability.

It is worth to note from such application scenarios that the group is one of important characters in IoT [5–7]. Therefore, several works have focused on this problem and proposed lots of solutions [8–20] most of which tried to apply the NEtwork MObility (NEMO) [21] into the IoT.

NEMO as a mobility support protocol for mobile network is derived from MIPv6 in which mobile router (MR) is introduced to deliver all the packets for mobile network nodes via the bidirectional tunnel between MR and its Home Agent (HA) [21]. When it is applied in IoT scenarios, the MR is generally used as the leader to perform the mobility signaling messages on behalf of all mobile network nodes. However, due to the frequent mobility, the mobile network nodes will change their attachment points dynamically which may introduce the additional signaling and transmission costs due to the nested tunnels operations according to standard NEMO protocol procedure.

In this work we focus on group characteristics of IoT devices while not only studying the dynamic group management mechanism, but also extending the bulk binding update



LMA: Local Mobility Anchor
MAG: Mobile Access Gateway

Figure 1: Application scenarios of IoT.

of PMIPv6 [22] to set up the dynamic groups for IoT devices. The contributions of our work are summarized as follows:

(1) propose the group binding update scheme based on PMIPv6 to update a group of IoT devices to reduce the signaling cost;

(2) introduce the extensions for group identifier and the operation considerations of Mobility Access Gateway (MAG) and Local Mobility Anchor (LMA) and give the detailed operation flow of group binding update procedure;

(3) derive the signaling cost reduction ratio and analyze the relationship between group size and signaling cost.

The organization of this paper is shown as follows. Section 2 introduces the related works of IPv6 mobility management solutions and their applications in IoT. Section 3 describes the PMIPv6-based group binding update solution in detail. Section 4 models and analyzes the impact of group binding update of PMIPv6, followed by deriving the cost reduction ratio and giving the numerical results. Finally, Section 5 concludes our work.

2. Related Work

2.1. MIPv6 and Extensions for IoT. MIPv6 is a basic mobility support solution, which introduces the HA to perform the handover and location managements. In MIPv6, the HAs forward the packets via the bidirectional tunnels for each registered mobile node (MN) to maintain the ongoing sessions. MIPv6 introduces two kinds of address, Home Address (HoA) and Care-of Address (CoA), which are used to separate identifier functions and routing functions of IP address. In specific, HoA is the permanent identifier of MN used to set up the upper layer sessions, while CoA is the location identifier used to transmit the packets. Compared with MIPv4, MIPv6 also provides optimal routing process by maintaining the binding cache of each MN in Corresponding Node (CN). Hierarchical MIPv6 (HMIPv6) [23] was further proposed to solve the signaling cost problem in MIPv6 caused by the frequent Ping-Pong movements between two adjacent subnets, in which the Mobility Anchor Point (MAP) is introduced to deal with the micromobility on behalf of HA. Besides, fast handover for MIPv6 (FMIPv6) [24] was also proposed to reduce the handover delay with the assist of L2 trigger, which supports preconfiguration via Handover Indication (HI) and Handover Acknowledgment (HAck) messages to reduce the packet loss by setting up a short-time bidirectional tunnel between the previous attachment and new attachment points. However, these improved solutions require the involvements of MNs which may result in the excessive energy consumption for the power limited mobile devices and also restrict the wide deployment due to the additional function extensions.

Proxy MIPv6 (PMIPv6) [25] was proposed to deal with the mobility-related signaling messages. In particular, PMIPv6 introduces the MAG on behalf of MNs and emulate the home link for its attached MNs by setting up the bidirectional tunnel between the serving MAG and its LMA. Some works [26, 27] have testified that the PMIPv6 can reduce the handover delay and signaling cost greatly, and it is easy to deploy. Besides, some solutions such as F-PMIPv6 [28] were proposed by combining PMIPv6 and FMIPv6 to achieve better performance.

The above analyses show that the PMIPv6 and its variants are more suitable for mobile devices. However, these solutions are designed for traditional mobile devices such as smartphones, mobile Internet devices with the objective to provide

the continual connections for MNs, which is different from the requirements of the IoT devices. For example, the IoT session mobility is not always required [29]. Therefore, to provide the mobility management for IoT devices, some IoT characteristics should be considered, such as limited computation capacity, limited network resource, and movement behaviors. Considering many IoT devices having different movement trajectory in different application scenarios such as Intelligent Transportation Systems (ITSs), Internet of Cars, and Wireless Body Networks, it becomes very important to analyze the group character to optimize the mobility management [30].

2.2. NEMO and Extensions for IoT. NEMO is a basic solution to provide the group mobility support. However, this solution will bring in high packet delivery cost with the increase of nodes and multiple nests of mobile networks. Therefore, several improved solutions have been proposed in the past few years.

2.2.1. FMIPv6-Based Schemes. FMIPv6-based schemes were proposed by combining FMIPv6 and NEMO, such as fNEMO [8, 9]. After that, EfNEMO [10] was proposed by adopting the tentative binding update which pre-registers a new address assigned by Next Access Router (NAR) for the MR with the HA before the layer-2 association. The tentative binding update sets up a path between NAR and HA to deliver the traffic without the additional encapsulation. Therefore, it can reduce the handover latency and also reduce the traffic delivery cost thanks to the tentative binding update. This idea is similar to the transient binding for PMIPv6 [11], which can mitigate the tunneling burden and handover latency. However, they lack the consideration of dynamic character of mobile networks and are inefficient to deal with the dynamic of the mobile network nodes.

2.2.2. PMIPv6-Based Schemes. PMIPv6-based solutions aim to extend the PMIPv6 to support the network mobility. The early solution was proposed by Soto et al. [12] called NEMO-enabled PMIPv6 (N-PMIPv6), which takes the MR as the mobile MAG (mMAG). However, it just simply combines the PMIPv6 and NEMO without the consideration of the character of IoT devices. After that, Jeon and Kim [13] employed a relay station to mitigate the packet tunneling costs, which is called proxy router-based NEMO (PR-NEMO) that introduces the proxy router instead of the MR as forwarder between MAG and MNN to reduce the cost.

Besides, Lee et al. [14] also proposed the PMIPv6-based NEMO (P-NEMO) scheme which takes MR as MN and extends MAG and LMA to maintain mobile network prefix (MNP) for each MR and advertises MNP to mobile network nodes via Router Advertisement (RA) message. Furthermore, they also proposed the Fast P-NEMO (FP-NEMO) to improve the handover delay by combining P-NEMO with fast handover of PMIPv6. In vehicular networks, Kim and Lee [15] proposed a group-based fast handover by extending PMIPv6-based NEMO with the assist of bulk binding to reduce the overhead, in which the bulk binding update is used to perform the binding update and revocation operations on behalf of a group of mobility sessions [22]. Some detail operations of this solution are still under the way. For example, it requires the neighboring MAG information to set up the tunnels in advance which are only applied in some specific scenarios. This idea absorbs the advantages of both PMIPv6 and FMIPv6, which can provide better mobility management for mobile networks. However, they are still limited by NEMO, which cannot support the mobile network nodes efficiently.

2.2.3. DMM-Based Schemes. DMM-related schemes apply MIPv6, PMIPv6, and NEMO in a distributed way. Based on the MIPv6, Sornlertlamvanich et al. [16] distributed the HA functions to each access router and set it as the default gateway of each attached mobile network. Besides, based on the PMIPv6, Do and Kim [17] combined the LMA and MAG functions into each access router and introduced a proxy router to manage MNs' mobility and set up a central session database to track MNs and the proxy router. Furthermore, Ernest et al. [18] improved these schemes by decomposing the logical functions of LMA to location management and resource management. It not only extends LMA to support the home network prefix allocation functions, but also extends MR to support the Delegating Router (DR) and Request Router (RR) functions. More recently, Nguyen and Bonnet proposed the H-NEMO solution which combines the centralized mobility management with the distributed mobility management [19]. This new solution assigns two prefixes for Mobile Anchor Router (MAR) called MAR-prefix and central mobility anchor called LMA-prefix and uses them for short-lived flow and long-lived flow, respectively. The DMM-based solutions can improve the scalability of the mobility scenarios with high session density; however, as for IoT devices, the session mobility is not always necessary.

2.2.4. ID/Locator Separation Based Schemes. Another solution is based on the splitting of identifier and location. Ishino et al. [29] proposed a routing-based mobility management which realized the routing aggregation by adopting the Bloom Filter to store routing information. This scheme separated the identifier and locator and provided the identifier routing based on Bloom Filter to support the mobility. Besides, Kim et al. [20] also suggested adopting the ID/locator separation to provide the network mobility, in which MR sends an aggregate location update message on behalf of the whole mobile network to reduce the signaling cost and handover delay. This kind of solution requires large modification of the existing network architecture, which cannot be easily deployed.

2.3. Group Mobility Management for IoT. Group-based approaches have been proposed to improve the mobility. Based on the movement trace similarity, Fu et al. [5] proposed a group mobility management mechanism, in which the global location database divides the nodes into different groups. The group leader performs the mobility management on behalf of the other group members. Moreover, Galluccio et al. [31, 32] derived an analytical framework to assess the impact of object group mobility when adopting MIPv6.

Based on the above analysis, we find these solutions to have the following characters:

(1) IoT devices are power limited and should support the bulk operations to reduce the cost to reduce the redundancy operations.

(2) NEMO-based solution defines a group by mobile network prefix and provides the mobility support via mobile router.

(3) The group-based method is dynamic, and the group is temporarily set up to realize the aggregative registrations. In another word, the group-based method can be viewed as a general case of NEMO.

In this paper, we propose a group binding update scheme based on PMIPv6 to update a group of IoT devices to reduce the signaling cost and analyze its impact on signaling cost.

3. PMIPv6-Based Group Binding Update Solution

The proposed scheme is based on the PMIPv6 bulk binding update mechanism which is designed to optimize the binding update and revocation operations for a group of mobility sessions by introducing group identifier. The group identifier can be assigned by MAG or LMA and will be exchanged via Proxy Binding Update (PBU) and Proxy Binding Acknowledgment (PBA) messages with "*B*" flag and is finally recorded in the Binding Cache Entry (BCE) of LMA. Therefore, the bulk binding is generally used to extend the lifetimes of multiple mobility sessions and revoke all the sessions hosted on the failed service card.

The proposed scheme extends the bulk binding not for session groups but for the node groups and its basic idea is to set up a group for IoT devices based on some metrics such as the movement similarity [5] or administrative domain [6, 7] to perform the binding update in form of node groups. By this way, the signaling cost therefore will be reduced. To provide this group binding update, we have to extend the mobile node group identifier option, binding information on MAG and LMA, and design the group creation procedure of IoT devices.

3.1. Mobile Node Group Identifier Option Extensions. The mobile node group identifier option (MNGIO) defined in [22] is used to carry the group identifier. Figure 2 shows the extended MNGIO format.

(i) The type field is 50 which has been assign by IANA to represent that this is a mobile node group identifier option.

(ii) The length field is 6 bytes which exclude the type and length field.

(iii) The subtype field is 8-bit which has 256 types. While the values of 0 and 255 have been reserved, the value of 1 has been assigned to bulk binding update group. In our work, we introduce a new subtype called group binding update and temporarily set its value of 2 for IoT devices group binding update.

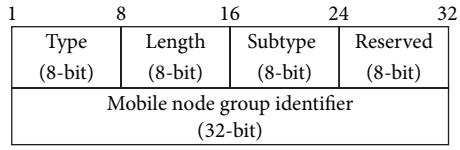
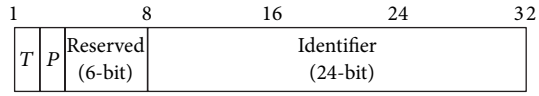


Figure 2: Mobile node group identifier option.



$T = 1$ indicates a group ID assigned by LMA
$T = 0$ indicates a group ID assigned by MAG
$P = 1$ indicates a permanently assigned ("well-known") group ID
$P = 0$ indicates a dynamically assigned group ID

Figure 3: Extended group identifier format.

(iv) The mobile node group identifier (noted as G-ID) is 32-bit, which can be assigned by MAG and LMA. The all 0 and all 1 are reserved.

Figure 3 shows the extended G-ID format which is 4 octets and divides into flag and identifier fields. The first 1 octet is set to different flags, in which we introduce two flags "*T*" and "*P*" and reserve 6 bits for future extensions.

"*T*" flag distinguishes the assigner of group ID. $T = 1$ indicates a group ID assigned by LMA (called LG-ID), and $T = 0$ indicates a group ID assigned by MAG (called MG-ID). More specifically, MG-ID represents the group of IoT devices that attaches to the same MAG, while LG-ID represents the group of IoT devices with the same HA or LMA. In the proposed scheme, LG-ID is predefined by HA or LMA to divide its IoT devices into different groups in advance, while MG-ID is used to record the attached IoT devices in the same access network. "*P*" flag indicates the well-known group ID. $P = 1$ indicates a permanently assigned group ID, while $P = 0$ indicates a transient or dynamical group ID. The following 3 octets identify different groups in which all zeros and all ones are revised.

3.2. MAG and LMA Operations Extensions. MAG extends its Binding Update List (BUL) to add the MG-ID and LG-ID fields, while LMA extends its BCE to include MG-ID and LG-ID. LMA predefines the groups of IoT devices and assigns a LG-ID for each group in advance, so that the IoT device in the same group will be marked by LG-ID in the BCE of LMA. MAG creates the groups of IoT devices dynamically according to the group policy (e.g., movement-based method [5]), assigns a MG-ID for each group, and records the MG-ID in the Binding Update List for each node in that group. LMA and MAG exchange LG-ID and MG-ID via PBU and PBA messages within the MNGIO. Based on (MG-ID, LG-ID) information, both MAG and LMA update their Binding Update List and binding cache, respectively. Similar to [22], the extensions of BCE and BUL use the MAG bulk binding update group ID to record the MG-ID received from MAG

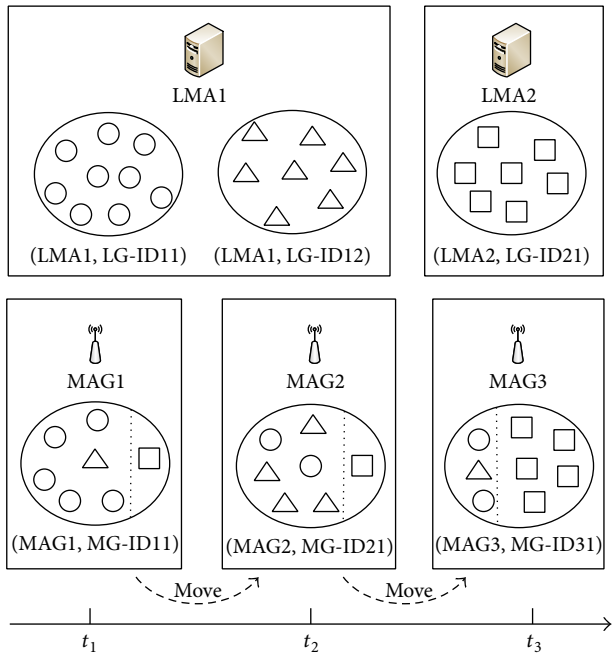


Figure 4: The IoT group creation procedure.

and use the LMA bulk binding update group ID to record the LG-ID received from LMA.

3.3. IoT Group Creation Procedure. We adopt Figure 4 to describe the IoT group creation procedure. Assume that there are three groups of IoT devices called LG-ID11, LG-ID12, and LG-ID21 initially. LG-ID11 and LG-ID12 are assigned by LMA1, and LG-ID21 is assigned by LMA2. Therefore, we can note them as (LMA1, LG-ID11), (LMA1, LG-ID12), and (LMA2, LG-ID21).

At t_1, several IoT devices attach to MAG1. Based on the movement trace character or other grouping policies, MAG1 sets up a group and assigns a group identifier MG-ID11. After exchanging the group information between MAG and LMA, the devices in MG-ID1 can be further divided into multiple subgroups based on their LMAs. As shown in Figure 4, the MG-ID11 group can be divided into two parts. The left part is the subgroup of devices belonging to LMA1, while the right part belongs to LMA2. The binding update will be performed in form of a group with same (MG-ID, LG-ID). To this end, the devices in the same subgroup belonging to LMA1 will only perform once registration with their LMA, while the other IoT devices in the group have to perform multiple registrations with their LMAs, respectively. By this way, the signaling cost can be greatly reduced especially when the devices' density is high. Once these IoT devices move into another access network, a new group will be created and perform similar procedure as t_1. By setting up the group, the IoT devices belonging to the same LMA will only update once with their LMAs, and therefore the signaling cost will be reduced.

3.4. Mobility Management Procedure. Assume that the LG-IDs are assigned by LMA in advance and stored in LMA's BCEs. Figure 5 shows the signaling flow of group binding update when the IoT devices enter the PMIPv6 domain, and its detailed steps are shown as follows.

(1) In the beginning, MAG1 detects the attachment events of IoT devices to acquire MN-IDs and their profiles. After that, MAG1 divides the attached IoT devices into different groups based on the grouping policy such as movement similarity and assigns the MG-IDs for these groups. MAG1 records the group members of each IoT group and updates the related BUL with MG-ID for each group member.

(2) IoT device (noted as MN1) belonging to MG-ID1 sends Router Solicitation message (noted as RS1) to MAG1 at any time after it attached to MAG1.

(3) After receiving RS1, MAG1 sends PBU message with "*B*" flag to MN1's LMA (noted as LMA1). The PBU message carries the MN1's ID (MN-ID1) and group ID (MG-ID1).

(4) Once LMA1 receives this PBU message, it will update the related BCE based on the MN-ID and update the MG-ID field, and then it will reply with a PBA message with "*B*" flag in which MN1's LG-ID1 is carried in the MNGIO field.

(5) Once MAG1 receives PBA, it will update the related BUL with the LG-ID1. In this way, the group information is exchanged.

(6) In a similar way, other IoT devices (such as MN2 and MN3) perform a similar binding update procedure.

(7) After finishing the initiation procedure, the MAG will perform the group binding update and update the group binding relationship of MG-ID1. For all the nodes with the same LMA, it only performs one binding update procedure.

(8) In a similar operation, MAG2 and MAG3 will perform the same procedure as MAG1. By this way, the IoT devices in the same group such as MG-ID and LG-ID will only perform once, and the overall binding update cost will be reduced.

4. Model and Analysis

We consider an IoT application scenario which consists of several IoT devices with different access technologies. Based on [5], we assume that an IoT device joins the group following a Poisson process with λ, and the dwell time of this IoT device in this group follows a general distribution with mean $1/\mu$. Based on the $M/G/\infty$ queuing model, the steady-state probability for the number M of IoT devices in this group MG-ID1 is

$$P_r[M = x] = \frac{\alpha^x e^{-\alpha}}{x!}, \tag{1}$$

where $\alpha = E[M] = \lambda/\mu$.

We denote by C_Δ as unit binding update cost of mobility management protocol and by C_o as the binding update cost of original mobility management protocol.

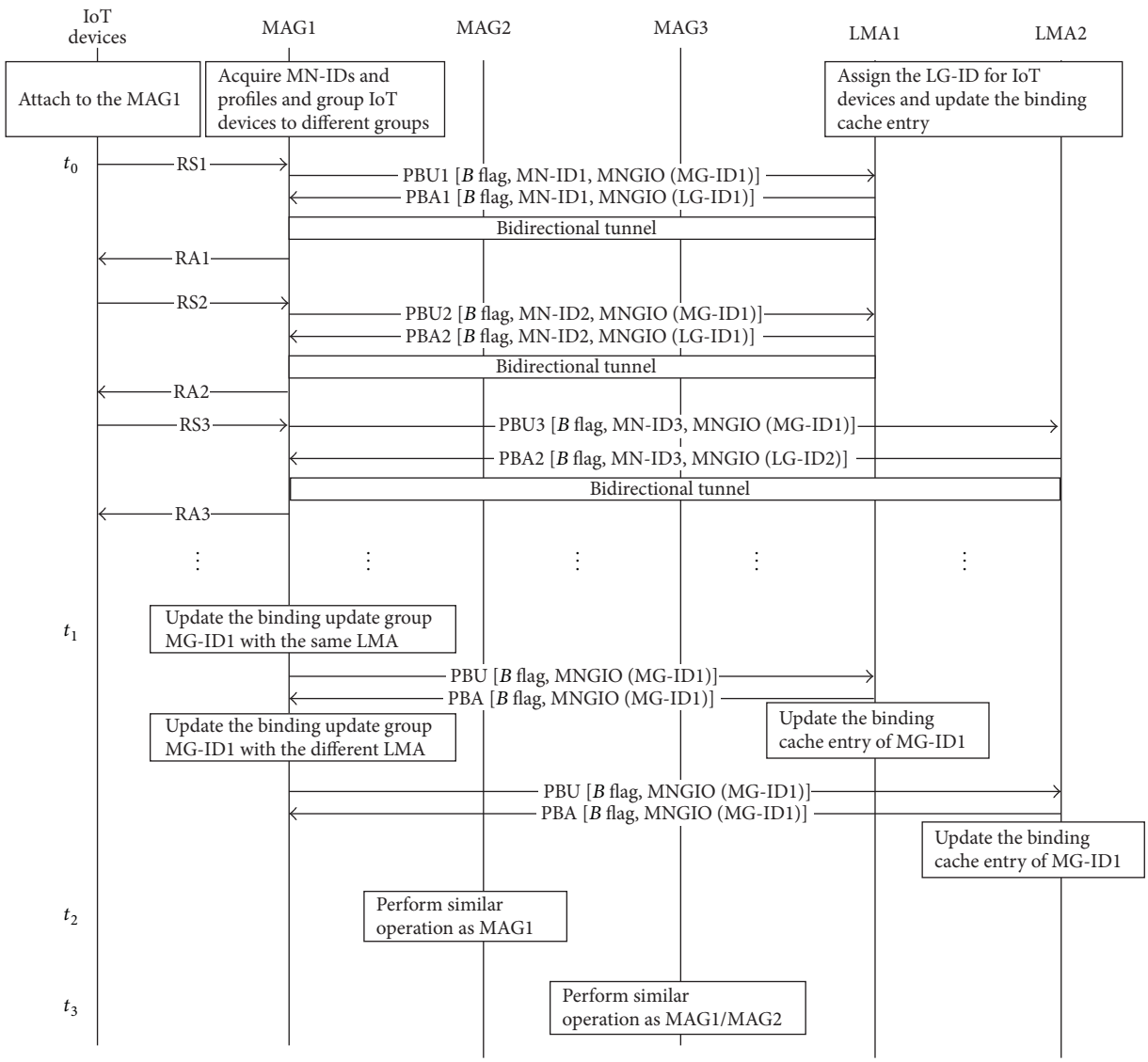


Figure 5: Group binding update operations for IoT devices.

To simplify the analysis, we divide the MG-ID into two subgroups and denote by C_1 as the binding update cost of subgroup in MG-ID with the same LMA and by C_2 as the binding update cost of IoT devices in MG-ID with different LMAs. Suppose that the probability of IoT devices located in the given group with the same LMA is ρ; we can get the cost of basic PMIPv6 and group binding as follows.

(1) Basic Binding Update Cost. For basic PMIPv6, the total update cost of x nodes is shown as follows:

$$C_{o|M=x} = xC_{\Delta}. \tag{2}$$

Then, we can get its expected value as follows:

$$E\left[C_o\right] = \sum_{x=1}^{\infty} xC_{\Delta}P_r\left[M = x\right] = E\left[M\right]C_{\Delta}. \tag{3}$$

(2) Group Binding Update Cost. For the IoT devices in the group with the same LMA, they only perform once binding update, so the cost of group mobility management cost can be expressed as

$$C = \rho C_1 + \left(1 - \rho\right)C_2. \tag{4}$$

Then, the expected value of C is

$$\begin{aligned} E\left(C\right) &= E\left(\rho C_1 + \left(1 - \rho\right)C_2\right) \\ &= \rho E\left(C_1\right) + \left(1 - \rho\right)E\left(C_2\right). \end{aligned} \tag{5}$$

Next, we can get the

$$\begin{aligned} E\left[C_1\right] &= \sum_{x=1}^{\infty} C_{\Delta}P_r\left[M = x\right] = C_{\Delta}\left(1 - P_r\left[M = 0\right]\right) \\ &= C_{\Delta}\left(1 - e^{-\alpha}\right). \end{aligned} \tag{6}$$

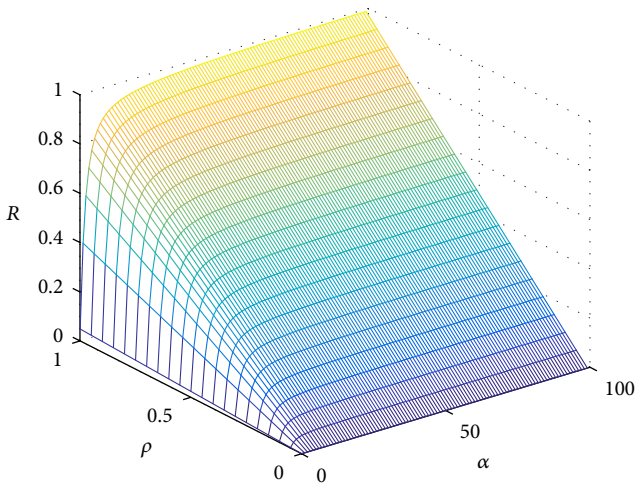


Figure 6: The reduction ratio of signaling cost.

As for the IoT devices in the group with different LMAs, they have to perform multiple binding update procedures, which can be expressed as follows:

$$E[C_2] = \sum_{x=1}^{\infty} xC_\Delta P_r[M = x] = C_\Delta E[M]. \quad (7)$$

Based on (6) and (7), the group-based signaling cost is

$$E(C) = \rho\left(1 - e^{-\alpha}\right)C_\Delta + (1-\rho)\alpha C_\Delta. \quad (8)$$

To evaluate the performance of group binding update, we adopt the cost reduction ratio, which is shown as

$$R = \frac{E(C_o) - E(C)}{E(C_o)} = \rho\left(\frac{e^{-\alpha}}{\alpha} - \frac{1}{\alpha} + 1\right). \quad (9)$$

Figure 6 shows the reduction ratio according to (9), and Figure 7 shows the signaling cost reduction ratio under different group sizes and probabilities. We can get that, with the increase of the group size, the R is increased and approaches to the ρ. With the increase of the ρ, the R is also increased. These results show that with the increase of number of IoT devices with the same LMA, the signaling cost will be reduced greatly.

5. Conclusions

To provide effective mobility management for large-scale IoT devices, new behavior features should be taken into consideration. In this paper, we have studied the IoT mobility management challenges and proposed a group binding updating scheme based on the PMIPv6 for making the best use of the group character of IoT. In particular, we extend the PMIPv6 by introducing the group identifier to support the group binding to reduce the signaling cost, and a complete operation flow is described in detail. The analytical results show that our proposed group binding can reduce the signaling cost and the group impact. The further work is to improve the analytical model and derive it in more complex network conditions.

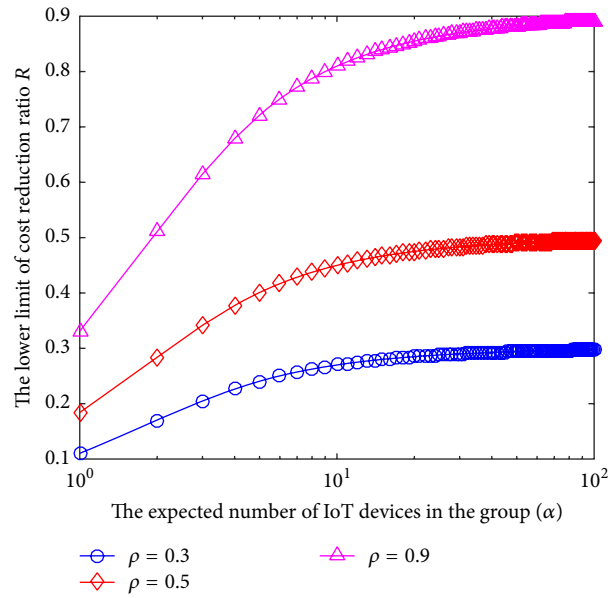


Figure 7: Impact of group size on the cost reduction.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was partially supported by the National Basic Research Program of China (973 Program) under Grant no. 2013CB329102, and also in part supported by Soonchunhyang University Research Funding.

References

[1] D. Evans, "The internet of things: how the next evolution of the internet is changing everything," Cisco IBSG-White Paper, 2011.

[2] H. Zhang, P. Dong, W. Quan, and B. Hu, "Promoting efficient communications for high speed railway using smart collaborative networking," *IEEE Wireless Communications*, vol. 22, no. 6, pp. 92–97, 2015.

[3] T. Savolainen, J. Soininen, and B. Silverajan, "IPv6 addressing strategies for IoT," *IEEE Sensors Journal*, vol. 13, no. 10, pp. 3511–3519, 2013.

[4] C. Perkins, D. Johnson, and J. Arkko, "Mobility support in IPv6," IETF RFC 6275, 2011.

[5] H.-L. Fu, P. Lin, H. Yue, G.-M. Huang, and C.-P. Lee, "Group mobility management for large-scale machine-to-machine mobile networking," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 3, pp. 1296–1305, 2014.

[6] X. Jiang and D. H. C. Du, "BUS-VANET: a BUS vehicular network integrated with traffic infrastructure," *IEEE Intelligent Transportation Systems Magazine*, vol. 7, no. 2, pp. 47–57, 2015.

[7] X. Jiang, X. Cao, and D. H. C. Du, "Multihop transmission and retransmission measurement of real-time video streaming over DSRC devices," in *Proceedings of the 15th IEEE International Symposium on a World of Wireless, Mobile and Multimedia*

Networks (WoWMoM '14), pp. 1–9, IEEE, Sydney, Australia, June 2014.

[8] L. Zhong, F. Liu, X. Wang, and Y. Ji, "Fast handover scheme for supporting network mobility in IEEE 802.16e BWA system," in *Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '07)*, pp. 1757–1760, Shanghai, China, September 2007.

[9] B. A. Mohammed and T. C. Wan, "Modified fast-integrated lightNEMOv6 handoff in IEEE 802.16e BWA networks," in *Proceedings of the 2nd International Conference on Network Applications, Protocols and Services (NETAPPS '10)*, pp. 182–187, Alor Setar, Malaysia, September 2010.

[10] S. Ryu, K.-J. Park, and J.-W. Choi, "Enhanced fast handover for network mobility in intelligent transportation systems," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 1, pp. 357–371, 2014.

[11] M. Liebsch, A. Muhanna, and O. Blume, "Transient binding for proxy mobile IPv6," RFC 6058, 2011.

[12] I. Soto, C. J. Bernardos, M. Calderon, A. Banchs, and A. Azcorra, "NEMO-enabled localized mobility support for internet access in automotive scenarios," *IEEE Communications Magazine*, vol. 47, no. 5, pp. 152–159, 2009.

[13] S. Jeon and Y. Kim, "Cost-efficient network mobility scheme over proxy mobile IPv6 network," *IET Communications*, vol. 5, no. 18, pp. 2656–2661, 2011.

[14] J.-H. Lee, T. Ernst, and N. Chilamkurti, "Performance analysis of PMIPv6-based network mobility for intelligent transportation systems," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 1, pp. 74–85, 2012.

[15] M.-S. Kim and S. Lee, "Group-based fast handover for PMIPv6-based network mobility in vehicular networks," in *Proceedings of the IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS '15)*, pp. 113–114, IEEE, Hong Kong, April-May 2015.

[16] P. Sornlertlamvanich, S. Kamolphiwong, R. Elz, and P. Pongpaibool, "NEMO-based distributed mobility management," in *Proceedings of the 26th IEEE International Conference on Advanced Information Networking and Applications Workshops (WAINA '12)*, pp. 645–650, IEEE, Fukuoka, Japan, March 2012.

[17] T.-X. Do and Y. Kim, "Distributed network mobility management," in *Proceedings of the 5th International Conference on Advanced Technologies for Communications (ATC '12)*, pp. 319–322, IEEE, Hanoi, Vietnam, October 2012.

[18] P. P. Ernest, H. A. Chan, O. E. Falowo, L. A. Magagula, and S. Céspedes, "Network-based distributed mobility management for network mobility," in *Proceedings of the IEEE 11th Consumer Communications and Networking Conference (CCNC '14)*, pp. 417–425, IEEE, Las Vegas, Nev, USA, January 2014.

[19] T.-T. Nguyen and C. Bonnet, "A hybrid centralized-distributed mobility management architecture for network mobility," in *Proceedings of the IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WOWMOM '15)*, pp. 1–9, Boston, Mass, USA, June 2015.

[20] Y. Kim, H. Ko, and S. Pack, "Network mobility support in distributed ID/locator separation architectures," in *Proceeding of IEEE 11th Consumer Communications and Networking Conference (CCNC '14)*, pp. 521–523, Las Vegas, Nev, USA, January 2014.

[21] V. Devarapalli, R. Wakikawa, A. Petrescu, and P. Thubert, "Network mobility (NEMO) basic support protocol," IETF RFC 3963, 2005.

[22] F. Abinader, S. Gundavelli, K. Leung, S. Krishnan, and D. Premec, "Bulk binding update support for proxy mobile IPv6," RFC 6602, 2012.

[23] H. Soliman, C. Castelluccia, K. El Malki, and L. Bellier, "Hierarchical mobile IPv6 mobility management (HMIPv6)," RFC 4140, 2005.

[24] R. Koodli, Ed., *Fast Handover for Mobile IPv6*, RFC 4068, 2005.

[25] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury, and B. Patil, "Proxy mobile IPv6," IETF RFC 5213, 2008.

[26] J. Guan, H. Zhou, W. Xiao, Z. Yan, Y. Qin, and H. Zhang, "Implementation and analysis of network-based mobility management protocol in WLAN environments," in *Proceedings of the International Conference on Mobile Technology, Applications, and Systems (Mobility '08)*, ACM, Yilan, Taiwan, September 2008.

[27] J. Guan, H. Zhou, Z. Yan, Y. Qin, and H. Zhang, "Implementation and analysis of proxy MIPv6," *Wireless Communications and Mobile Computing*, vol. 11, no. 4, pp. 477–490, 2011.

[28] H. Yokota, K. Chowdhury, R. Koodli, B. Patil, and F. Xia, "Fast handovers for proxy mobile IPv6," RFC 5949, 2010, http://tools.ietf.org/html/rfc5949.

[29] M. Ishino, Y. Koizumi, and T. Hasegawa, "A study on a routing-based mobility management architecture for IoT devices," in *Proceedings of the 22nd IEEE International Conference on Network Protocols (ICNP '14)*, pp. 498–500, IEEE, Raleigh, NC, USA, October 2014.

[30] 3rd Generation Partnership Project (3GPP), "3rd Generation Partnership Project; technical specification group services and system aspects; service requirements for Machine-Type Communications (MTC); stage 1 (Release 11)," Tech. Rep. 3G TS 22.368, Sophia-Antipolis Cedex, Paris, France, 2011.

[31] L. Galluccio, G. Morabito, and S. Palazzo, "On the potentials of object group localization in the internet of things," in *Proceedings of the IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM '11)*, Lucca, Italy, June 2011.

[32] S. D'Oro, L. Galluccio, G. Morabito, and S. Palazzo, "Exploiting object group localization in the internet of things: performance analysis," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 8, pp. 3645–3656, 2015.

Hindawi Publishing Corporation
Mobile Information Systems
Volume 2016, Article ID 9684032, 9 pages
http://dx.doi.org/10.1155/2016/9684032

Research Article

Preallocated Duplicate Name Prefix Detection Mechanism Using Naming Pool in CCN Based Mobile IoT Networks

Juyong Lee and Jihoon Lee

Department of Information and Telecommunication Engineering, Sangmyung University, Cheonan, Republic of Korea

Correspondence should be addressed to Jihoon Lee; vincent@smu.ac.kr

Received 7 August 2015; Revised 24 December 2015; Accepted 29 December 2015

Academic Editor: Kamal Deep Singh

As the number of mobile devices and IoT (internet of things) devices has explosively increased, various contents are created anytime and anywhere. To meet such trend, current internet architecture has exposed many limitations such as high control overhead due to multistaged address resolution, frequent location updates, and network congestion. Recently, information centric networking (ICN) is considered as new networking architecture to redesign current internet's content exchange paradigm. In current ICN architecture, whenever a mobile node moves to a new domain, it needs long latency to configure and confirm the temporary name prefix. So, this paper presents an efficient name prefix configuration mechanism in mobile CCN to reduce the latency needed for the name prefix configuration during MCS's handover. From the performance analysis, the proposed mechanism is shown to provide lower control overhead and lower resource consumption.

1. Introduction

As the volumes of various internet of things (IoT) devices and mobile devices have rapidly grown, the device movements have become the key research point. Mobile devices may link to information about IoT devices or may transmit real-time sensing data from them anytime and anywhere. It is expected that global mobile data traffic will increase 18 times from 2011 to 2016 [1]. Through such trend, mobile-connected devices lead to huge volume of data traffic and network resource consumption. However, the current internet architecture based on the host-based model cannot keep pace with such trend due to various additional operations (i.e., frequent location updates, multilevel address resolution, etc.).

To solve such limitations, many researchers have paid attention to the new networking architecture like information centric networking (ICN), at which content is accessed by a content name itself instead of IP address for the node having the content. There are typical examples for ICN: DONA [2], 4WARD [3], NetInf [4], PSIRP [5], and CCN [6]. Among these studies, this paper focuses on the content-centric networking (CCN) architecture because it is regarded as efficient networking architecture for content sharing [7].

However, there has been little consideration for the effects due to device movement. In CCN, content consumer mobility can be handled well because of CCN's inherent receiver-driven natures, at which there are no needs for location updates. That is, after content consumer moves into another domain, it just retransmits interest packets relevant to data packets that have not been downloaded yet. Meanwhile, the content source mobility in CCN causes many problems such as frequent routing updates, repeated transmission of interest packets, and long content download time because it requires fully updating routing tables of all content routers.

To overcome these problems, some solutions have been proposed. The tunnel-based approach [8, 9] is based on MIPv6. The locator/ID-based approach [10, 11] uses the concept of location and identifier separation to support the provider mobility in the ICN. The forwarding information base (FIB) update-based approach [12, 13] is also based on the MIPv6 concept, but instead of using a tunnel to redirect interest packets to the new location of the content source, it uses the FIB updates to redirect interest packets. However, these approaches still remain inefficient in terms of handover latency and high interest packet drop rate due to long latency for the new name prefix configuration. That is, such redirection schemes need the new name prefix to consistently receive

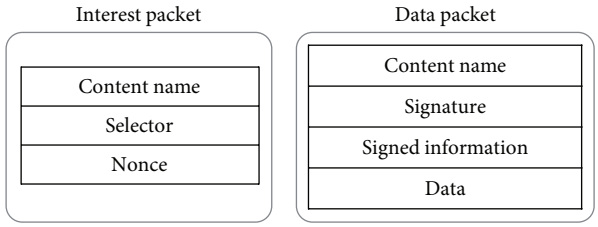


Figure 1: Interest packet and data packet.

the interest packets from content consumers. However, current CCN architecture assumes interest-data exchange model for the new name configuration and confirmation process to guarantee the name uniqueness, which results in long latency (i.e., over 5 seconds). So, fast duplicate name prefix detection (fDND) mechanism is presented to reduce the latency for the new name configuration [14]. But, it also requires over second unit for the new name configuration. So, this paper proposes an efficient preallocated name configuration mechanism using a naming pool architecture to reduce long handover latency of mobile content sources. The proposed scheme does not need any additional latency to check the name uniqueness by using a stateful name pool architecture.

The rest of the paper is organized as follows. Section 2 briefly describes CCN architecture and then shows what happens while mobile content source moves into another domain. Section 3 presents the design and detailed operations of the proposed mechanism. Then we present the evaluation results and then make a conclusion.

2. Mobility Management in CCN

2.1. Basic CCN. Basically, CCN is composed of two packet types, interest and data (Figure 1). Interest packet acts as a query for content. Data packet is utilized to carry the actual content. Interest packet contains a unique identifier (content name), a set of parameters such as the order preference (selector), and a random nonce value to prevent the packet from looping and so on. Each name prefix has a hierarchical structure, and "/" character represents delimiter between different components (i.e., "*/smu.ac.kr/jyl/doc/pDND.txt*").

Content name is utilized for content identifier and routing lookup.

Each CCN node basically has three functional blocks for packet forwarding: content store (CS), pending interest table (PIT), and forwarding information base (FIB). The FIB is utilized to forward interest packets toward potential content holder(s) with matching data. It is almost similar to an FIB table of existing IP routers, at which the critical difference is that the object identifier is the content name itself, not IP address. The CS has the role as the content cache. It stores data packets to be used in future by other content requesters. So, CCN can provide localized transmission near content requesters. The PIT keeps track of interests forwarded toward content source(s) so that returned data can be sent back to its requester(s). PIT entries are eliminated as soon as they have been consumed to forward a matching data packet. PIT entries for interests that never find matching data are eventually timed out and naturally eliminated.

The content retrieval and forwarding procedures are depicted in Figure 2. When an interest packet arrives at Face 0, a content store lookup is conducted based on its content name. If matched data is found in CS, the data packet is delivered. Otherwise, it searches a PIT entry to check whether it has already received another request for the same content. If a matched entry is found in the PIT, the CCN node adds into the existing entry the face on which the new interest packet arrived. When the data packet arrives at any node, it is replicated and sent out on all faces in the PIT entry for the content. If the content name in interest packet does not match either the CS or the PIT, the FIB is looked up to determine the outgoing face where the interest packet should be forwarded. Furthermore, the PIT adds one entry for the forwarded interest packet. In other words, data packet just simply follows the order of PIT entries back to the content requester.

2.2. The Relevance of CCN with Mobile IoT. In IoT world, numerous heterogeneous devices are interconnected and transfer data to each other over a network [15–18]. So, large quantities of heterogeneous data that need to be processed in real time are dynamically generated. Also, the smart devices with various sensors (i.e., accelerometer, GPS, gyroscope, magnetometer, etc.) generate lots of information keeping in motion. Therefore, additional networking and service mechanisms are required to handle the huge volume of various kinds of heterogeneous information from IoT devices. Thus, inherent characteristics of CCN like hierarchical content name prefix and cache-and-forward architecture can efficiently handle the relevant traffic and service cases in IoT environment.

However, in mobile environment, it hardly appears possible that mobile nodes know the exact name of the content data that it needs when interacting with local networks. Content name prefix in CCN environment is configured as hierarchical structure with various components and application-relevant naming. It facilitates more rapid content lookup. Content consumers do not care about where and how to obtain a piece of content data. So, to keep continuity of content delivery, new CCN naming configuration mechanism is required to handle various mobile IoT environment.

2.3. The Problems due to Content Source Movements. Although CCN architecture is efficient for content sharing, some problems may take place to support mobile content sources. That is, as content consumer does not know the content source movement, mobile CCN requires an update of the FIB entry to forward interest packet towards the current location of mobile content source. It takes much time to update the FIB entry of all content routers and causes too many dynamic routing updates. So, interest packets may not be delivered due to long latency when the route to related content source is changed.

Redirection based schemes like TBR [8, 9] are utilized to redirect interest packets and data packets between MCS's home domain and the content router in the moved domain.

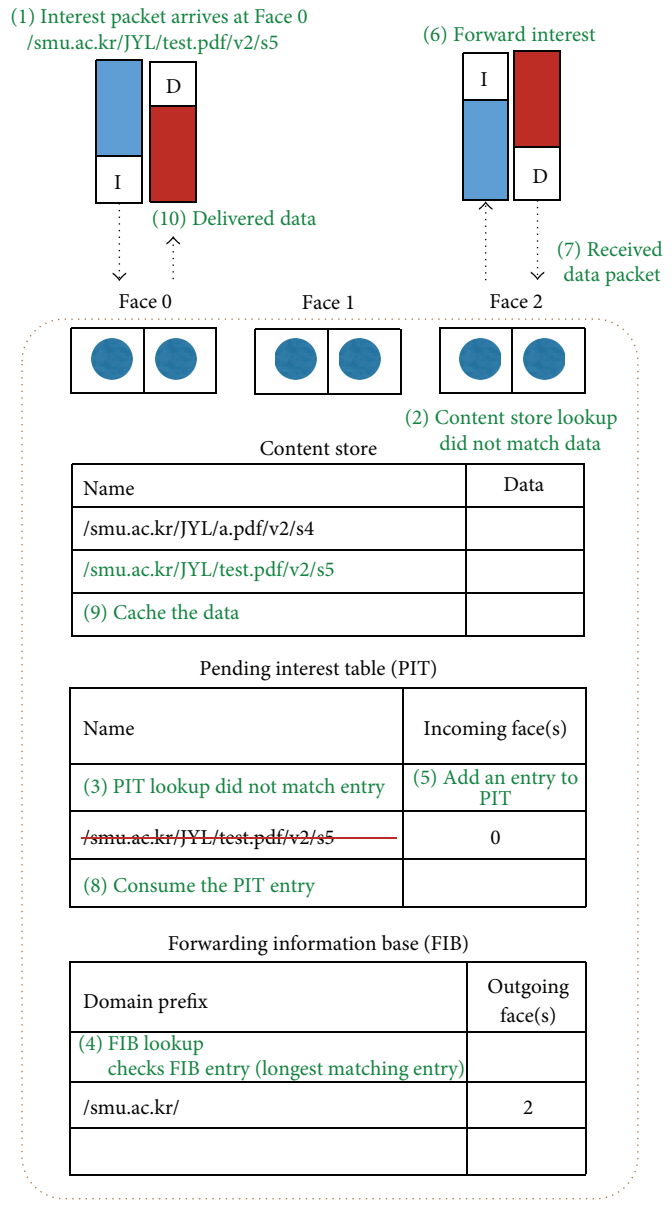


Figure 2: CCN forwarding engine.

However, existing redirection based schemes just consider normal interest-data sharing based name configuration and confirmation procedure, at which they do not care about the long latency required to guarantee name uniqueness. In the IoT environment where up-to-date information especially is very important, long handover latency incurs unnecessary deliveries of content request and provides bad experience to users.

3. The New Preallocated Duplicate Name Detection Mechanism

The main point of the proposed mechanism is to use a preallocated prefix from naming pool to provide fast and seamless content source mobility in CCN environments in order to reduce long handover latency.

The proposed mechanism is based on the redirection model to handle content source mobility, which redirects incoming interest packets towards MCS from the MCS's home content router (CR_h) to content router (CR) in the currently located domain. For that, it is assumed that CRs periodically broadcast their information within their service range in order to decide whether MCS sends the "name request (NR)" message to configure a new tentative name prefix. Through the preallocation of tentative name prefix from new content router (CR_n), the proposed scheme reduces the handover latency due to new name configuration and confirmation procedures. Thus, it provides MCS's seamless handover and low network resource consumption.

CR_n	NR	CS prefix	Tunnel	Current tentative prefix

Example of the name request message

c.com	*NR*	*smu.ac.kr/JYL*	*Tunnel*	*smu.ac.kr/JYL*

Figure 3: Name request message.

CR_n	NR	CS prefix	Tunnel	Current tentative prefix	NRA	New tentative prefix

Example of the name request acknowledgement message

c.com	*NR*	*smu.ac.kr/JYL*	*Tunnel*	*smu.ac.kr/JYL*	*NRA*	*c.com/pre01/smu.ac.kr/JYL/*

Figure 4: Name request acknowledgement message.

The proposed preallocated duplicate name prefix detection (pDND) mechanism starts when a MCS sends a NR message to CR_n to request tentative name prefix. It is assumed that each CR creates and manages a unique naming pool. CRs randomly generate globally routable name prefixes as background process. On generating a new name prefix, CR performs DND on the name prefix within its domain area. Only after successful DND does CR reserve the name prefix into its naming pool and regard the reserved name prefix as a unique name prefix. Afterwards, CR when receiving NR message sends the "name request acknowledgement (NRA)" message toward MCS to provide the uniqueness of tentative name prefix from naming pool and the allocated tentative name prefix. Also, when receiving the NRA message, intermediate CRs manage a binding table that records the relevance between the MCS's prefix and tentative prefix. So, MCS's current content router (CR_p) and intermediate content routers can simultaneously deliver incoming interest packets towards the MCS to both CR_n and its serving range. The proposed pDND mechanism using naming pool is composed of 3 steps. The detailed operation procedures are as follows.

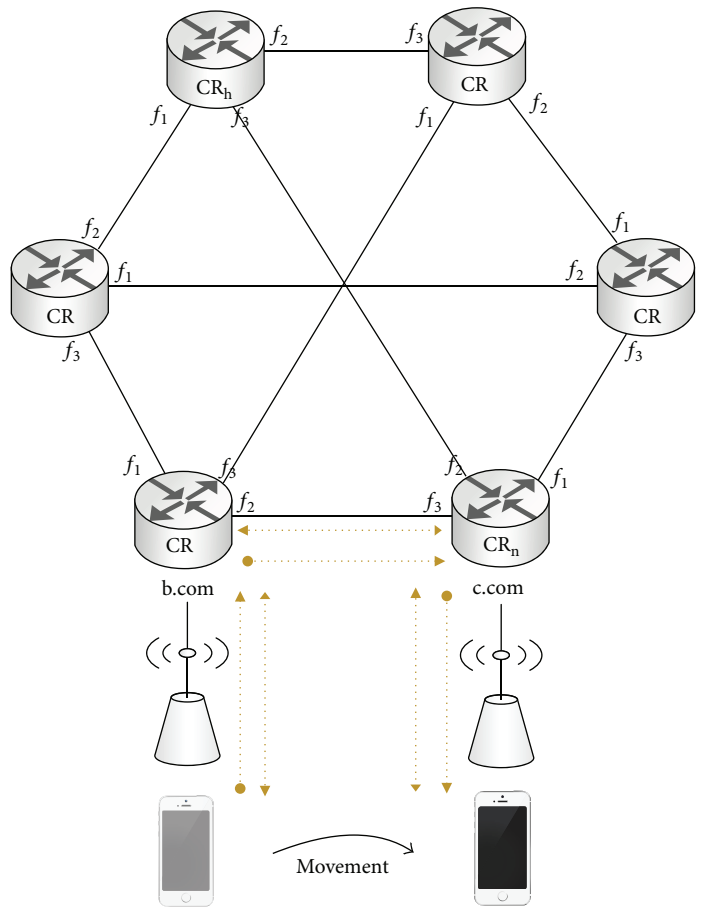
Step 1 (name configuration). To perform the pDND mechanism, it is assumed that the initialization data like signal strength or physical link information or network domain name containing a collection of network nodes is provided to MCS by the wireless access point. Through the report of the received signal strength, the MCS decides whether it initiates the transmission of the NR message towards CR_n to ask for a new tentative prefix. If it is imminent to move into another network domain, the MCS asks for the new name prefix to the content router in CR_n by sending NR message. The new NR message format is shown in Figure 3. CR_n field indicates the target location name prefix to route the NR message. For the packet identifier, *NR* field is designed. *CS prefix* field is utilized to keep track of the MCS's prefix. Finally, *current tentative prefix* is to indicate MCS's current tentative name prefix to identify the binding between current tentative prefix and new tentative prefix at intermediate CRs.

The content router in the new domain (CR_n) receiving NR message sends the NRA message toward MCS's current domain to guarantee the uniqueness of tentative name prefix from naming pool and the allocated new name prefix. Also, intermediate content router (CR_i) receiving the NRA message manages a binding table that records the MCS's home prefix and two tentative prefixes (i.e., current tentative prefix and new allocated tentative prefix). Thus, all interest packets toward the MCS can be delivered to CR_n through the preallocated name prefix routing information. The message format of NRA packet is shown in Figure 4.

After a successful name configuration process, MCS continuously checks the current location from initialization data by the wireless access point. If the movement into another domain network is detected, the MCS sends a PU message indicating its new location to CR_h. That is, the PU message includes the tentative name prefix as well as MCS's signature to validate the PU originator.

The process of the name configuration is shown in Figure 5. When the MCS detects when the movement is imminent, the MCS sends the NR message to the c.com to receive a new tentative prefix. The content router in c.com domain receiving NR message sends the NRA message toward *b.com* to provide the uniqueness of tentative name prefix from the naming pool and the allocated new name prefix. Also, b.com receiving the NRA message manages a binding table that records the MCS home prefix and two tentative prefixes. So, b.com can simultaneously deliver interests to both b.com and c.com. Thus, all interest packets toward the MCS can be delivered to CR_n through the preallocated name prefix routing information.

Step 2 (path redirection). The CR receiving a PU message compares its name prefix with that of the PU message. On receiving the PU message, CR_n checks the name prefix information from the received PU message and then looks up its routing table about the entry for CR_h. Each intermediate CR sends the PU message toward CR_h based on its FIB reference if its domain prefix is different from that of the



<Binding table at b.com>

Before NR reception

MCS prefix	Tentative prefix
smu.ac.kr/JYL	b.com/pre05/smu.ac.kr/JYL

After NR reception

MCS prefix	Tentative prefix
smu.ac.kr/JYL	b.com/pre05/smu.ac.kr/JYL
	c.com/pre07/smu.ac.kr/JYL

Figure 5: Example of name configuration.

PU message. Otherwise, it does not forward the PU message and then sends a PUA message to the MCS. That is, when receiving the PU message, the intermediate CRs identify the domain name prefix information from received PU message and check its routing table to decide whether the PU message is for its domain or not. Then the PIT entry is created with CR_h prefix in PU message. CR_h receiving PU message sends a PUA message indicating that the prefix update is successful. When receiving the PUA message, intermediate CRs look up the routing table again. If the relevant PIT entry exists that corresponds with the name prefix in the received PUA message, intermediate CR consumes the PIT entry and forwards the PUA message to the next hop.

Step 3 (interest redirection). After that, a content consumer requests content data generated by the MCS. The interest packets are delivered to CR_h. CR_h redirects incoming interest packets from the lookup of its routing table towards the domain where the MCS is currently located. In other words, CR_h generates new interest packet encapsulating received interest packet and then sends the encapsulated interest packet to MCS's current location. CR_i between CR_h

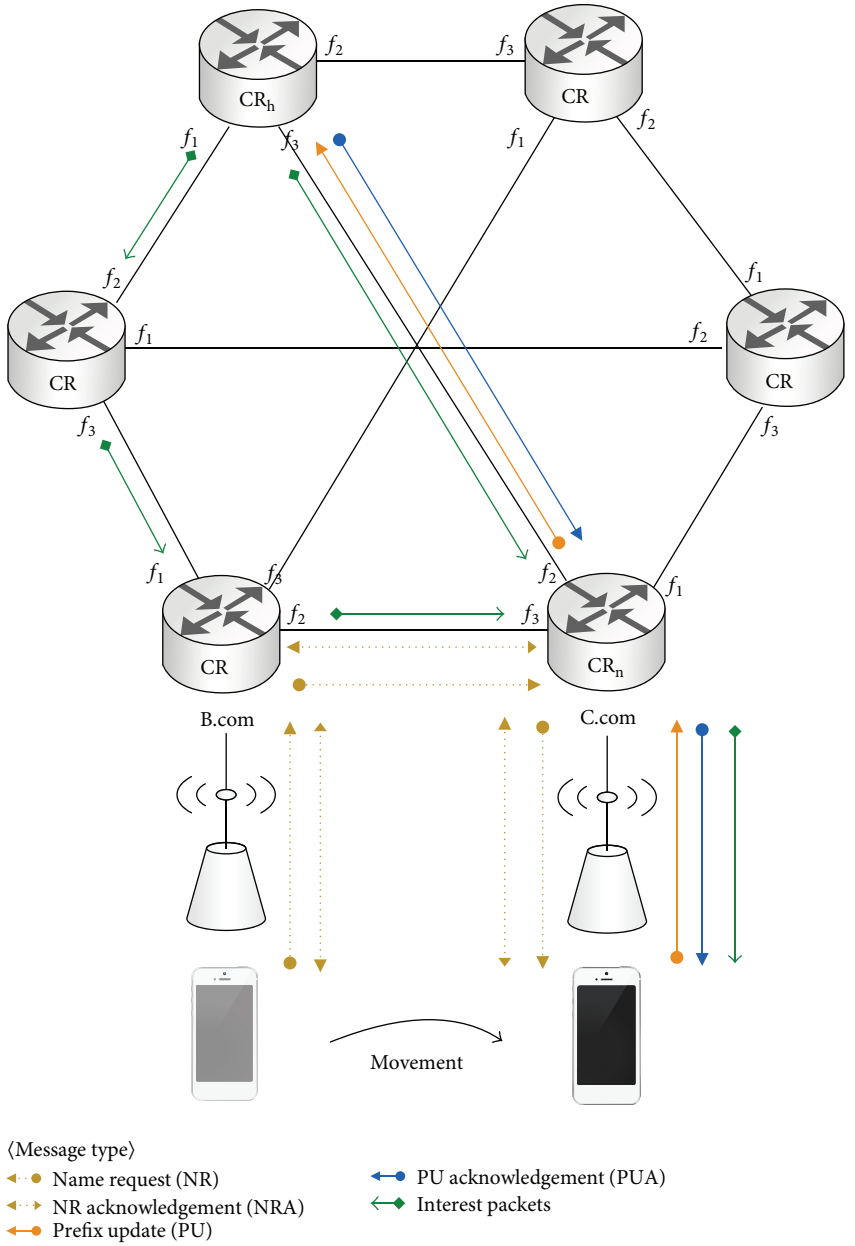


Figure 6: Example of the pDND mechanism.

and moved domain sends the encapsulated interest packet according to the FIB reference. After receiving the encapsulated interest packet, the MCS decapsulates it. The MCS transmits the data packet in response to the interest packet to CR_h. Intermediate CRs decapsulate the tunnel header and then do content caching with the original name prefix of the requested content object. Therefore, the proposed scheme can keep localizing traffic delivery through CCN's content caching.

Figure 6 shows the simplified example of the proposed pDND operation. It shows that the new name prefix allocation is done through the name configuration procedure to new content router and then contents are shared by encapsulating and decapsulating both interest and data packets at CR_h and content routers at foreign domain (CR_p and CR_n).

4. Performance Evaluation

To evaluate the performance of the proposed pDND, a series of simulations are conducted through discrete event simulator based on the named data networking (NDN) module in NS-3 [19, 20]. However, as the current NS-3 NDN module does not support a device mobility management functionality, a tunnel-based mobility management model is supplemented consisting of various content routers (i.e., 10 to 200 content routers), in which each content router is connected to wireless router. The initial location of content consumers is randomly selected at the considered network topology and MCS frequently changes its location. That is, when a MCS moves between CR_i and CR_n, it makes handover

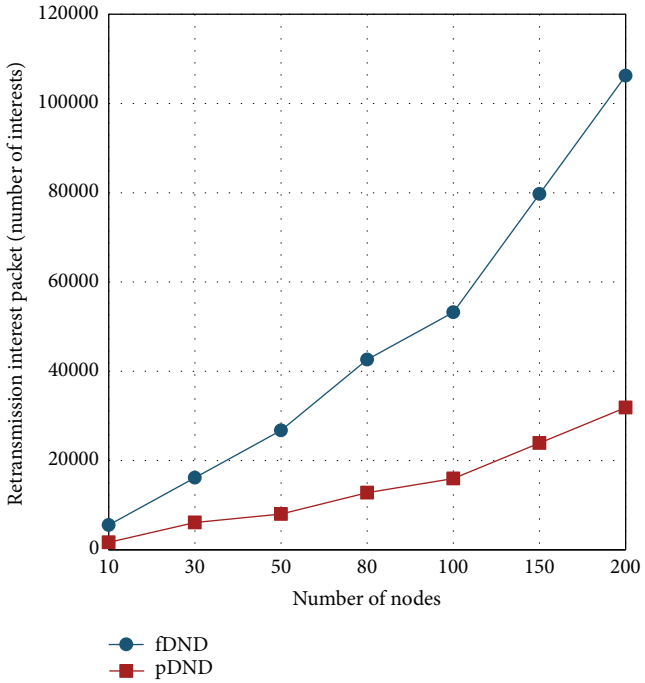


Figure 7: The number of interest message overhead Message.

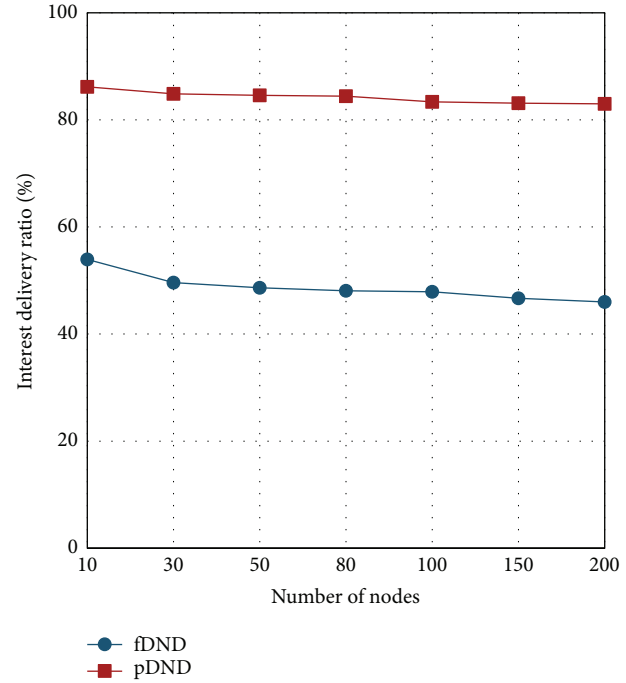


Figure 8: Interest delivery ratio at the mobile content source.

between content routers. As the simulation proceeds, an MCS moves according to the modified vehicle-to-vehicle (v2v) mobility model [21–23] because it makes user movement faster, smoother, and more realistic than other models. To evaluate the impact of the movements for the MCS, we check the interest retransmission packets, delivery ratio at the MCS, and content retrieval time against varying network sizes through the comparison with existing fDND mechanism.

Figure 7 shows the comparison results of the proposed pDND with the fDND in terms of the interest message overhead against varying network nodes. Interest overhead means total interest packets that include the number of original interests and retransmitted interests, which is presented to indicate the waste of network resource consumption. As the network topology size (i.e., the number of content routers) increases, the efficiency of existing fDND degrades rapidly when compared to the proposed pDND. As fDND requires uniform second unit latency for duplicate name detection whenever a movement into another domain happens, it leads to high drop rate of interest packets during MCS handover. So, it causes large number of retransmitted interest packets. Moreover, as the number of network size increases, the delivery latency of interest packets is also increased, which results in the increase of interest retransmission as well as high network resource consumption. Meanwhile, the proposed pDND scheme does not require any latency for DND operation because it conducts the new name configuration and confirmation before a handover event happens. So, the proposed mechanism leads to shorter handover latency and lower transmission of interest packets including retransmitted interest packets.

The interest packet delivery ratio is presented to indicate the efficiency of the name configuration mechanism in mobile CCN environment as shown in Figure 8. The delivery ratio of interest packet is the ratio of the number of interest packets received by MCS over the total interest packets (including original and retransmitted interest packets) sent by all content consumers. As the network size increases, the interest delivery ratio of fDND is decreased when compared to the pDND. Because the proposed pDND mechanism uses a preallocated naming pool architecture for tentative name prefix, it does not generate any additional latency for a DND operation, which leads to reduced delivery latency as well as reduced drop rate. However, fDND assumes a uniform waiting latency so it causes high drop rate of interest packets. Moreover, as the number of network nodes increases, the successful rate for interest packets in fDND becomes lower. It leads to the fact that the proposed pDND mechanism supports lower network resource consumption than fDND.

Figure 9 shows the contents retrieval time. It is shown to indicate how fast the proposed mechanism is to support the movement of the mobile contents source. It shows the average content retrieval time against various network sizes. The content retrieval time means the total time until a content consumer receives all data packets. As shown in Figure 9, the proposed pDND becomes stable irrespective of the number of network nodes because it needs low handover latency by using the preallocated name prefix. On the other hand, existing fDND scheme requires any second unit DND operation time. This leads to the high drop rate of interest packets during handover and high volume of interest retransmission. It shows that the proposed pDND mechanism can provide

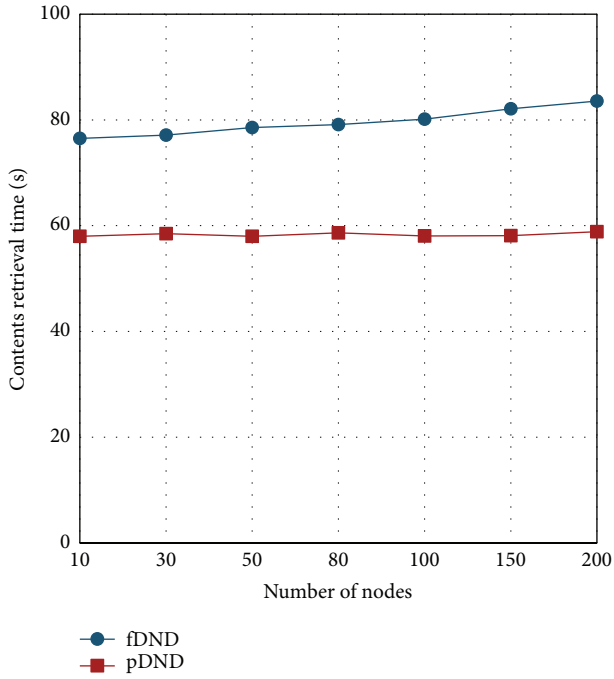


Figure 9: Contents retrieval time.

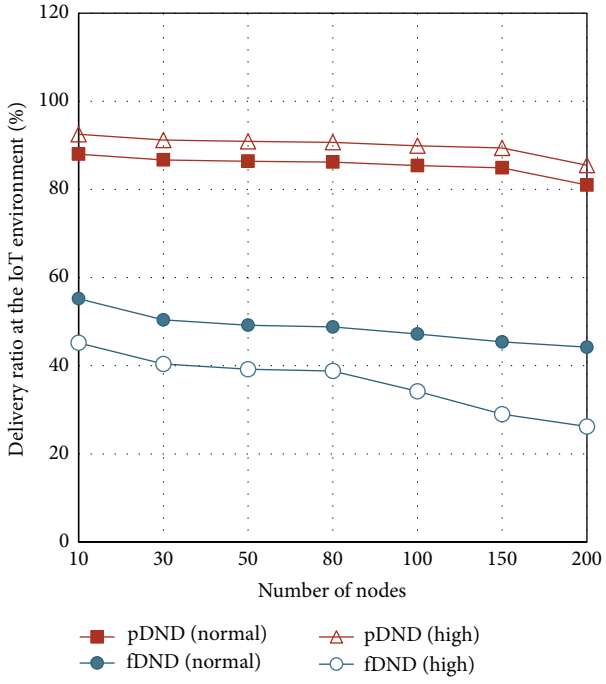


Figure 10: Interest delivery ratio at IoT environment.

stable network performance for the mobile content source in CCN environment irrespective of the network sizes.

Finally, the delivery ratio is shown to indicate how efficient the proposed scheme is to support the types of mobile content sources. Figure 10 shows the average delivery ratio in mobile IoT environment against the frequency of movement of mobile content sources. Normal is defined as average pedestrian walking speed and high as average vehicle driving speed. As shown in Figure 10, existing fDND shows lower delivery ratio as the frequency of a movement is high. This is due to the fact that existing fDND mechanism requires a uniform second unit when a handover event happens. On the contrary, the proposed pDND mechanism shows little performance difference irrespective of the moving frequency. That is, the proposed scheme is influenced only on the number of the handover events because it conducts the new naming configuration in advance.

5. Conclusion

This paper shows the following points. First, it indicates the possibility as to whether mobile CCN scheme can be considered into high movement scenarios like mobile IoT environments. Second, it shows that the frequent network change of MCS's movement causes waste of high network resource consumption and the network's performance drop and high control overhead in CCN environments. Third, to solve such problems for supporting contents source's movements, the proposed mechanism is based on TBR scheme and pDND mechanism. The proposed pDND mechanism utilizes the preallocated name prefix using a naming pool architecture to provide seamless MCS's movement in mobile CCN environment. It can save resource consumption of network devices by limiting the range of routing updates only to the home domain CR. Also, it can save network resource consumption due to low interest message overhead compared to fDND mechanism. This feature is important in CCN based mobile IoT networks since various mobile content sources inherently cause dynamic topology change, high resource consumption, and low route aggregation.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This research was supported by a 2015 Research Grant from Sangmyung University.

References

[1] S. Sen, C. Joe-Wong, S. Ha, and M. Chiang, "Incentivizing time-shifting of data: a survey of time-dependent pricing for internet access," *IEEE Communications Magazine*, vol. 50, no. 11, pp. 91–99, 2012.

[2] T. Koponen, M. Chawla, B.-G. Chun et al., "A data-oriented (and beyond) network architecture," *ACM SIGCOMM Computer Communication Review*, vol. 37, no. 4, pp. 181–192, 2007.

[3] M. Brunner, H. Abramowicz, N. Niebert, and L. M. Correia, "4WARD: a European perspective towards the future internet," *IEICE Transactions on Communications*, vol. E93-B, no. 3, pp. 442–445, 2010.

[4] C. Dannewitz, "Netinf: an information-centric design for the future internet," in *Proceedings of the 3rd GI/ITG KuVS Workshop on the Future Internet*, May 2009.

[5] V. Dimitrov and V. Koptchev, "PSIRP project: publish-subscribe internet routing paradigm," in *Proceedings of the 11th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing on International Conference on Computer Systems and Technologies (CompSysTech '10)*, pp. 167–171, June 2010.

[6] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, "Networking named content," in *Proceedings of the ACM Conference on Emerging Networking Experiments and Technologies (CoNEXT '09)*, pp. 1–12, Rome, Italy, December 2009.

[7] Y. Luo, J. Eymann, K. Angrishi et al., "Mobility support for content centric networking: case study," in *Mobile Networks and Management*, vol. 97 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pp. 76–89, Springer, Berlin, Germany, 2012.

[8] J. Lee, S. Cho, and D. Kim, "Device mobility management in content-centric networking," *IEEE Communications Magazine*, vol. 50, no. 12, pp. 28–34, 2012.

[9] F. Hermans, E. Ngai, and P. Gunningberg, "Mobile sources in an information-centric network with hierarchical names: an indirection approach," in *Proceedings of the 7th Swedish National Computer Networking Workshop (SNCNW '11)*, Linköping, Sweden, June 2011.

[10] F. Hermans, E. Ngai, and P. Gunningberg, "Global source mobility in the content-centric networking architecture," in *Proceedings of the 1st ACM Workshop on Emerging Name-Oriented Mobile Networking Design—Architecture, Algorithms, and Applications (NoM '12)*, pp. 13–18, ACM, Hilton Head Island, SC, USA, June 2012.

[11] Z. Yan, H. Zhou, H. Zhang, and S. Zhang, "A novel mobility management mechanism based on an efficient Locator/ID separation scheme," in *Proceedings of the IEEE 1st International Conference on Future Information Networks (ICFIN '09)*, pp. 11–16, Beijing, China, October 2009.

[12] D. Han, M. Lee, K. Cho, T. Kwon, and Y. Choi, "Publisher mobility support in content centric networks," in *Proceedings of the 28th International Conference on Information Networking (ICOIN '14)*, pp. 214–219, IEEE, Phuket, Thailand, February 2014.

[13] D.-H. Kim, J.-H. Kim, Y.-S. Kim, H.-S. Yoon, and I. Yeom, "End-to-end mobility support in content centric networks," *International Journal of Communication Systems*, vol. 28, no. 6, pp. 1151–1167, 2015.

[14] J. Lee and J. Lee, "Efficient duplicate name prefix detection mechanism in content-centric network," *Advanced Science and Technology Letters*, vol. 54, pp. 45–48, 2014.

[15] G. C. Polyzos and N. Fotiou, "Building a reliable internet of things using information-centric networking," *Journal of Reliable Intelligent Environments*, vol. 1, no. 1, pp. 47–58, 2015.

[16] C.-W. Tsai, C.-F. Lai, and A. V. Vasilakos, "Future Internet of Things: open issues and challenges," *Wireless Networks*, vol. 20, no. 8, pp. 2201–2217, 2014.

[17] J. François, T. Cholez, and T. Engel, "CCN traffic optimization for IoT," in *Proceedings of the 4th International Conference on the Network of the Future (NoF '13)*, pp. 1–5, Pohang, South Korea, October 2013.

[18] E. Baccelli, C. Mehlis, O. Hahm, T. C. Schmidt, and M. Wählisch, "Information centric networking in the IoT: experiments with NDN in the wild," in *Proceedings of the 1st International Conference on Information-Centric Networking (ICN '14)*, pp. 77–86, ACM, Paris, France, September 2014.

[19] S. Mastorakis, A. Afanasyev, I. Moiseenko, and L. Zhang, "ndnSIM 2.0: a new version of the NDN simulator for NS-3," Tech. Rep. NDN-0028, NDN, 2015.

[20] A. Afanasyev, I. Moiseenko, and L. Zhang, "ndnSIM: NDN simulator for NS-3," Tech. Rep. NDN-0005, NDN, 2012.

[21] L. Wang, A. Afanasyev, R. Kuntz, R. Vuyyuru, R. Wakikawa, and L. Zhang, "Rapid traffic information dissemination using named data," in *Proceedings of the 1st ACM Workshop on Emerging Name-Oriented Mobile Networking Design—Architecture, Algorithms, and Applications (NoM '12)*, pp. 7–12, Hilton Head, SC, USA, June 2012.

[22] Y. Abidy, B. Saadallahy, A. Lahmadi, and O. Festor, "Named data aggregation in wireless sensor networks," in *Proceedings of the IEEE Network Operations and Management Symposium (NOMS '14)*, pp. 1–8, Krakow, Poland, May 2014.

[23] G. Han, Z. Zhou, J. Rodrigues, L. Dong, and K. Namuduri, "Mobility support for next-generation wireless sensor networks," *International Journal of Distributed Sensor Networks*, 2015.

Hindawi Publishing Corporation
Mobile Information Systems
Volume 2016, Article ID 9202151, 9 pages
http://dx.doi.org/10.1155/2016/9202151

Research Article

Group-Interest-Based Verifiable CCN

DaeYoub Kim

Department of Information Security, Suwon University, 17 Wauan-gil, Bongdam-eup, Haseong-si, Gyeonggi-do 445-743, Republic of Korea

Correspondence should be addressed to DaeYoub Kim; daeyoub69@suwon.ac.kr

Received 11 September 2015; Accepted 29 December 2015

Academic Editor: Kamal Deep Singh

To solve various problems of the Internet, content centric networking (CCN), one of information centric networking architectures (ICN), provides both an in-network content caching scheme and a built-in content verification scheme. However, a user is still asked to generate many request messages when retrieving fragmented content through CCN. This model can seriously increase the amount of network traffic. Furthermore, when receiving content, a user is asked to verify the received content before using it. This verification process can cause a serious service delay. To improve such inefficiencies, this paper proposes a transmission process to handle request messages at one time. Also, it suggests an efficient content verification method using both hash chains and Merkel-hash tree.

1. Introduction

The Internet was originally designed to establish reliable connections between remotely located hosts [1]. The initial designers of the Internet did not expect that the Internet would be utilized for various services/applications as now. Also, they did not consider various problems which are currently faced by the Internet: as various services/applications begin to utilize the Internet, the amount of network traffic rapidly increases, which leads to serious network congestion [2]. For example, as mobile/smart consumer devices are popularized, it becomes trendy for users to actively generate/share content from their daily lives with others using their own mobile/smart devices. Also, the quality of shared content has become higher than that of the past [3]. Furthermore, various IoT (Internet of Things) services like a vehicle communication system gather/provide massive amounts of information through the Internet [4].

Besides a network congestion problem, the weak security of the Internet is also a serious problem which should be improved [5, 6].

To solve such problems of the Internet, various future Internet architectures/technologies like information centric networking architecture (ICN) are introduced [7]. Specially, since ICN is focusing on contents itself, not on a host providing the content, ICN can make it possible that a user receives content from several possible hosts caching the content. So a user can access content through ICN more efficiently as well as more rapidly than through the Internet [8–10].

Content centric networking architecture (CCN) is one of ICN [11, 12]. CCN has several distinguishing characteristics as follows:

(i) It is designed as a request-driven communication model.

(ii) It utilizes in-network caching functionality to enhance network efficiency.

(iii) It delivers network packets referring to a content identity, not to a device identity (e.g., IP/MAC address) so as to efficiently use cached content.

(iv) It provides a built-in content verification mechanism to authenticate both received content and the original publisher of the content.

However, such characteristics of CCN still cause network/computation inefficiencies. Actually, to distribute content through CCN, the content is fragmented into several segments with small size, and each segment of the content is handled as an independent data in CCN. Hence, when requesting the content, a user should generate a request

message for each segment of the content. These request messages can increase the amount of network traffic so as to be misused by denial-of-service (DoS) attackers [13, 14].

Furthermore, since CCN utilizes content cached in intermediated nodes, it is possible that a user receives content from unknown (malicious) nodes, not from the original publisher of the content. So CCN highly recommends that a user verifies received content before using the content. This content verification process could solve certain security problems of the Internet such as malware and man-in-the-middle attacks [15]. However, since a user is asked to verify all segments of content, such a recursive verification process can cause long service delays [16, 17].

Hence, utilizing CCN for various IT services like multimedia content distribution services as well as various IoT services, the transmission and computation overheads of CCN should be improved [18]. Hence, this paper proposes a process to handle a set of serial request messages at one time to enhance the network efficiency of CCN as well as an improved content verification mechanism to reduce the service latency of CCN.

2. Content Centric Networking

To enhance network efficiency, CCN implements a content-caching functionality on network nodes. Then if a node caching content receives a request packet (Interest) for the cached content, the node transmits the cached content as a response packet (data) to the sender of the Interest and then finishes forwarding the Interest. Hence, a user can receive the content more rapidly than when receiving the content from the original provider of the content. Also, since request messages that converged to the original provider of the content can be handled by intermediated nodes, CCN can solve a network congestion problem which can happen close to the content provider.

Also, to efficiently use the cached content, CCN utilizes the hierarchical identity of content as a packet forwarding address. Since this hierarchical identity of content should be uniquely defined in network, when receiving Interest, an intermediated node can search cached content in its storage (content store, CS) just analyzing the forwarding address of the Interest. The hierarchical identity of content is called a content name.

Figure 1 describes CCN process to handle Interest/data:

(1) If a user generates/sends Interest for a segment of content (e.g., a.mpg), an intermediated node receives the Interest through its interface (e.g., Face 1).

(2) The node checks whether the requested segment has been cached in CS. If it has been, the node sends back the cached segment through Face 1 as data. Then the node completes the processing of the received Interest.

(3) If the requested segment is not cached in CS, the node checks its pending Interest table (PIT) to confirm whether it has already forwarded the same Interest. If the node did, since the content name of the Interest has been recorded in its PIT, the node can find an entry of PIT which is relevant to the Interest. In this case, the node just adds Face 1 on the found entry of PIT, and then stops handling the Interest.

(4) If there is no found entry of PIT, the node compares the content name of the Interest with the entries of its forwarding information based (FIB) table using the longest prefix match in order to select a proper interface (e.g., Face 3) through which it will forward the Interest.

(5) The node records both the content name of the Interest and the incoming interface (Face 1) of the Interest on its PIT.

(6) The node forwards the Interest through Face 3.

(7-8) When receiving data, the node checks whether there is an entry of PIT which is matched to the content name of the data. If there is no proper entry of PIT, the node discards the data and then stops handling the data.

(9-10) If there is a proper entry of PIT, the node saves the data in CS and then forwards the data through the faces of the found entry of PIT. Specially, if the node is an end-user's device, it should first check the validity of the data and then save the data in CS only if the data is valid. Finally, the node deletes the found entry of PIT.

Table 1: The structure of a (Group-) Interest.

Interest structure		
BYTE []	name;	
INT	*n*Seg;	
INT	*g*Seg;	// optional
INT	*t*Seg;	// optional
INT	version;	

3. Group-Interest Operation

As shown in Figure 1, to transmit content, it is first required to generate/send Interest. Specially, CCN asks a content publisher to fragment content into several segments with small size to distribute the content. Then CCN deals with each segment of the content as a single data. So, for receiving the content, a user should generate/send many Interests, even though the only difference of these Interests is just the number of segment. This requesting process may increase the amount of network traffic. Also, after receiving the $i-1$th segment of content, a user can generate/transmit Interest for the ith segment of the content. This linear process can lead to long content retrieval latency.

To improve such problems, we suggest a Group-Interest for requesting m serial segments at one time. Table 1 shows an Interest structure for a Group-Interest:

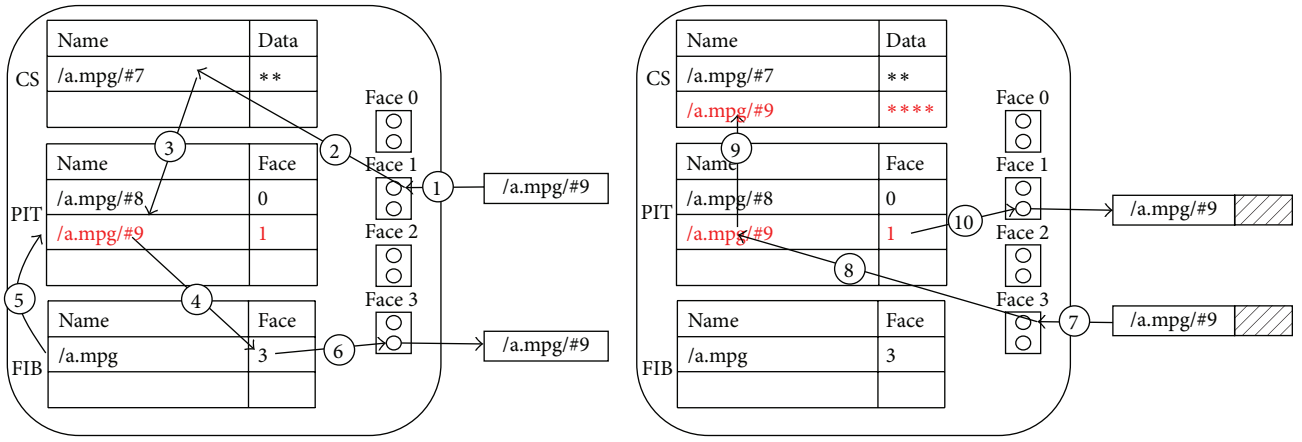
(i) [name] is the hierarchical prefix identities of content.

(ii) [*n*Seg] is a serial number of the segment of the content.

(iii) [version] is the publication time of the content.

Actually, these three fields are the original fields of Interest. The following two optional fields are added for a Group-Interest:

(i) [*g*Seg] describes the number of segments which this Group-Interest requests. That is, this Group-Interest



(a) Interest process (b) Data process

Figure 1: CCN forwarding model using Interest and data.

Table 2: The structure of PIT to handle a (Group-) Interest.

PIT_Entry structure		
BYTE []	name;	
INT	num;	//segment num.
INT	version;	
INT []	face;	
INT	time;	//expire time
INT	gFlag	//group flag
INT	sFlag	//group size
BOOL	rFlag	//response flag

requests serial segments with identities from nSeg to nSeg + gSeg − 1. If gSeg = 1, this Interest is a general Interest, not a Group-Interest.

(ii) [tSeg] is the total number of segments which consists of the requested content. This field is used for verifying gSeg. That is, if nSeg + gSeg − 1 > tSeg, gSeg is invalid.

To handle a Group-Interest, it is necessary to modify the structure of PIT entry as shown in Table 2: [gFlag], [sFlag], and [rFlag] are added.

(i) [gFalg] describes whether this PIT entry is relevant to a Group-Interest or not. If gFlag = 0, this PIT entry is for a general Interest. In this case, both sFlag and rFlag are unmeaning. Otherwise, it means that this PIT entry is relevant to a Group-Interest. Specifically, the value of gFlag is the number of the first segment of the Group-Interest.

(ii) [sFlag] is the number of segments which the Group-Interest requires. This field can be used to delete entry of PIT.

(iii) [rFlag] describes whether the relevant data has been received or not. If rFlag = 1, it means that the data has been transmitted and then forwarded toward requesters.

```
Interest Operation Code
Input: Interest, Face
Output: void
delete expired entries of PIT; // call DeleteEntryOfPIT( );
set f-Flag = 0; // forwardingFlag
for each index i from nSeg to nSeg + gSeg −1 {
  generate Interest[i] such that
    Interest[i].name == Interest.name and
    Interest[i].nSeg == Interest.nSeg;
  find an entry (E[i]) of PIT relevant to Interest[i];
  if there is no, add a new entry to its PIT for Interest[i];
  else {
    if Face isn't in E[i].face, add Face to E[i].face;
    if E[i].gFlag > 0 and E[i].rFlag == 1, set E[i].rFlag = 0;
    else stop handling Interest[i];
  }
  find an entry (C[i]) of CS relevant to Interest[i]
  if there is C[i], then {
    transmit C[i] through E[i].face;
    if E[i].gFlag == 1, then set E[i].rFlag = 1.
    else delete E[i] from PIT; // general Interest
  }
    else set f-Flag = 1;
    if E[i].gFlag > 0 and E[i].rFlag == 0, set f-Flag = 1;
}
find a proper forwarding face referring to FIB table;
if gSeg == 1, forward Interest through the face;
else if f-Flag == 1, forward Interest via the face;
else stop handling Interest;
```

Pseudocode 1: The pseudocode to handle a Group-Interest.

Pseudocodes 1 and 2 are pseudocodes describing how to handle both a general Interest and a Group-Interest. As described in Pseudocode 1, the major differences between the

```
Delete Entry Of PIT Code
Input: void
output: void
for each index i from 1 to sizeOfPIT {
   read the ith entry (E[i]) of PIT;
   if E[i] is expired, delete E[i];
   if E[i].gFlag > 0 and E[i].rFlag == 1, {
      set deleteFlag = 1;
      for each k from 1 to sizeOfPIT {
         if E[k].name == E[i].name and E[k].gFlag ==
         E[i].gFlag and E[k].rFlag == 0, then deleteFlag = 0;
      }
      if deleteFlag == 1, delete E[i];
   }
}
```

Pseudocode 2: The pseudocode to handle PIT.

processes of a general Interest and of a Group-Interest are as follows:

(i) A Group-Interest is disassembled to generate general Interests. These Interests are, respectively, corresponding to serial segments requested by the Group-Interest. These disassembled Interests are only internally used for managing PIT. That is, when handling PIT, a node uses these disassembled Interests, not the original Group-Interest.

(ii) Each entry of PIT generated from a Group-Interest is deleted when either the entry has expired or after all segments requested by the Group-Interest have been forwarded to requestors.

(iii) A Group-Interest is forwarded until all segments requested by the Group-Interest have been transmitted to requestors.

Since the proposed process of Interest as shown in Pseudocode 1 can handle a Group-Interest as well as a general Interest, a user can selectively generate either a general Interest or a Group-Interest considering response status. That is, after receiving some data packets relevant to a Group-Interest, to request remaining data again, a user can selectively generate either a general Interest or a Group-Interest.

Also, it is necessary to modify the process of data in order to handle a Group-Interest. Specially, it is needed to prevent duplicated packet transmission. For that, Pseudocode 3 shows a modified process. The major differences of the modified data process are as follows:

(i) If data is relevant to a Group-Interest and the same data has been forwarded already, a node does not forward the data again even though the relevant entry of PIT exists.

(ii) If data is relevant to a Group-Interest, relevant PIT entry is not instantly deleted from PIT.

```
Data Operation Code
Input: Data
output: void
find an entry (E) of PIT corresponding to Data;
if there is no, stop this process;
else {
   if E.gFlag and E.rFlag are all 1, stop this process;
   save Data in CS;
   forward Data through E.face;
   if E.gFlag is 0, delete E from PIT;
   else set E.rFlag = 1;
}
```

Pseudocode 3: The pseudocode to handle data.

4. Content Verification

In CCN, since a node can receive a segment of content from an anonymous network node caching the segment as well as from the original publisher of the segment, it is possible that malicious nodes send a forged segment.

Hence, a content verification process is one of the essential requirements of CCN. However, since a user should recursively verify each segment of content whenever the user receives the segment, this recursive verification process can cause another inefficiency of CCN.

4.1. MHT-Based Content Verification Scheme. To efficiently verify both the segments of content as well as the original publisher of the content, CCN utilizes a Merkel-hash tree (MHT) [11, 19–21]. Figure 2 shows an example of a MHT-based content verification scheme: assume that a content-publisher fragments content into 7 segments $\{S_2, \ldots, S_8\}$ and then generates meta-data S_1 describing the structure of the segments of the content. From now on, we assume that content consists of 8 segments including a metadata segment.

Step 1 (constructing MHT). A content-publisher builds a binary tree with 8 leaf nodes and then assigns $\{S_1, \ldots, S_7, S_8\}$ to leaf nodes in numerical order. Then the publisher computes the hash value $H(S_i)$ of each segment S_i using the one-way hash function H. The publisher uses $H(S_i)$ as the node value V_k of a leaf node N_k which is assigned to S_i.

Step 2 (computing node values). For each node N_j except for leaf nodes, the publisher computes a node value $V_j = H(V_{2j} \parallel V_{2j+1})$, where $\parallel$ is a concatenation operation and N_j is the parent node of two child nodes, N_{2j} and N_{2j+1}.

Step 3 (signing a root node value). After computing all node values of the binary tree, the publisher signs a root node value V_1 with its signature key SK to generate a signature value (sign).

Step 4 (generating a witness of a segment S_i). For each segment S_i, let N_ks be the sibling nodes of the nodes on the path, from a leaf node assigned to S_i to the root node N_1. The

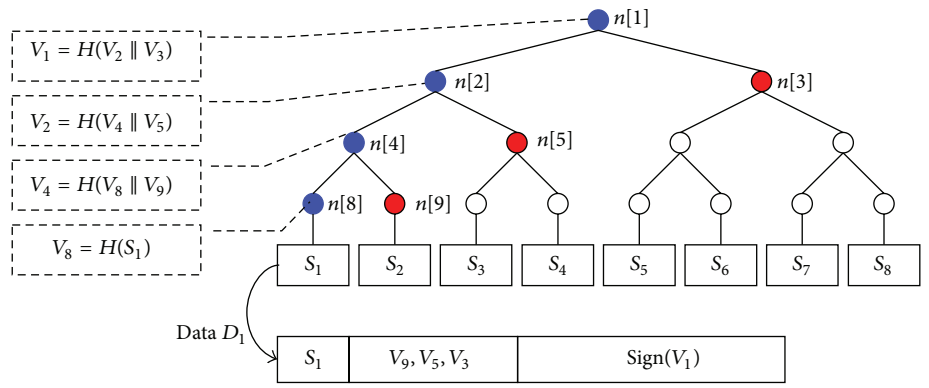
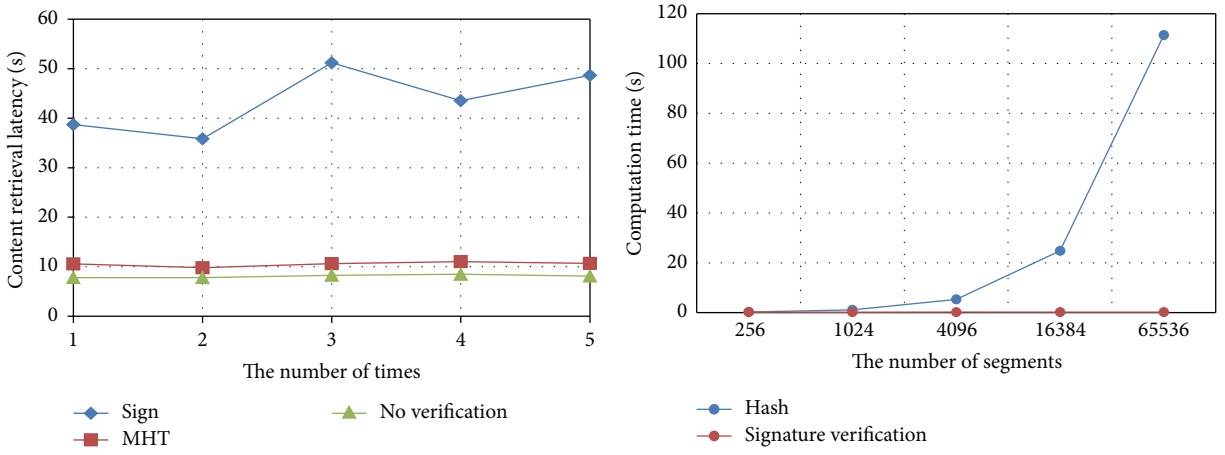


FIGURE 2: MHT-based contents verification: each CCN data contains a segment, a relevant witness, and a signature.



(a) Comparison of content verification schemes

(b) Performance of MHT-based scheme

FIGURE 3: Operation delay. (a) It describes the comparison of response time to share contents fragmented into 256 segments between two smart phones. (b) In the case of applying MHT-based scheme for content.

publisher generates a witness W_i which consists of the node values, V_ks, of N_ks. For example, the W_1 of S_1 is $\{V_9, V_5, V_3\}$ in Figure 2. The witness is needed to verify the sign. That is, using both S_i and W_i, any verifier can compute the same V_1 and then verify the sign.

Step 5 (packaging as data). The publisher generates data (D_i) packaging S_i, W_i, and sign.

If a user receives D_i, the user recursively computes the necessary hash values using both S_i and W_i to compute the root node value, V_1. Then the user verifies the packaged sign using the computed V_1. In practice, after verifying the sign packaged in the first type of data, D_1, the user temporarily saves the computed V_1. Then, the user does not need to verify the sign again for verifying S_i ($i > 1$). Instead, it is sufficient that the user just compares the computed V_1 with the previously saved V_1. Hence, it is possible to reduce the operation time of a segment verification process.

However, as shown in Figure 3, the operation delay of a MHT-based verification scheme is still a burden to CCN. Figure 3(a) shows the comparison result of response times when sharing 256 segments of content between two smart phones over WLAN using three different methods:

(i) [No verification] is a case that a user does not verify received data at all.

(ii) [Sign] is a case that each data has a relevant signature value in order that any user receiving the data can instantly verify the data.

(iii) [MHT] is a case to verify data using MHT.

The result shows that a MHT-based verification scheme can reduce the response time needed to verify segments as compared with [Sign]. That is, [MHT] is more efficient than [Sign] by about 75%.

However, as shown in Figure 3(b), the computation overhead of MHT has increased proportionally to the number of segments, that is, to the size of content. It means that a MHT-based verification process can still cause a serious service delay when distributing high-quality, large-size content. This overhead is due to the fact that the number of recursive hash operations of a MHT-based scheme increases. Hence, to improve the performance of the content verification process

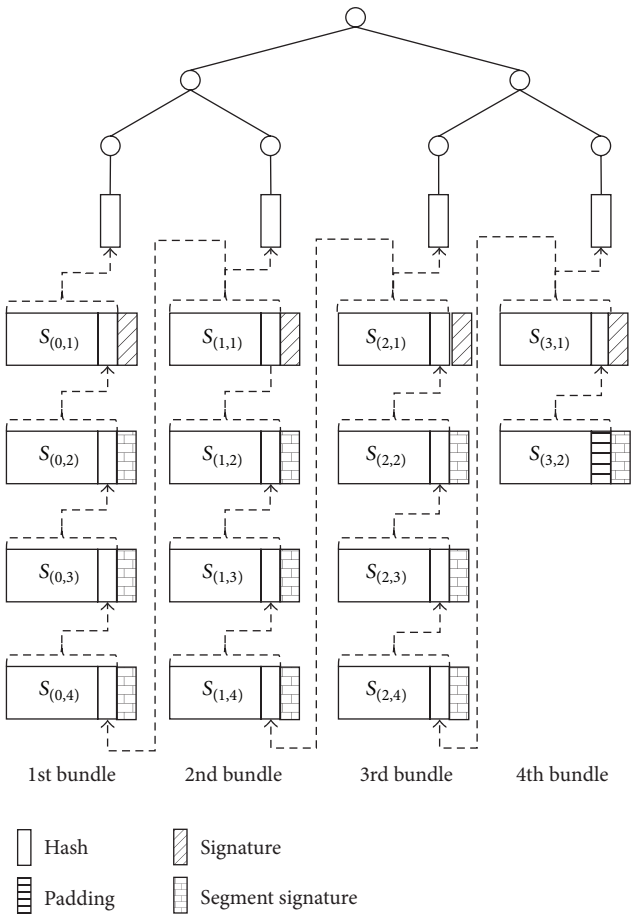


Figure 4: Bundle Interest-based CCN content verification using H-MHT.

of CCN, it is necessary to reduce the total number of hash operations needed to verify the segments of content.

4.2. Hash Chain Based MHT Verification. To reduce an operation delay caused by the content verification process of CCN, in this section, it is proposed to combine MHT with a hash chain which is a general approach to reduce the amount of verification data. It is called a hash chain based MHT scheme (H-MHT).

4.2.1. Verifiable Data Generation. As shown in Figure 4, H-MHT utilizes both MHT and hash value chains: let the number ($N = 2^n$) of the leaf nodes of MHT be 4. Let the number (S) of the segments of content be 14. That is, the content consists of 14 segments, $\{S_1, \ldots, S_{14}\}$ including a metadata segment.

Step 1 (bundling segments). The content-publisher divides 14 ($= S$) segments into 4 ($= N$) segment bundles $\{B^0, B^1, B^2, B^3\}$. Let b_i be the size of a bundle B^i. Let $S_{(i,j)}$ be the jth element of B^i. For each k, segment S_k is assigned to $S_{(i,j)}$, where $k = i \times N + j$:

(a) If $1 \le i < N-1$, $b_i = \lceil S/N \rceil$. In Figure 4, B^i $(1 \le i < 3)$ consists of 4 segments in order.

(b) Otherwise, $b_i \le \lceil S/N \rceil$. In Figure 4, the final segment bundle B^3 consists of balanced segments.

Step 2 (attaching the hash value of the next segment). For each k $(1 \le k < S)$, it computes $H(S_{(k+1)})$ and then concatenates the computed hash value to $S_{(k)}$. Let $S'_{(k)} = S_{(k)} \parallel H(S_{(k+1)})$. If $k = S$, $S'_{(k)} = S_{(k)} \parallel$ null_padding.

Step 3 (constructing MHT). For each i, it computes $H(S'_{(i,1)})$ and then assigns the computed hash value to a leaf node of MHT as its node value in order. Also, it computes the witness W_i of $S'_{(i,1)}$. Finally, it signs the V_1 of MHT with its private key SK. Let the generated signature value be sign.

Step 4 (generating $D_{(k)} = D_{(i,j)}$). For each $j > 1$, let $\text{sign}_{(k)}$ be the signature value of $H(S_{(i,j)})$ generated with SK. The publisher generates $D_{(i,j)}$ which is data for delivering a segment $S_{(i,j)}$ as follows:

(a) If $j = 1$, $D_{(i,1)} = \{S'_{(i,1)}, W_i, \text{sign}\}$.

(b) If $1 < j < b_i$, $D_{(i,j)} = \{S'_{(i,j)}, \text{sign}_{(k)}\}$.

(c) In the case of $j = b_i$, if $k = N$, $D_{(i,j)} = \{S_{(i,j)}, \text{padding}, \text{sign}_{(k)}\}$. Otherwise, $D_{(i,j)} = \{S'_{(i,j)}, \text{sign}_{(k)}\}$.

The $\text{sign}_{(k)}$ attached to $D_{(i,j)}$ is an optional field considering packet loss situation. Hence, if packet loss rate is negligible or the impact of packet loss is not serious, $\text{sign}_{(k)}$ can be removed from $D_{(i,j)}$.

4.2.2. Data Verification. When receiving $D_{(k)} = D_{(i,j)}$, a user verifies the $S_{(i,j)}$ of $D_{(i,j)}$ as follows.

Case 1 ($k = 1$). If $D_{(i,j)}$ is for the first segment, that is, $D_{(i,j)} = D_{(0,1)}$, the user computes V_1 using both S_1 and W_1 and then verifies the sign of $D_{(0,1)}$. If valid, it regards $D_{(0,1)}$ as valid data and then temporarily saves both $H(S_{(0,2)})$ packaged in $D_{(0,1)}$ and the computed V_1 to verify the next data, $D_{(0,2)}$, and $D_{(r,1)}$, respectively.

Case 2 ($k > 1$ and $H(S_{(k)})$ has been saved). If $H(S_{(k)})$ has previously been saved when handling $D_{(k-1)}$, the user computes the hash value of $S_{(k)}$ packaged in $D_{(k)}$ and then compares the computed hash value with the saved $H(S_{(k)})$. If the two values are the same, the user regards $D_{(k)}$ as valid data. Then the user temporarily saves $H(S_{(k+1)})$ packaged in $D_{(k)}$ to verify the next data, $D_{(k+1)}$.

Case 3 ($k > 1$ but $H(S_{(k)})$ has not been saved). (a) If $j = 1$, the user computes V_1 using W_i and then compares the computed V_1 with the previously saved V_1 in Case 1. If the two values are equal, the user regards $D_{(i,1)}$ as valid and then temporarily saves $H(S_{(k+1)})$ packaged in $D_{(k)} = D_{(i,1)}$ to verify the next data, $D_{(k+1)} = D_{(i,2)}$, if $D_{(k)}$ is not the final segment of content.

(b) Otherwise, the user verifies $\text{sign}_{(k)}$ attached in $D_{(k)}$. If valid, it temporarily regards $D_{(k)}$ as a valid data and then saves

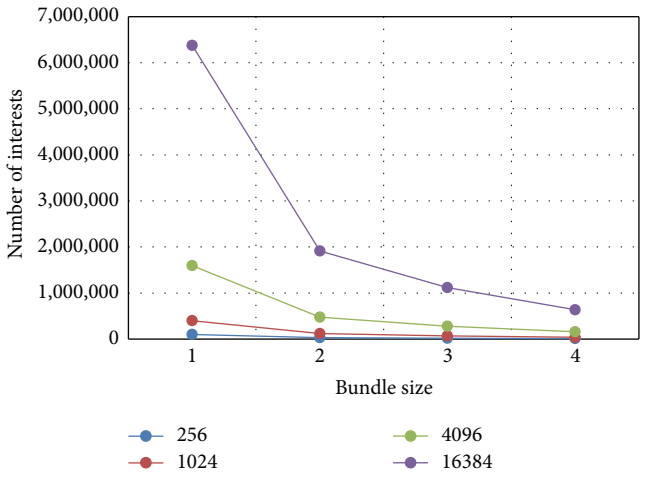


Figure 5: Transmission overheads for forwarding Interest.

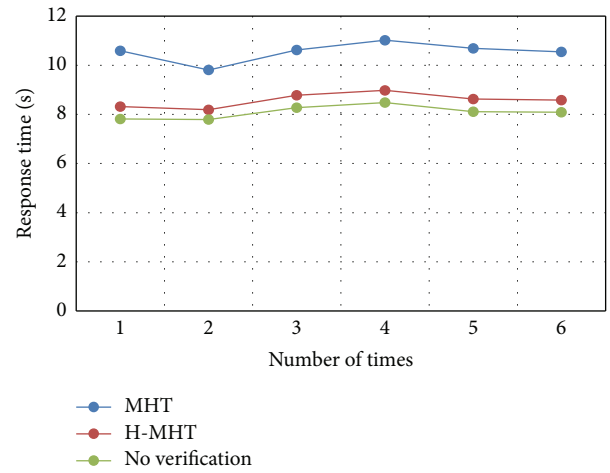


Figure 6: A comparison of response times to share a content fragmented into 256 segments between two smart phones.

both $H(S_{(k)})$ and $H(S_{(k+1)})$. The $D_{(k)}$ will finally be verified after achieving the verification of $D_{(k-1)}$ and comparing the saved $H(S_{(k)})$ with the $H(S_{(k)})$ attached in the valid $D_{(k-1)}$.

5. Performance Evaluation

5.1. Group-Interest Performance. To evaluate the transmission overheads of a Group-Interest, we assume the following:

(i) There are 5 networks connected by 5 border gateways. Each network has a binary tree-shaped network topology with depth 3 consisting of multiple CCN routers. Each end-user is initially placed and then is connected to CCN router, respectively.

(ii) During this simulation, a user utilizes only Group-Interests for requesting content and keeps trying to send *Interests* at a predefined sending rate.

(iii) There are 100 content files which users can access. Each content consists of N (= 256, 1024, 4096, and 16384) segments including meta-data. Each bundle consists of m (= 1, 4, 8, and 16) segments in order.

Then we measure the total amount of transmitted Interest.

Figure 5 shows two results. First, if a bundle size is 1, it means a Group-Interest is actually a general Interest. So when using a Group-Interest ($m > 1$), the transmission overheads of CCN can meaningfully be reduced. Second, the larger the size of the bundle of segments is, the more the transmission overheads of Interest are improved. That is, a Group-Interest is especially efficient when being applied to large size content. However, even if some segments requested by a Group-Interest have been responded, the Group-Interest is continuously forwarded until all requested segments are retrieved. Hence, when utilizing a Group-Interest with size m, transmission performance is not enhanced proportionally to m.

5.2. Content Verification Performance. To analyze the performance of the proposed content verification scheme, we

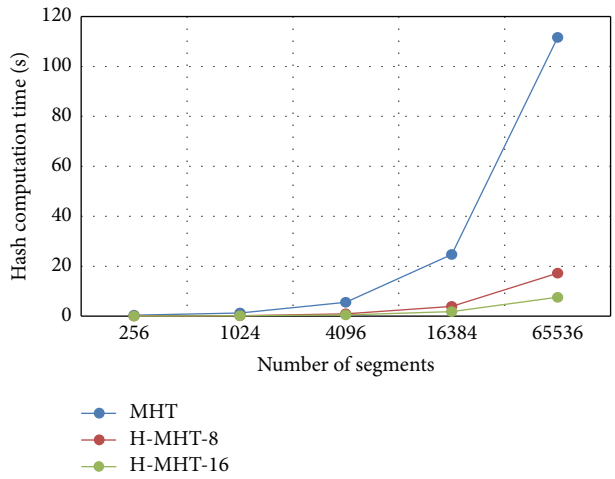


Figure 7: The computation overhead of H-MHT.

assume the configuration of simulation as described in the previous evaluation. Also, H-MHT and MHT use a binary tree with 8 and 64 leaf nodes, respectively. And we use general Interests, not Group-Interests. Then we measure the time for retrieving content. As shown in Figure 6, the response time is improved by about 20%.

Also, we measure the computation overheads of computing hash values for verifying content. Figure 7 shows results considering the cases in which content is fragmented into 256, 1024, 4096, 16384, and 65536 segments, respectively. Then we measure the average time required to compute all hash values for verifying the content:

(i) [MHT] shows the computation overhead of the case of bundle size 1.

(ii) [H-MHT-n] show the results of the cases of bundle size n.

As shown in Figure 7, the larger the bundle size is (as well as the more segments the content is fragmented into), the more efficient the communication overhead is.

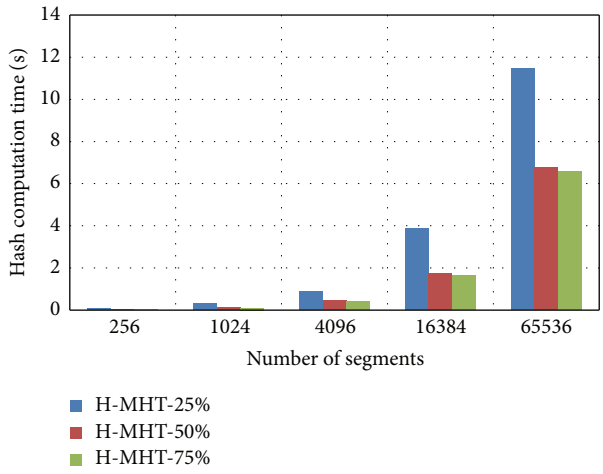


Figure 8: The computation overhead considering different bundle sizes.

Figure 8 is the result of performance measurement of H-MHT considering different bundle sizes. For that, we assume that content is fragmented into 256, 1024, 4096, 16384, and 65536 segments. Let the number of segments be 2^n. For each case, we consider the bundle size as $2^{n\times0.25}$, $2^{n\times0.5}$, and $2^{n\times0.75}$, respectively. As shown in Figure 8, when using the case of a bundle size $2^{n\times0.5}$, the computation efficiency of computing hash values is dramatically improved. But it becomes less effective when using larger bundle sizes than $2^{n\times0.5}$.

6. Conclusion

This paper makes two main points to enhance the performance of CCN. First, since CCN is designed as a request-driven communication model and utilizes fragmented content segments, when requesting content, a user should generate a number of similar Interests to retrieve the content. Using a Group-Interest, it is possible to reduce transmission overheads for forwarding Interests.

Second, an enhanced content verification process is proposed to reduce service latency due to the content verification process of CCN. For that, it is proposed to utilize hash chains. However, when applying a hash chain, it should be considered how to handle packet loss situation. So we also use both MHT and the signature of each segment. Combining MHT to hash chains, it is possible to reduce the computation overheads of a content verification process as well as to limit the effect of packet loss situation. Also, the proposed scheme is designed as being suitable for a Group-Interest. The proposed scheme can provide improved service scalability and low computation costs by reducing the number of hash operations.

These features are important in mobile consumer environments since most mobile consumer devices inherently have limited resource capability. Specially, since various IoT services utilize thin devices like a sensor, these features are meaningful to such services.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This work was supported in part by NRF, Republic of Korea, under Grant no. NRF-2013R1A1A2008389.

References

[1] D. Clark, "The design philosophy of the DARPA internet protocols," *ACM SIGCOMM Computer Communication Review*, vol. 18, no. 4, pp. 106–114, 1988.

[2] Visual Networking Index (VNI), *Cisco Visual Networking Index: Forecast and Methodology, 2014–2019*, Cisco White Paper, 2015.

[3] L. Y. Huanga, Y. J. Hsiehb, and Y. C. Wu, "Gratifications and social network service usage: the mediating role of online experience," *Information Management*, vol. 51, no. 6, pp. 774–778, 2014.

[4] L. Delgrossi and T. Zhang, *Vehicle Safety Communications*, John Wiley & Sons, 2012.

[5] A. Feldmann, "Internet clean-slate design: what and why," *ACM SIGCOMM Computer Communication Review*, vol. 37, no. 3, pp. 59–64, 2007.

[6] R. H. Weber, "Internet of things—new security and privacy challenges," *Computer Law & Security Review*, vol. 26, no. 1, pp. 23–30, 2010.

[7] J. Pan, S. Paul, and R. Jain, "A survey of the research on future internet architectures," *IEEE Communications Magazine*, vol. 49, no. 7, pp. 26–36, 2011.

[8] B. Ahlgren, C. Dannewitz, C. Imbrenda, D. Kutscher, and B. Ohlman, "A survey of information-centric networking," *IEEE Communications Magazine*, vol. 50, no. 7, pp. 26–36, 2012.

[9] G. Xylomenos, C. N. Ververidis, V. A. Siris et al., "A survey of information-centric networking research," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 2, pp. 1024–1049, 2014.

[10] K. Pentikousis, B. Ohlman, D. Corujo et al., "Information-centric networking: baseline scenarios," RFC 7476, March 2015.

[11] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, "Networking named content," in *Proceedings of the ACM Conference on Emerging Networking Experiments and Technologies (CoNEXT '09)*, pp. 1–12, Rome, Italy, December 2009.

[12] V. Jacobson, R. L. Braynard, T. Diebert et al., "Custodian-based information sharing," *IEEE Communications Magazine*, vol. 50, no. 7, pp. 38–43, 2012.

[13] C. Park, T. Kwon, and Y. Choi, "Scalability problem for interest diffusion in content-centric network," in *Proceedings of the 14th Conference on Next Generation Communication Software (NCS '10)*, Pyeongchang, Republic of Korea, December 2010.

[14] S. Choi, K. Kim, S. Kim, and B.-H. Roh, "Threat of DoS by interest flooding attack in content-centric networking," in *Proceedings of the 27th International Conference on Information Networking (ICOIN '13)*, pp. 315–319, Bangkok, Thailand, January 2013.

[15] C. A. Shue, M. Gupta, and M. P. Davy, "Packet forwarding with source verification," *Computer Networks*, vol. 52, no. 8, pp. 1567–1582, 2008.

[16] G. Ma and Z. Chen, "Comparative study on CCN and CDN," in *Proceedings of the IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS '14)*, pp. 169–170, IEEE, Toronto, Canada, May 2014.

[17] C. Ghali, A. Narayanan, D. Oran, G. Tsudik, and C. A. Wood, "Secure fragmentation for content-centric networks," in *Proceedings of the IEEE 14th International Symposium on Network Computing and Applications (NCA '15)*, pp. 47–56, Cambridge, Mass, USA, September 2015.

[18] M. Amadeo, C. Campolo, and A. Molinaro, "Internet of things via named data networking: the support of push traffic," in *Proceedings of the IEEE International Conference and Workshop on the Network of the Future (NOF '14)*, pp. 1–5, Paris, France, December 2014.

[19] R. C. Merkle, "Protocols for public key cryptosystems," in *Proceedings of the IEEE Symposium on Security and Privacy (SP '80)*, pp. 122–134, Oakland, Calif, USA, April 1980.

[20] R. J. Bayardo and J. Sorensen, "Merkle tree authentication of HTTP responses," in *Proceedings of the Special Interest Tracks and Posters of the14th International World Wide Web Conference (WWW '05)*, pp. 1182–1183, ACM, May 2005.

[21] K. Ren, W. Lou, K. Zeng, and P. J. Moran, "On broadcast authentication in wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 11, pp. 4136–4144, 2007.

Hindawi Publishing Corporation
Mobile Information Systems
Volume 2016, Article ID 5232846, 9 pages
http://dx.doi.org/10.1155/2016/5232846

Research Article

F2AC: A Lightweight, Fine-Grained, and Flexible Access Control Scheme for File Storage in Mobile Cloud Computing

Wei Ren, Lingling Zeng, Ran Liu, and Chi Cheng

School of Computer Science, China University of Geosciences, Wuhan 430074, China

Correspondence should be addressed to Wei Ren; weirencs@cug.edu.cn

Received 6 September 2015; Revised 11 November 2015; Accepted 19 November 2015

Academic Editor: Jong-Hyouk Lee

Current file storage service models for cloud servers assume that users either belong to single layer with different privileges or cannot authorize privileges iteratively. Thus, the access control is not fine-grained and flexible. Besides, most access control methods at cloud servers mainly rely on computationally intensive cryptographic algorithms and, especially, may not be able to support highly dynamic ad hoc groups with addition and removal of group members. In this paper, we propose a scheme called F2AC, which is a lightweight, fine-grained, and flexible access control scheme for file storage in mobile cloud computing. F2AC can not only achieve iterative authorization, authentication with tailored policies, and access control for dynamically changing accessing groups, but also provide access privilege transition and revocation. A new access control model called directed tree with linked leaf model is proposed for further implementations in data structures and algorithms. The extensive analysis is given for justifying the soundness and completeness of F2AC.

1. Introduction

With the pervasive usage of mobile handheld computing devices such as mobile phones, tablets, and laptops, mobile business processing becomes possible and grows largely during commercial traveling. As storage services in cloud, such as Google, Alibaba, and Baidu, are freely provided, users may rely on these services to share and edit business files with others remotely and cooperatively. For example, after one user creates or uploads a file into cloud servers, the others can access the file remotely and edit the file cooperatively.

Currently, mobile storage cloud services impose a typical security problem, access control for distributed users who will access the shared file in cloud servers. In particular, the related access control policies should be determined by user themselves, which result in sophisticated requirements. Moreover, many service providers simply assume that the other users have almost the same access privileges as the original user who creates and uploads files. It can simplify management logics at cloud but raise the risks of file leakage at client. That is, the other users can arbitrarily read, modify, and update uploaded files, which usually imposes security risks in mobile scenarios.

For example, a user (called A) uploads a file into a storage cloud. She tells the other two users (called B and C) user name and password for logging into the storage cloud. B and C can use the user name and password to successfully log into the storage cloud and access the file, like A. In other words, A, B, and C have the same privilege for the file, which is not fine-grained. It may result in security risks such as the leakage or damage of files.

When the number of uploaded files and shared users increases, the fine-grained access control of those files for those users is mandatory. It is worth to note that because access control policies are not determined by the administrator of cloud servers, control methods should be so easy that ordinary cloud users can understand and conduct straightforwardly. It should be flexible in that access control policies are user-centric and can be defined on demand via operational interfaces provided by cloud servers. Moreover, the access control should be lightweight; otherwise, the response delay upon accessing will not be endured and user experiences will be worse to damage the QoS (Quality of Service) of storage cloud services.

Recently, access control in storage cloud has attracted more and more attention [1–3]. However, in those methods

a fine-grained, flexible, and lightweight solution has not been thoroughly explored. We make the first attempt to solve it in this regard. However, such an access control scheme poses three challenges as follows:

(1) The flexible property requires that the proposed method must tackle complicated situations such as highly dynamic ad hoc groups, in which accessing users can join and leave groups conveniently. Moreover, all users may not be in the same layer (in terms of file sharing relationships). That is, a leader of a group can be authorized by the initiator and the leader can further assign and revoke privileges to other group members. A subgroup (or sublayer) can be formed in a group or a layer, so the authorization can be conducted in an iterative manner.

(2) The fine-grained property requires that user privileges should be easily defined, changed, revoked, verified, and managed via various rules and policies. And especially, those can be determined and controlled by cloud user themselves. The access control mechanism should smoothly make it possible for users to edit shared files remotely, cooperatively, and securely.

(3) The lightweight property requires that computational overhead should be managed in cloud side, as the total number of users in storage cloud is always huge. Thus, cryptographic algorithms should be avoided as least as possible. The computational overhead at mobile clients should be as least as possible for better performance in energy consumption and user experiences.

In this paper, we propose a scheme called F2AC (lightweight, fine-grained, and flexible access control) to tackle the above challenges. We present and analyze the design rationale in an incremental way for better understanding. Some formal presentations are presented for better clarity and rigorous generality.

The contributions of the paper are listed as follows.

(1) F2AC can create, add, and delete users in a group (and a subgroup iteratively), authorize a user as a group leader who can authorize privileges to other group users iteratively, and revoke privileges for a user in the group.

(2) F2AC can manage access control such as merge, delete, and retrieve user or privileges in a lightweight manner via a proposed access control model, directed tree with linked leaf model.

(3) F2AC can permit users to define various access control rules as they demand and separate the access control for system users and file users, which simplifies user experiences and management flows in cloud.

The rest of the paper is organized as follows. Section 2 gives an overview on relevant prior work. In Section 3 we discuss the basic assumption used throughout the paper. Section 4 provides the detailed description of our proposed models and analysis. Finally, Section 5 concludes the paper.

2. Related Work

Access control methods for mobile cloud have attracted more and more attention [1, 3, 4]. Ghafoor et al. [5] suggested to enable users to create and manage access control policies on their resources according to their own security and access control requirements. They proposed a framework, standard policy definition language, and user interface to specify and manage access control. Tang et al. [6] designed and implemented a secure overlay cloud storage system that achieves fine-grained, policy-based access control and file assured deletion. Habiba et al. [7] presented framework and different modules along with their functionalities. They also described Multiagent Based System (MAS) and an enhanced authorization scheme. Ruj et al. [4] proposed a new decentralized access control scheme that supports anonymous authentication. Their scheme prevents replay attacks and supports creation, modification, and reading data stored in the cloud. Wan et al. [8] argued that attribute-based encryption (ABE) suffers from inflexibility in implementing complex access control policies. They thus proposed hierarchical attribute-set-based encryption (HASBE) by extending ciphertext-policy attribute-set-based encryption (ASBE) with a hierarchical structure of users. Hajivali et al. [9] proposed agent-based user authentication and access control algorithm based on discretionary and role-based access control model for increasing the reliability and rate of trust in cloud computing environments. Their model uses a cloud-based software-as-a-service application and a client-based user authentication application. Ortiz et al. [10] discussed the industrial application of the extensions to traditional role-based access control to enable secure and mobile collaboration among manufacturing enterprisers. Lv et al. [11] proposed a modified CP-ABE algorithm to set up a fine-grained access control method, in which user revocation is achieved based on the theory of Shamir's secret sharing. Yang et al. [12] proposed to delegate the computation intensive task, such as data reencryption, key distribution, and derivation to cloud servers. Their scheme required bilinear pairing and random padding. Yao et al. [2] proposed a lightweight cipher-text access control mechanism based on authorization certificates and secret sharing for mobile cloud computing. Jung et al. [3] proposed to control privilege and anonymity by fully anonymous ABE. Shen et al. [13] studied the problem of keyword search with access control over encrypted data in cloud. They proposed a framework where user can use his attribute values and a search query to locally derive a search capability, and a file can be retrieved only when its keywords match the query and the user's attribute values can pass the policy check. They also proposed a scheme that utilizes HPE to enforce fine-grained access control, and support the derivation of the search capability. We argue that most of current works extensively rely on ABE, which may not be lightweight due to encryption operations. Moreover, ABE-based schemes cannot support flexible self-defined access policies or cannot be fine-grained in terms of dynamical and iterative privilege authorization (and revocation) when ad hoc groups change and group members vary.

3. Problem Formulation

3.1. Network Model. There exist at least three major entities in file storage cloud: an entity A who uploads files and shares

those files, an entity B who is asked to access these files, and a storage server C who is at cloud side. Simply speaking, A uploads files into C. A gives login account information to B. B logs in and accesses files according to her privilege that is set by A.

In this paper, we will tackle more complicated scenarios and manage access privileges in a fine-grained, flexible, and lightweight manner.

Definition 1 (original user). This is the user who applies for an account from a cloud storage service for uploading files, for example, Alibaba's AliYun. She uploads files into the cloud server. She is denoted by $User_A$ in further examples.

Definition 2 (accessing user). After original user logs in to cloud and uploads files, she will be asked to assign access privileges to some accessing users for those files. Original user also needs to specify accessing user's account information for logging in to cloud systems and token information for distinguishing different users for file accessing.

Definition 3 (team leader). In accessing users, some are team leaders who can assign privileges for shared files in this team. Team leader can also add and remove accessing users in this team.

Definition 4 (team member). Accessing user who is added by a team leader or original user is a team member in this team.

Definition 5 (privilege). An accessing user accesses an uploading file according to her privilege that is authorized by team leader or original user.

Definition 6 (authorize privilege). Original user and team leader can have authorize privilege, who can assign privileges to other users, and add or remove accessing users.

3.2. Design Goals. The design of F2AC should tackle the following situations: (1) Original user and team leader can authorize privileges for a highly dynamic group in which users can be added or removed. (2) Those privileges can be flexibly managed by original user or team leader in an iterative way and accessing hierarchical layers can be diverse and arbitrary. (3) The access control can be conducted in a lightweight manner at cloud, especially when system access control is separated from file access control, which is a realistic deployment requirement in storage cloud.

4. Proposed Scheme: F2AC

4.1. Basic Setting. There exist two tables for facilitating control mechanism at *Cloud*. We denote them as $Cloud ::= \langle ACL, UCL\rangle$, where $::=$ can be looked as "is defined to." ACL is used for access control; UCL is used for user authentication.

(1) ACL is a table for access control that has four fields $\langle File_*, User_*, P, C\rangle$, where $File_*$ denotes a file name, $User_*$ denotes a user name, P denotes a privilege, and C denotes one or more conditions. We denote it as $ACL ::= \langle File_*, User_*, P, C\rangle$.

The privileges usually have four types, *Create*, *Update*, *Modify*, and *Read*, which denote create, update, modify, and read, respectively. Users who are assigned *Read* privilege can only view files. Users who are assigned *Modify* can read and modify files but not update files. Users who are assigned *Update* can read, modify, and update the modified content into files. Users who are assigned *Create* can read, modify, and update files and, especially, can assign privileges to other users. We denote it as $P ::= \langle Create, Update, Modify, Read\rangle$. The former one is the superset of the latter one. That is, *Create* contains the privilege *Update*, *Update* contains *Modify*, and *Modify* contains *Read*. In shorthand, $Read \subset Modify \subset Update \subset Create$.

Extra conditions (column C) can include more requirements for access, for example, locations, device MAC addresses, or any other requirements. The default value for it is Null, denoted as $*$. We denote it as $C ::= \langle *, Loc, MAC\rangle$.

(2) UCL is a table for user control that has two fields $\langle User_*, Token_*\rangle$, where $User_*$ is a username and $Token_*$ is a password for user authentication. We denoted it as $UCL ::= \langle User_*, Token_*\rangle$.

F2AC is composed of the following steps.

Step 1 (user: firstly login and upload). For the first time login, $User_A$ can upload one or more than one files into $Cloud_C$.

Step 2 (user: assign one user and one privilege for one file). Once $User_A$ uploads a file successfully (e.g., $File_{A1}$); *Cloud* will add two records into ACL as follows.

Step 2.1. One record is $\langle File_{A1}, User_A, Create, *\rangle$ that is added into ACL automatically.

Step 2.2. $User_A$ will be asked to set up the access privilege of the file ($File_{A1}$) by *Cloud*. That is, $User_A$ specifies the privilege (e.g., *Read*) of $File_{A1}$ to $User_B$ by checking the options (in user interface) provided by *Cloud*. The setting results will be stored in ACL. That is, the record $\langle File_{A1}, User_B, Read, *\rangle$ is added into ACL.

Step 3 (user: assign more users and their privileges for one file). $User_A$ continues to set up more users and more privileges for this file. *Cloud* add more records in ACL correspondingly. That is, Step 2.2 will be conducted iteratively for $File_{A1}$ for assigning more users and privileges.

Step 4 (user: upload more files and assign more users and their privileges for more files). If $User_A$ uploads additional files, for example, $File_{A2}$, $File_{A3}$, Steps 2 and 3 will be reconducted iteratively for those files by *Cloud* and $User_A$.

Step 5 (cloud: store records into ACL). After $User_A$ uploads all files and sets up all options for access control, *Cloud* will store corresponding records into ACL.

Step 6 (user: assign UCL). $User_A$ sets up corresponding token for each user in ACL including herself by checking options (in user interface) provided by *Cloud*.

Step 7 (cloud: store records into *UCL*). After that, *Cloud* stores corresponding records into *UCL*. For example, $\langle User_A, Token_A\rangle$, $\langle User_B, Token_B\rangle$, $\langle User_C, Token_C\rangle$.

Step 8 (user: send account information and tokens to accessing users). $User_A$ gives her account and password for *Cloud*, together with $Token_B$, to $User_B$; $User_A$ gives her account and password for *Cloud*, together with $Token_C$, to $User_C$. The account and password of $User_B$ and $User_C$ are the same as those of $User_A$. The token is used for authentication and distinction of different accessing users.

Step 9 (user: secondly login and present token). After $User_A$ logs in to *Cloud* at the second time, she will be asked for her token by *Cloud*. The reason is that $User_A$ is in *UCL* at this time.

Step 10 (cloud: retrieve *UCL* and *ACL*). After $User_A$ responds her token, *Cloud* will retrieve *UCL* to get the user name, namely $User_A$. *Cloud* lists all files for $User_A$ by retrieving *ACL* according to her privileges.

Step 11 (other users: login and present token). After $User_B$ logs in to *Cloud*, she will be asked for showing her token by *Cloud*. After $User_B$ responds with her token, *Cloud* will conduct processes similar to Step 10.

Remarks 1. (1) It is required to separate the authentication for logging in to *Cloud* and the authentication for access control, namely, user name in *UCL*. It will simplify the implementation and management of user authentication at *Cloud*. The security of original authentication system of *Cloud* will not be damaged. It is also easy for $User_A$ to understand (for better user experience) and conduct user addition for file accessing.

(2) In Step 2.2, one privilege is already enough, as relationships between privileges are subset (or superset). The highest (or largest) privilege is set for one user aiming at one file.

(3) The condition C could be extra accessing rules on demand, which is set "$*$" in the description for simplicity. Indeed, it can be set as *Loc*, *MAC*, and so on. Hereby, $C ::= \langle *, Loc, MAC\rangle$. C can be considered as the extension of P.

(4) Step 8 is usually accomplished offline or out of the channel, which is out of the scope of the scheme.

4.2. Notations. The major notations used in the remainder of the paper are listed as follows for better understanding.

Cloud: File Storage Cloud

P: Privilege, $P ::= \langle Create, Update, Modify, Read\rangle$

$Login(Cloud, User_*)$: User Login Function. Login into *Cloud* as $User_*$

$Upload(File_*)$: User Upload File Function. Upload file $File_*$

$SaveACL(\langle File_*, User_*, P, *\rangle)$: Cloud Save ACL Function

$SetupACL(\langle File_*, User_*, P, *\rangle)$: User Setup ACL Function

$SetupUCL(User_*, Token_*)$: User Setup UCL Function

$SaveUCL(\langle User_*, Token_*\rangle)$: Cloud Save UCL Function

$RequestToken()$: Function Request Token by Cloud after User Login Returning $Token_*$

$RetrieveUCL(Token_*)$: UCL Retrieve Function by Cloud, Taking as input $Token_*$, Returning $User_*$

$RetrieveACL(User_*)$: ACL Retrieve Function by Cloud, Taking as input $User_*$, Returning $File_*, P, C$.

4.3. Abstract Model. The overall procedures are stated as follows in a shorthand for simplicity and clarity as follows.

Abstract Model for F2AC

Step 1

$$User_A : Login(Cloud, User_A),\ User_A : Upload(File_{A1});$$

Step 2.1

$$Cloud : SaveACL(\langle File_{A1}, User_A, Create, *\rangle);$$

Step 2.2

$$User_A : SetupACL(\langle File_{A1}, User_B, Read, *\rangle),$$
$$Cloud : SaveACL(\langle File_{A1}, User_B, Read, *\rangle);$$

Step 3

$$User_A : SetupACL(\langle File_{A1}, User_C, Modify, *\rangle),$$
$$Cloud : SaveACL(\langle File_{A1}, User_C, Modify, *\rangle);$$
$$User_A : SetupACL(\langle File_{A1}, User_D, Update, *\rangle),$$
$$Cloud : SaveACL(\langle File_{A1}, User_D, Update, *\rangle); \ldots$$

Step 4

$$User_A : Upload(File_{A2});$$
$$User_A : Upload(File_{A3}); \ldots$$

Step 5

Redo Steps 2.1, 2.2, and 3;

Step 6

$$User_A : SetupUCL(User_B, Token_B);$$
$$User_A : SetupUCL(User_C, Token_C);$$

Step 7

$$Cloud : SaveUCL(\langle User_A, Token_A\rangle);$$
$$Cloud : SaveUCL(\langle User_B, Token_B\rangle);$$
$$Cloud : SaveUCL(\langle User_C, Token_C\rangle);$$

Step 8

Off-line Operations.

Step 9

$User_A : Login(Cloud, User_A)$;

Step 10

$Cloud : Token_A \Leftarrow RequestToken()$;
$Cloud : U \Leftarrow RetrieveUCL(Token_A)$;
$Cloud : File_*, P, C \Leftarrow RetrieveACL(U)$;

Step 11

$User_B : Login(Cloud, User_A)$;

Step 12

$Cloud : Token_B \Leftarrow AskToken()$;
$Cloud : User_B \Leftarrow RetrieveUCL(Token_B)$;
$Cloud : File_*, P, C \Leftarrow RetrieveACL(User_B)$;
...

Remarks 2. (1) In Step 1, $User_A$ logs in for file uploading. After this login, *UCL* is created. In $User_A$'s next login (e.g., in Step 9), $User_A$ will be requested for her token. If this time no file is uploaded, *UCL* will not be created. That is, once a file is uploaded by a user, *UCL* and *ACL* will be created at *Cloud*.

(2) In Step 11, if a user (e.g., $User_X$) logs in via not using $User_A$'s account (including user name and password), *Cloud* will respond that this user (namely, $User_X$) is not a file sharing user related to $User_A$. It is an independent login, not being related to the file sharing of $User_A$ (i.e., this login may be used for $User_X$'s file uploading and sharing).

In contrast, if $User_A$ logs in with $User_B$'s account, instead of $User_A$'s, $User_A$ will be regarded as $User_B$'s file sharer and be requested for showing the token that is set by $User_B$.

(3) Usually, *Create* can be only assigned to the user who uploads the file into *Cloud*, by *Cloud* automatically in Step 2.1. In the following sections we will extend it for more flexible functions in more complicated application scenarios.

4.4. Extension for Flexibility. In the above abstract model, the major steps for F2AC are described. In this setting, only user that uploads the file can obtain *Create* privilege and can authorize privileges to others. It can simplify the management of privileges, but it may not be convenient when the file is shared in a dynamic group whose members are changed frequently, or a group with a large number of users or different layers.

For example, $User_A$ uploads $File_A$, but $User_A$ is willing to let $User_B$ be a proxy of her for managing the file editing. Thus, $User_A$ hopes that $User_B$ can authorize privileges to other users as a team leader and represent as a proxy of $User_A$; for example, $User_B$ can further assign privilege such as *Read* to $User_C$. To extend this flexibility, we propose an extension of basic settings and abstract model.

Firstly, $P ::= \langle Create, Update, Modify, Read\rangle$ in basic settings will be extended to $P ::= \langle Create, Authorize, Update, Modify, Read\rangle$. That is, the user with *Authorize* privilege will be able to authorize privileges to other users like $User_A$. Certainly, *Authorize* is a superset of *Update*, *Modify,* and *Read*.

Secondly, in abstract model after Step 12, if P that is returned by $RetrieveACL(User_B)$ is Authorize, $User_B$ will be able to set up *ACL*. Four policies are proposed hereby on methods for setting up *ACL* in the following.

Policy 1 (none-transitive *Authorize* privilege and none-additive users). After $User_B$ with the privilege *Authorize* logs in into *Cloud* as $User_A$, all files are listed. $User_B$ selects a file, for example, $File_{A1}$, and the related users and their privileges are listed. $User_B$ can change those privileges that are *Update, Modify,* and *Read*. That is, only the user with *Create* can assign *Authorize* privilege; the user with *Authorize* cannot assign *Authorize* privilege to others but can change privileges such as *Update*, *Modify*, and *Read*. After any chance happens, *ACL* will be updated for column P for corresponding records.

Policy 2 (none-transitive of *Authorize* privilege and additive users). The difference between this policy and Policy 1 is that the user with *Authorize* privilege can add more users for current files. If so, *ACL* will not only be updated at column P, but also be appended more records.

If the trace back on who adds the user into the group is required, *ACL* will be extended to $\langle File_*, User_*, P, C, ByWhom\rangle$. *ByWhom* will record the user who adds $User_*$. For example, if $User_A$ adds $User_B$ into *ACL*, *ByWhom* will be $User_A$. If $User_B$ with privilege *Authorize* adds $User_C$ into *ACL*, *ByWhom* in this row will be set as $User_B$.

If a new user is appended into *ACL*, this user must be appended into *UCL*. The token of the new user is set up by the user who adds the new user. For example, if $User_B$ with privilege *Authorize* adds $User_C$ into *ACL*, $User_B$ should set up $\langle User_C, Token_C\rangle$ in *UCL*.

The deletion of users in *ACL* is only permitted by the user who matches *ByWhom* in this row. That is, if *ByWhom* is $User_A$ in this row in *ACL*, only $User_A$ can delete this row (namely, the user) from *ACL* (by user interface provided by *Cloud*). Once a user is deleted in *ACL*, the user will be also deleted in *UCL* automatically. For example, if $ProjectColumn(ACL, User_*) != User_B$ (i.e., all records like $\langle File_*, User_B, *, *, ByWhom = User_A\rangle$ are deleted in *ACL* by $User_A$), $\langle User_B, Token_B\rangle$ will be deleted from *UCL*.

Policy 3 (transitive of *Authorize* privilege and nonadditive users). The difference between this policy and Policy 1 is that the user with *Authorize* privilege can change privileges *Update, Modify,* and *Read* into *Authorize* in *ACL*. That is, not only the user with *Create* can assign *Authorize* privilege, but also the user with *Authorize* can assign *Authorize* privilege to others. If so, the column P in *ACL* will not only be updated

among *Update, Modify,* and *Read*, but also be updated from *Update*, *Modify*, and *Read* into *Authorize*. However, the user with *Authorize* privilege cannot add more users for current files, namely, nonadditive users.

Policy 4 (transitive of *Authorize* privilege and additive users). This one will be most complicated and flexible. It combines functions in Policies 2 and 3. That is, the user with *Authorize* can assign *Authorize* privilege to others, and the user with *Authorize* privilege can also add more users for current shared files.

Remarks 3. (1) Each user has only one possible *ByWhom* in *ACL*. Once a user is added into *ACL*, the *ByWhom* field will be set. For better understanding, if we consider *ByWhom* as a parent (namely, father) and consider each user as a node, user graph will be considered as a directed tree. The direction is from a child or all children to the father. Recall that, for user deletion, only the father can delete directed child or children in the tree (if user *B*'s *ByWhom* in *ACL* is user *A* who currently logs in to *Cloud*, user *A* will be presented an option on whether deleting *B* or not in user interface).

Note that, for consistence, if a none-leaf node is deleted, the node's all children will replace their fathers by the node's father. (All children's *ByWhom* will be replaced by the node's *ByWhom*.)

(2) A user's privilege for a file can be modified by anyone who has *Create* or *Authorize* privilege for this file. For better understanding, we can look users with *Creat* and *Authorize* for a file as team leaders for this file. Other users with privileges such as *Update*, *Modify*, or *Read* for this file will be considered as group members for this file.

(3) Once a user is added into *ACL*, she will be added into *UCL*. The same user name presents only once in *UCL*. For better understanding, all users in *UCL* can be considered as a group whose members share the same login account information (namely, login user name and password) in *Cloud* as the original user who uploads shared files.

(4) For each file, only one user can have *Create* privilege. The user who has *Create* privilege is the original user (the one who uploads the file into *Cloud*). All users in *UCL* will share the same account information with this user who has *Create* privilege.

(5) A user's father (according to *ByWhom* field) has responsibility to deliver the account information and token to the user, usually offline.

4.5. Directed Tree with Linked Leaf Model. Policy 4 is the most flexible (or powerful) one in all policies for file sharing in a dynamic group or a group with a large number of members. Intuitively, someone may suspect that this policy may result in some inconsistency due to the complexity of user addition, deletion, and transitive authorization. To make it clear, we propose a concept and implementation model for F2AC, which presents as a directed tree with linked leaf (leaves).

(1) Users are organized in a tree structure. Each user presents as one node at the tree. The root of the tree is the user who uploads all current sharing files, namely, one who has *Create* privilege for those files. All users in the tree share the same login account information as the root, but they are distinguished via their tokens.

(2) Each nonroot node has one parent (namely, father). Anyone of these nodes has a directed edge pointed to its father. The father is the user who adds a child node or children nodes in the tree. Only users who have *Authenticate* or *Create* privilege can add a child or children. The father of a child can be fetched from *ACL* by looking up *ByWhom* column.

(3) Each node has three properties, namely, *UserName*, *Token*, *OneParent*. *UserName* and *Token* are assigned by node's father. The property *OneParent* points to node's father.

(4) Each node links to one or multiple properties, called *NodeLink*. Each link has two related properties. One is *AccessFileSet*; the other is *OnePrivilege*. *AccessFileSet* property is a set consisting of one or more files; *OnePrivilege* property is one privilege for files in *AccessFileSet*. Each node may have more than one *AccessFileSet*, but each *AccessFileSet* has one corresponding *OnePrivilege* (namely, the largest privilege). That is, each node links to one or more sets that represent access files, and each set has one bit to represent the privilege for this set. We look on one or multiple properties as linked leaf (leaves) for this node. The model is thus called a tree with linked leaf (leaves). Note that linked leaf is not a leaf node; leaf node is a node that has no child, but linked leaf is a property for every node.

(5) The file set (usually consisting of multiple files) at the father node should be the superset of that (namely, file set) at a child node or children. It is nontrivial to be aware of but can be understood for the reason that father's files are reassigned to a child or children to edit. Thus, a merging set of a child's or children's *AccessFileSet* is upper-bounded by the father's *AccessFileSet*.

(6) *ACL* and *UCL* can be constructed by the above tree model. The retrieval of *ACL* and *UCL* can be accomplished by underlying tree data structures and algorithms.

We denote the tree model in the following.

Directed Tree with Linked Leaf Model

$$
\begin{aligned}
&Tree ::= \langle Root, NodeSet, Edges\rangle,\\
&FileSet ::= \{files \mid uploadedbyRoot\},\\
&Root ::= \langle UserName, Token, OneParent = NULL, RootLink\rangle,\\
&RootLink ::= \langle AccessFileSet = FileSet, OnePrivilege = Create\rangle,\\
&NodeSet ::= \langle Node\rangle, \ldots, \langle Node\rangle\\
&Node ::= \langle UserName, Token, OneParent, NodeLink\rangle,\\
&NodeLink ::= \langle AccessFileSet, OnePrivilege\rangle, \ldots, \langle AccessFileSet, OnePrivilege\rangle,\\
&AccessFileSet ::= \{onefile \in FileSet\} \parallel morefiles \subseteq FileSet,\\
&OnePrivilege ::= \{P \mid P = Authorize \parallel Update \parallel Modify \parallel Read\},\\
&Edges ::= \langle Edge\rangle, \ldots, \langle Edge\rangle,\\
&Edge ::= \langle Node_1, Node_2\rangle \mid \{Node_1.OneParent = Node_2,
\end{aligned}
$$

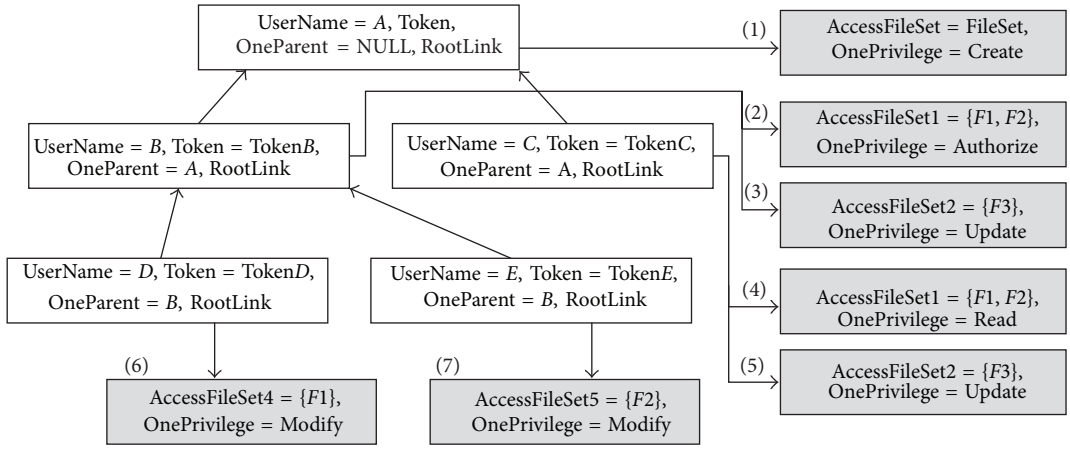


Figure 1: An example for illustration the logic of setup.

$Node_1, Node_2 \in NodeSet$, $Node_2.OnePrivilege = \{Create \parallel Authorize\}\}$,

Example 7. We explain the logic in following example, which follows the sequence number listed in Figure 1 (grey nodes are linked leaves).

(1) $FileSet = F1, F2, F3$. User A (e.g., original user, manager) uploads all three files into *Cloud*. A's privilege for all files is *Create*. A's login account information for *Cloud* will be used for all other users, namely, B, C, D, and E.

(2) $F1$ and $F2$ are two files that will be edited by a team led by B (B is the leader of team 1, e.g., technical team).

(3) $F3$ can be edited by B (by herself, e.g., $F3$ is written by B and will be reported to manager A).

(4) $F1$ and $F2$ can only be read by C (C is from another team, e.g., testing team).

(5) $F3$ can be edited by C (by herself, as $F3$ is written by C and will be reported to manager A). Note that, hereby $F3$ can be edited by B and C cooperatively.

(6) B is the team leader. She adds two users into her team, namely, D and E. B assigns D to edit $F1$, but D cannot update $F1$. The modification on $F1$ by D can be reviewed by B and updated by B.

(7) B assigns E to edit $F2$, but E cannot update $F2$. The modification on $F2$ by E will be reviewed by B and updated by B.

4.6. Constraints and Principles. Next, we point out some nontrivial insightful constraints to depict some inner principles in F2AC to evaluate its soundness.

Principle 1. The number of leaves at *Root* is one. The total number of leaves ($|NodeLink|$, namely, $\langle AccessFileSet, OnePrivilege\rangle$ pairs) at a child of *Root* is upper bounded by $|FileSet|$. The upper bound can be achieved, because each one file has one privilege. We will explain how to merge the leaves that have the same privilege. Thus, the number of total possible leaves for a node would be the number of possible privileges (namely, four, they are *Authorize*, *Update*, *Modify*, and *Read*).

Principle 2. The merging set of *AccessFileSet* in leaves at a child of *Root* is upper bounded by *FileSet*.

Principle 3. The merging set of *AccessFileSet* in leaves at all children of *Root* is upper bounded by *FileSet*.

Principle 4. The number of children of *Root* (namely, the number of users added by *Root*) has no limit. It depends on the editing logic of shared files; thus, the number of members in an initial group could be very large.

Principle 5. The total number of leaves on a child of a node (not *Root*) is upper bounded by $|AccessFileSet|$ in the node's leaf whose *OnePrivilege* is *Authorize*. It is similar to Principle 1.

Principle 6. The merging set of *AccessFileSet* in leaves on a child of a node (not *Root*) is upper bounded by *AccessFileSet* in the node's leaf whose *OnePrivilege* is *Authorize*. It is similar to Principle 2.

Principle 7. The merging set of *AccessFileSet* in leaves at all children of a node (not *Root*) is upper bounded by *AccessFileSet* in the node's leaf whose *OnePrivilege* is *Authorize*. It is similar to Principle 3.

Principle 8. The number of children of a node has no limit. It is similar to Principle 4.

Principle 9. The birth sequence of nodes in the tree is from an upper layer to a lower layer (for a subtree). That is, the growth of the tree is from upper to lower. Thus, users are added into the tree from upper to lower sequentially.

Principle 10. Usually, a physical person has one *Token* (and corresponding *UserName*). Our scheme has the flexibility that one physical person can be given more than one *Token*. There is another way to achieve the functionality that more *Token*s are held by one physical person. That is, usually *Token*s in $\langle UserName, Token\rangle$ are distinct for different *UserName*, but same *Token* for different *UserName* will achieve the goal

of more *Token*s at a physical person. (The reason is that *Token* is used for the authentication of access control.)

Principle 11. One node cannot have two parents (fathers). That is, if one *UserName* with corresponding *Token* is created and assigned, the *UserName* will not be reused by other nodes for assigning. This principle will simplify the tree. Nonetheless, the F2AC can also achieve that one node has more than one father. The method is that you add the node at first and let two children's *Token* be the same.

Principle 12. The deletion of a user can be done by the deletion of a node. If the node has children (namely, not a leaf node), the deletion of the node will let the node's father be the node's children's father. It seems to replace the team leader. If the node has no children (at lowest layer), the deletion of the node will be done directly. Certainly, the deletion of a node will remove the node's linked leaves together.

Principle 13. The modification of user's privilege can be done by modifying the *OnePrivilege* on the linked leaves of the node. *Update*, *Modify*, and *Read* can be changed into each other among them. *Update*, *Modify*, and *Read* can be changed into *Authorize*. We reserve the flexibility that *Authorize* can also be changed into *Update*, *Modify*, and *Read*, but it may need to delete subtree of this node for the consistence of principles.

Principle 14. For simplicity and better understanding, we do not include the further accessing conditions, denoted as C, in the design of linked leaves. Indeed, it is without loss of generality. The C can be considered as an extension of P, as C specifies more requirements in accessing policies. Therefore, *NodeLink* can be extended into $\langle AccessFileSet, OnePrivilege, ExtraConditions\rangle$ if required.

4.7. Further Extension for Lightweight. As stated in aforementioned Principle 1, linked leaves of a node may be too many. We propose several algorithms to merge leaves in the tree structure or merge records in *ACL*. We take *ACL* as an example to describe our algorithms. The algorithm for tree structure can be done accordingly.

ACL has four tuples $\langle File_*, User_*, P, C\rangle$. We change it into $\langle User_*, File_*, P \parallel C\rangle$ for matching with tree model. Next, we will propose algorithms to accelerate the retrieval delay in *ACL* (and also in the tree).

$\langle User_*, File_*\rangle$ is cartesian product (simply speaking, many-to-many). $\langle File_*, P\rangle$ is many-to-one. $\langle User_* \times File_*, P\rangle$ is many-to-one. Thus, we can combine files that have the same privilege for a single user as one record to decrease the number of records in *ACL* (and retrieval delay). It is the so-called $\langle AccessFileSet\rangle$ in the tree. For each user (except for original user), there are at most four file sets, as the maximal number of privileges are four (namely, *Authorize*, *Update*, *Modify*, and *Read*). Thus, the linked leaves for each node will be at most four.

If we look *AccessFileSet* as a single item in the field $File_*$, the number of records in *ACL* can further be decreased by merging users. $\langle User_*, File_* \times P\rangle$ is cartesian product. As the intersection of *AccessFileSet* is empty set, $File_* \times P$ may be the same for different users. In this situation, those $User_*$ fields can be combined into one item. This can be looked as the merging of nodes that have the same linked leaves in the tree.

The retrieval of *ACL* can be considered as the traverse in the tree. After a user logs in and presents her token, the node is determined in the tree. The linked leaves can be fast fetched for listing (or looking up) files that can be accessed together with corresponding privileges. Also, the node's children can be listed for addition and removal, if applicable. Thus, the proposed tree-based model provides critical support for algorithms (and functions) in the access control mechanism.

5. Conclusions

In this paper, we proposed a lightweight, fine-grained, and flexible scheme, called F2AC, for access control in multiple-party file editing and sharing in mobile cloud computing. F2AC can support dynamically adding and deleting users in an ad hoc group, privilege self-defining as a creator's proxy or team leader, transitively authorizing privileges for members in subteams, transitively revoking privileges, and separating of access authentication from system authentication. The directed tree with linked leaf (leaves) model is proposed for lightweight implementation and verification. The leaf merging and node merging are described for lightweight storage and fast retrieval of privileges. The future work could be the evaluation of linked leaf model in some mainstream cloud services such as Apple iCloud, Baidu Cloud, and Alibaba Cloud.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The research was financially supported by the National Natural Science Foundation of China under Grant nos. 61170217, 61502440, 61301166, and 61363069, the Fundamental Research Funds for the Central Universities, and China University of Geosciences (Wuhan) (Grants nos. CUGL150831 and CUGL150416).

References

[1] X. Li and X. Zhao, "Survey on access control model in cloud computing environment," in *Proceedings of the International Conference on Cloud Computing and Big Data (CloudCom-Asia '13)*, pp. 340–345, IEEE, Fuzhou, China, December 2013.

[2] X. Yao, X. Han, and X. Du, "A lightweight access control mechanism for mobile cloud computing," in *Proceedings of the IEEE Conference on Computer Communications Workshops (INFOCOM '14)*, pp. 380–385, April 2014.

[3] T. Jung, X.-Y. Li, Z. Wan, and M. Wan, "Control cloud data access privilege and anonymity with fully anonymous attribute-based encryption," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 1, pp. 190–199, 2015.

[4] S. Ruj, M. Stojmenovic, and A. Nayak, "Decentralized access control with anonymous authentication of data stored in clouds," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 2, pp. 384–394, 2014.

[5] A. Ghafoor, M. Irum, and M. Qaisar, "User centric access control policy management framework for cloud applications," in *Proceedings of the 2nd National Conference on Information Assurance (NCIA '13)*, pp. 135–140, IEEE, Rawalpindi, Pakistan, December 2013.

[6] Y. Tang, P. P. C. Lee, J. C. S. Lui, and R. Perlman, "Secure overlay cloud storage with access control and assured deletion," *IEEE Transactions on Dependable and Secure Computing*, vol. 9, no. 6, pp. 903–916, 2012.

[7] M. Habiba, M. R. Islam, and A. B. M. S. Ali, "Access control management for cloud," in *Proceedings of the 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom '13)*, pp. 485–492, IEEE, Melbourne, VIC, Australia, July 2013.

[8] Z. Wan, J. Liu, and R. H. Deng, "HASBE: a hierarchical attribute-based solution for flexible and scalable access control in cloud computing," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 743–754, 2012.

[9] M. Hajivali, M. T. Alrashdan, F. Fatemi Moghaddam, and A. Z. M. Alothmani, "Applying an agent-based user authentication and access control model for cloud servers," in *Proceedings of the International Conference on ICT Convergence (ICTC '13)*, pp. 807–812, IEEE, Jeju, South Korea, October 2013.

[10] P. Ortiz, O. Lazaro, M. Uriarte, and M. Carnerero, "Enhanced multi-domain access control for secure mobile collaboration through linked data cloud in manufacturing," in *Proceedings of the IEEE 14th International Symposium and Workshops on a World of Wireless, Mobile and Multimedia Networks (WoWMoM '13)*, pp. 1–9, IEEE, Madrid, Spain, June 2013.

[11] Z. Lv, C. Hong, M. Zhang, and D. Feng, "A secure and efficient revocation scheme for fine-grained access control in cloud storage," in *Proceedings of the 4th IEEE International Conference on Cloud Computing Technology and Science (CloudCom '12)*, pp. 545–550, Taipei, Taiwan, December 2012.

[12] R. Yang, C. Lin, and Y. Jiang, "Enforcing scalable and dynamic hierarchical access control in cloud computing," in *Proceedings of the IEEE International Conference on Communications (ICC '12)*, pp. 923–927, IEEE, Ottawa, Canada, June 2012.

[13] Z. Shen, J. Shu, and W. Xue, "Keyword search with access control over encrypted data in cloud computing," in *Proceedings of the 22nd IEEE International Symposium of Quality of Service (IWQoS '14)*, pp. 87–92, Hong Kong, May 2014.

Hindawi Publishing Corporation
Mobile Information Systems
Volume 2016, Article ID 5032365, 12 pages
http://dx.doi.org/10.1155/2016/5032365

Research Article

Seamless Guidance System Combining GPS, BLE Beacon, and NFC Technologies

Rung-Shiang Cheng,[1] Wei-Jun Hong,[2] Jheng-Syun Wang,[1] and Kawuu W. Lin[2]

[1]*Department of Computer and Communication, Kun Shan University, No. 195, Kunda Road, Yongkang District, Tainan City 710, Taiwan*
[2]*Department of Computer Science and Information Engineering, National Kaohsiung University of Applied Sciences, No. 415, Jiangong Road, Sanmin District, Kaohsiung City 807, Taiwan*

Correspondence should be addressed to Rung-Shiang Cheng; rscheng@mail.ksu.edu.tw

Received 2 October 2015; Accepted 7 December 2015

Academic Editor: Jong-Hyouk Lee

Users rely increasingly on Location-Based Services (LBS) and automated navigation/guidance systems nowadays. However, while such services are easily implemented in outdoor environments using Global Positioning System (GPS) technology, a requirement still exists for accurate localization and guidance schemes in indoor settings. Accordingly, the present study proposes a system based on GPS, Bluetooth Low Energy (BLE) beacons, and Near Field Communication (NFC) technology. Through establishing graphic information and the design of algorithm, this study develops a guidance system for indoors and outdoors on smart phones, wishing to give user perfect smart life through this system. The proposed system is implemented on a smart phone and evaluated on a student campus environment. The experimental results confirm the ability of the proposed app to switch automatically from an outdoor mode to an indoor mode and to guide the user to requested target destination via the shortest possible route.

1. Introduction

According to statistics published by the International Telecommunication Union (ITU), the number of mobile devices in the world reached 6.835 billion at the end of 2013. Furthermore, the number of devices is still growing. As wireless technology continues to improve and wireless networks are ever more extensively deployed, the feasibility of developing Location-Based Services (LBS) has attracted growing interest [1]. LBS have many advantages from a user perspective, including convenience, efficiency, and fun. As a result, they are now widely applied in social networks, traffic and geographic search systems, and even public safety [2, 3]. Of the many functions offered by LBS, those of positioning localization and guidance are some of the most useful. According to previous research, adults spend around 86.9% of their time indoors, 5.5% in vehicles, and 7.6% outdoors [4]. Thus, in realizing seamless LBS applications, it is necessary to develop localization and guidance schemes capable of working in both indoor and outdoor environments and switching between the two modes automatically as required.

Global Positioning System (GPS) technology is used widely in the navigation, tourism, measurement, and engineering fields. However, the success of GPS depends on a strong signal between the user and the navigational satellite. Thus, while GPS functions well in open outdoor environments, its performance suffers dramatically in mountainous areas or build-up urban areas. Furthermore, the signals are unable to penetrate through building structures, and hence GPS is of only limited use in indoor environments.

The literature thus contains various alternative proposals for performing indoor localization. For example, in [5, 6], the user position is estimated using a wireless communication signal, while, in [7], a visible light communication system is used. The authors in [8] performed indoor positioning using radiofrequency identification (RFID) tags. In [9], user

positioning was performed by mapping the activities of the user to the positions in the indoor environment at which these activities were known to be performed. Finally, in [10], multiple indoor positioning technologies were combined in order to improve the localization accuracy.

Although mobile devices are invaluable in daily life nowadays, their usefulness is limited by their short battery lives, which prompts the need for frequent recharging. To address this problem, many mobile devices use Bluetooth Low Energy (BLE) technology to realize wireless communication connections. BLE has many advantages as a connection technology, including a stable signal, an ease of distribution, a low cost, and widespread compatibility with existing wireless devices. Furthermore, BLE beacons have an operating life of several months using only a simple button cell battery [11]. As a result, BLE has significant potential as an enabling technology for indoor LBS applications.

Owing to indoor environment being a complicated space, GPS signals can not be used and the spreading of wireless is easily affected by the interior structure of the building; therefore, it urgently needs precise technologies and a lot of spherical auxiliary electronic devices, leading to the difficulty in indoor precise positioning. Thus, for many years, these well-known positioning technologies such as GPS or Google Map services are usually applied in public construction fundamental facilities, road, or outdoor large area; however, along with the urbanization of living environment, there are more and larger complex buildings, increasing the needs for indoor positioning. Near Field Communication (NFC) is a short-range wireless connectivity standard which enables communications to be achieved between devices simply by touching them together or bringing them into very close proximity of one another (typically, less than 10 cm). NFC has found widespread use nowadays for such applications as loyalty schemes, home healthcare, public transport payment, ticketing, mobile workforce management, and so on. With the ability it provides to infer the user position with an ultra-high degree of precision, NFC also has significant potential for indoor localization purposes.

Thus, to enhance the application of positioning system and enable smart phones to become key equipment of augmented humanity which can effectively enhance the convenience, this study is based on short-distance wireless communication technology, combining with NFC, BLE beacon, and GPS technology to develop a positioning guidance system which is suitable for indoor buildings. This study combines GPS, BLE, and NFC technologies to realize a seamless indoor-outdoor user localization and guidance app suitable for implementation on a smart mobile device. In the proposed scheme, user localization is performed using conventional GPS technology in the outdoor environment. However, when the user enters an indoor space, the app switches automatically to an indoor mode, and user positioning is performed by means of BLE beacons and NFC tags. The positioning information obtained via the various technologies is combined with map information (outdoor and indoor) to realize a guidance system capable of leading the user to the requested target destination via the shortest available route.

The design of the system includes "storage data design," "positioning method design," "algorithm design," and "graphic design." In between, for indoor map establishment, this study designs a method which can assist in establishing positioning point storage function through wireless signal exploration (indoor space area as unit positioning base). Outdoor map uses GPS as positioning base to obtain the location information, using route map to display the guidance results. When users are in outdoors, this app will automatically adopt "outdoor mode" to obtain GPS positioning to provide users with guidance information.

The remainder of this paper is organized as follows. Section 2 reviews the GPS, BLE, and NFC technologies used for localization purposes in this study and introduces the path-finding algorithm used to realize the indoor guidance system. Section 3 describes the system framework and implementation. Section 4 presents and discusses the experimental evaluation results. Finally, Section 5 provides some brief concluding remarks.

2. Background Knowledge

This section commences by describing the GPS, BLE beacon, and NFC technologies used in the present study to develop the proposed positioning and guidance system. The shortest-path algorithm used to accomplish indoor guidance is then briefly introduced. Figure 1 presents a schematic illustration of the respective communication distances of the GPS, beacon, and NFC technologies.

2.1. GPS. GPS is a middle-distance global tracking satellite guidance system with a coverage area of more than 98% of the earth's surface. GPS can be used by any enabled device capable of receiving its signals and has the advantage of anonymity in that the user's position is not recorded as part of the communication process. However, GPS relies on the availability of a clear Line of Sight (LOS) between the user device and the satellite system. As a result, it provides only a limited positioning capability in indoor environments [12].

2.2. BLE. Bluetooth Low Energy (BLE) is a communication standard designed to enable short-range wireless devices to operate for months or even years on a single coin cell battery. When combined with beacon technology, BLE provides a highly effective method for estimating the position of the user relative to certain predefined monitoring spots. BLE operates over a distance of up to 50 m and provides the means to customize the LBS offered to the user based on their physical location. For example, certain ads can be pushed to the user device as the user approaches a particular sales counter in a store. Similarly, the user may be presented with different notification messages and application events as he or she moves across the boundary separating one monitored area from another.

The literature contains various proposals for integrating the BLE standard with beacon technology in order to support user localization, including iBeacon [13], Gimbal [14], and AltBeacon [15]. The beacons used in such systems periodically broadcast a wireless radio signal advertising

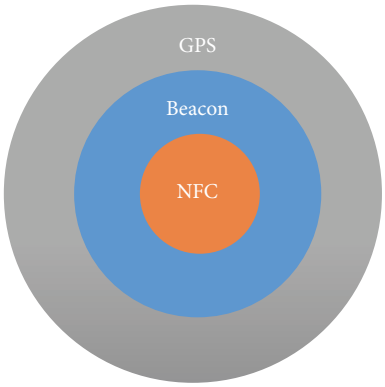


Figure 1: Communication distances of three positioning technologies.

Figure 2: Gimbal beacon.

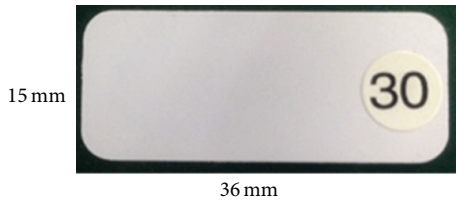


Figure 3: NFC tag.

their presence. As described above, BLE operates over a range of up to 50 m. Hence, in the event that the signal is detected (sighted) by a proximity-enabled user device, the user position can be inferred with an error of no more than 50 m. As a result, BLE/beacon technology provides a low-cost and energy-efficient solution for performing user localization with a medium degree of accuracy. The localization system proposed in the present study utilizes the Gimbal Series 10 beacon produced by Qualcomm (see Figure 2). The periodic message broadcast by the beacon contains many items of information, including the Factory ID, RSSI value, name, battery capacity, and temperature. The Factory ID uniquely identifies the beacon. Thus, by storing the Factor IDs and physical beacon locations in a database, the ID information contained in the message provides an efficient and reliable method for estimating the position of the user.

2.3. NFC. Near Field Communication (NFC) is an ultra-short distance wireless communication technology based on radiofrequency identification (RFID). NFC utilizes signal attenuation technology to enable devices to conduct noncontact point-to-point data transmissions over distances of up to approximately 10 cm (3.9 inches). NFC is currently used for such applications as automated payment, ticketing, loyalty schemes, and so forth. However, with its high bandwidth and low energy consumption [16], NFC (see Figure 3) also has significant potential for highly precise indoor positioning. As with the Gimbal beacon, each NFC chip has a unique ID number assigned to it by the manufacturer. Thus, by associating the ID with a physical location and storing this information in a database, the position of the user can be inferred with an extremely high degree of precision each time a sensing event occurs.

2.4. Shortest-Path Algorithm. Determining the shortest route between a start point and a target end point given the availability of multiple paths between them is a common problem in many walks of life [17]. The guidance system proposed in this study utilizes the algorithm proposed by Dijkstra [18]. Since, of the various algorithms available, it has the advantages of concise algorithm, the optimal solution can be obtained.

3. System Framework and Implementation

The app proposed in this study provides the user with a seamless positioning and guidance service as he or she moves from an outdoor environment to a target destination in an indoor environment or vice versa. In other words, the app switches automatically not only from an outdoor mode to an indoor mode, but also from an indoor mode to an outdoor mode. As described in Introduction, positioning in the outdoor environment is performed using conventional GPS technology, while that in the indoor environment is performed using BLE beacon and NFC technologies. For both environments, the guidance function is achieved using map information stored in a remote server and downloaded to the user device as required. For illustration purposes, the present study considers the localization/guidance problem for the case of a student campus environment containing many buildings scattered over a wide geographic area with many floors and rooms within each building. As described in the following sections, the system framework comprises four design components, namely, (1) storage data design, (2) positioning method design, (3) shortest-path algorithm design, and (4) map structure design.

3.1. Storage Data Design. To simplify the data storage and management task, four different data structures are used to support the different functionalities of the system, namely, an outdoor map structure, an indoor map structure, a beacon positioning data structure, and an NFC positioning data structure. Table 1 shows the data structure used to store and maintain the outdoor map information. As shown, the structure comprises six fields, namely, id (the primary index key); name (used to store the name of the destination); address (used to store the address of the destination); latitude and longitude (used to store the geographical coordinates

Table 1: Data format for outdoor map.

Name	Type	Remark
id	Int(11)	Primary
name	Text	Destination name
address	Text	Address
latitude	Text	Latitude
longitude	Text	Longitude
info	Text	Brief info

Table 2: Data format for indoor map.

Name	Type	Remark
id	int(11)	Primary
map_name	Text	map code
node	int(11)	node code
x	int(11)	X-axis coordinate
y	int(11)	Y-axis coordinate

of the destination); and information (used to store a brief description of the building).

Table 2 shows the data structure used to store the indoor map information. The structure contains five fields, namely, id (the primary index key); *map-name* (used to store the code of the indoor map); node (used to indicate the name of the nodes (rooms/locations) in the indoor map); and x and y (used to store the x- and y-coordinates of the nodes in the indoor map).

Table 3 shows the data structure used to store the beacon location information. As shown, the structure contains six fields, namely, *id* (the primary search index); *node* (the node identifier in the indoor map); *fid* (the unique ID assigned to the beacon by the factory); *name* (the name of the room/location in which the tag is located); *middle* (the name of the building in which the tag is located); and *far* (the name of the general campus area in which the building is located). Table 4 represents the indoor map information used to store graphic object-oriented information; clicking each point represents a position on the screen. Data list design includes id primary index key and *map_name* is the symbol of the map for the usage of server inquiry. Node is the present representing node symbol and x and y are the coordinates of the center of the node.

3.2. Positioning Method Design. As shown in Figure 4, the localization/guidance app proposed in this study resides by default in the "outdoor mode" and uses conventional GPS technology to locate the position of the user. More specifically, the system acquires the current latitude and longitude information from the GPS and uploads this information together with the Device ID to a remote server (see Figure 5). On receiving this information, the server interrogates the coordinate information and returns the appropriate outdoor map to the user device using the JSON format shown in Algorithm 1.

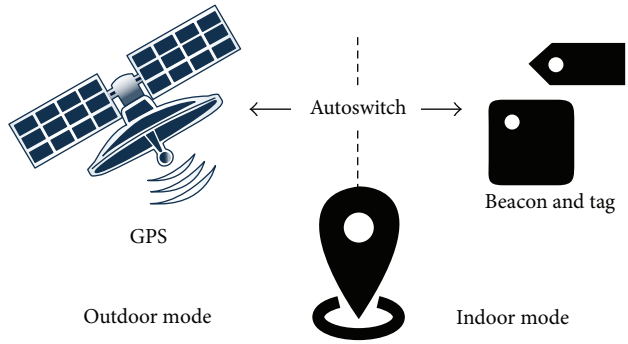


Figure 4: Automatic mode switching.

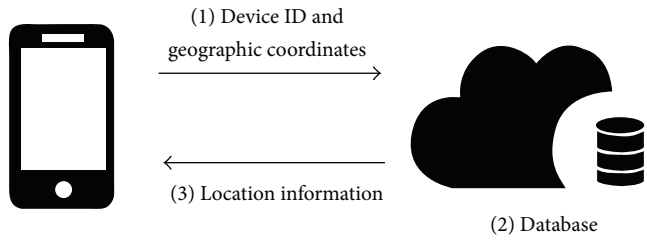


Figure 5: Outdoor map data flow.

Table 3: Beacon data format.

Name	Type	Remark
id	int(11)	Primary
node	int(11)	node
fid	text	Factory ID
name	text	Room name
middle	text	Building name
far	text	Big area name

Table 4: NFC data format.

Name	Type	Remark
id	int(11)	Primary
node	int(11)	node
uid	text	Tag uid
name	text	Room name
middle	text	Building name
far	text	Big area name

JSON message sent back includes the name, address, latitude and longitude, and building introduction of the destination whose format is shown in Algorithm 1.

When the user moves from the outdoor environment to an indoor environment, the app switches automatically to an "indoor mode" and launches an indoor positioning routine. If the user is within wireless range of a beacon, a sighting event occurs and the device uploads both its own ID and that of the beacon to the server. Utilizing the Factor ID as a key, the server retrieves the approximate location of the user and

```
{
"id": 1,
"name": "School",
"address": "No.195, Kunda Rd.,",
"latitude": "22.996175",
"longitude": "120.252970",
"info": "Kun Shan University"
}
```

Algorithm 1: JSON data format.

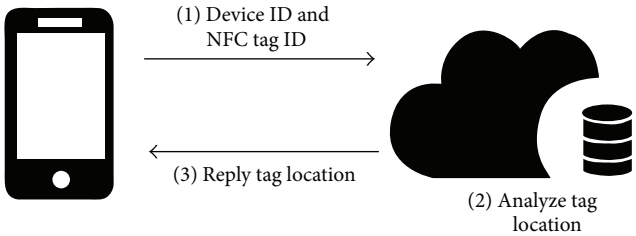


Figure 6: NFC positioning process.

returns this information to the user device. When the user's smart phone approaches a NFC tag, the user can send tag ID along with Device ID back to the server after the mobile reads the set NFC tag. The server will conduct index from the NFC data list and send the index results back the app to obtain the user's location and conduct indoor positioning as shown in Figure 6.

3.3. Shortest-Path Algorithm Design. After obtaining the user location, the app downloads the map from the server. Specifically, the app informs the server of the Device ID and the required map ID and the server searches its database for the corresponding map and returns it to the user device.

Let $G = (V, E)$ denote an indoor map, where V is the set of nodes in the map and E is the set of connecting edges (paths). In constructing the map, the server hosts a database with four columns, namely, ID (primary index key), node ID, adjacent node ID, and distance between adjacent nodes. Taking Node 2 in Figure 8 as an example, let the nodes adjacent to Node 2 be denoted as Node 0, Node 1, Node 3, and Node 4, respectively. Furthermore, let the distances of these nodes from Node 0 be equal to 4, 2, 9, and 2, respectively. The position of Node 2 relative to its neighboring nodes can therefore be expressed in the form shown in Table 5.

The node positioning information is communicated from the server to the device using the JSON format shown in Algorithm 2. Note that the information object includes such information as the name of the map, the number of nodes in the map, and the number of adjacent roads in the map. Similarly, the algorithm object includes the adjacent nodes, the costs of the adjacent paths, and the ID of the present node.

The information provided in the JSON message in Algorithm 2 provides the app with the relevant indoor shortest-path map. However, the Dijkstra shortest-path algorithm requires the input information to be presented in the

```
{
  "information": {
    "name": "Example",
    "node": 5,
    "road": 14
  },
  "algorithm": [
    {
      "neighbor": [
        0,
        1,
        3,
        4
      ],
      "cost": [
        4,
        2,
        9,
        2
      ],
      "this": 2
    }
  ]
}
```

Algorithm 2: JSON format for Node 2.

Table 5: Node positioning information.

Name	Type	Storage results
node	int(11)	2
neighbor_A	text	0
distance_A	int(11)	4
neighbor_B	text	1
distance_B	int(11)	2
neighbor_C	text	3
distance_C	int(11)	9
neighbor_D	text	4
distance_D	int(11)	2

form of a matrix. Therefore, in implementing the indoor guidance function, the JSON map information must first be converted into a matrix form. For example, the illustrative layout in Figure 8 contains five nodes and should therefore be converted to a 5×5 matrix of the form shown in Table 6.

For a large map comprising multiple nodes, creating the map and maintaining it over time as additional nodes are added or existing nodes are removed represent a huge task if performed manually. For example, for a map with n nodes, maintaining the corresponding adjacent matrix given a change in any one of the nodes incurs a time complexity of $O(n2)$. Accordingly, in the app proposed in this study, the matrix construction task is performed automatically using the function shown in (1), in which i is the number of the current node, v is the number of the adjacent node, and e is the cost of the path between them. When presented with the node map (constructed manually by the system developer),

Table 6: Dijkstra adjacent matrix.

Node	0	1	2	3	4
0	0	8	4	INF	INF
1	8	0	2	7	INF
2	4	2	0	9	2
3	INF	7	9	0	2
4	INF	INF	2	6	0

the matrix construction algorithm takes the current node and adjacent node information as an input and uses (1) to automatically generate the corresponding $n \times n$ adjacent matrix with a time complexity of $O(n)$:

$$\text{graph}[i].\text{adjacentEdge}(v, e). \tag{1}$$

For example, taking Node 2 in Figure 8 for illustration purposes once again, the relation between Node 2 and its adjacent nodes has the form shown in (2). Taking the i, v and e information given in (2), the algorithm automatically constructs the matrix shown in Figure 9:

$$\begin{aligned} &\text{graph}[2].\text{adjacentEdge}(0, 4), \\ &\text{graph}[2].\text{adjacentEdge}(1, 2), \\ &\text{graph}[2].\text{adjacentEdge}(3, 9), \\ &\text{graph}[2].\text{adjacentEdge}(4, 2). \end{aligned} \tag{2}$$

Taking Figure 8 as example, the app reads the above adjacent relation and automatically converts to the matrix adopted by the algorithm as shown in Figure 9. It inputs the converted matrix into Dijkstra's algorithm to estimate the shortest path by Dijkstra's algorithm.

Dijkstra's shortest-path algorithm calculates the shortest path from one point in a node set to any other node in the same set. Assume that Node 2 in Figure 8 is chosen as the start point and Node 3 is chosen as the destination. Three feasible paths exist between the two nodes, namely, a direct path from Node 2 to Node 3 and two two-node paths via Nodes 1 and 4, respectively. The three paths have total distances of 9, 9, and 8, respectively. Thus, in accordance with Dijkstra's algorithm, path 2-4-3 is chosen as the shortest path (see Figure 10).

3.4. Map Structure Design. In general, the success of any app is determined to a large extent by the appearance and intuitiveness of its graphical user interface (GUI). For a guidance system such as that proposed in the present study, a pictorial map with too much detailed information will serve simply to confuse the user. Consequently, in the proposed app, the indoor and outdoor maps are presented in the form of metro-like maps, in which the key locations (e.g., buildings, offices, classrooms, and toilets) are represented as nodes and the distances between them are indicated by numerals placed alongside the corresponding paths (see Figure 11).

As discussed above, having determined the user's current location, the app requests the appropriate map from the server and then stores the received map in the device (see Figure 7). Notably, the map is downloaded in its entirety

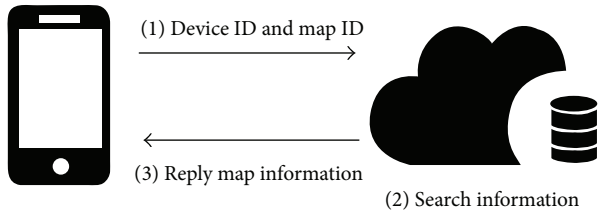


Figure 7: Map algorithm download process.

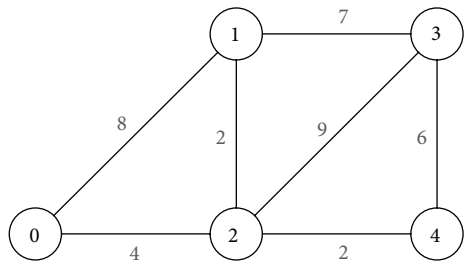


Figure 8: Node layout scheme.

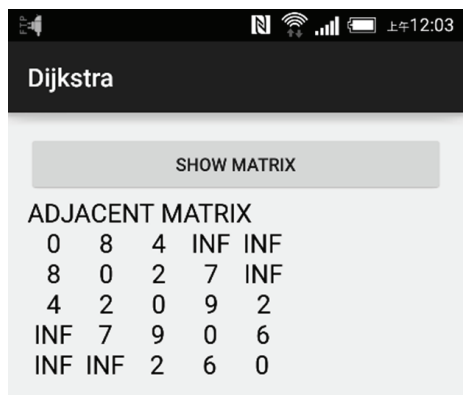


Figure 9: Automatic conversion of adjacent node relations to matrix form.

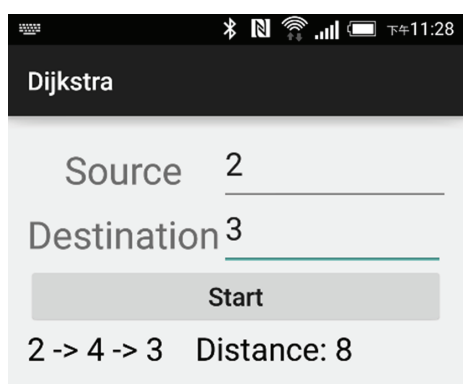


Figure 10: Output of Dijkstra shortest-path algorithm.

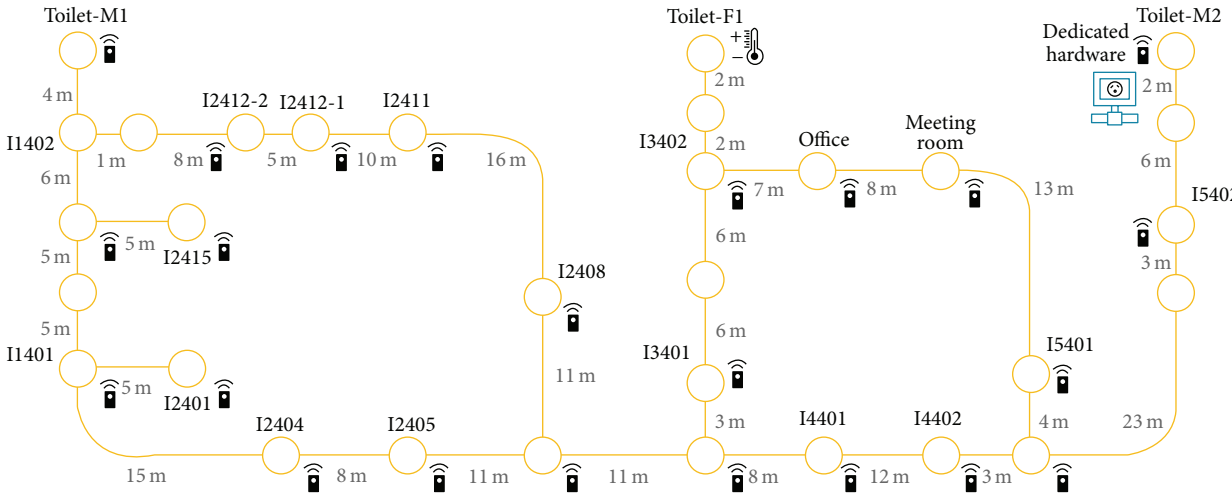


Figure 11: Map representation.

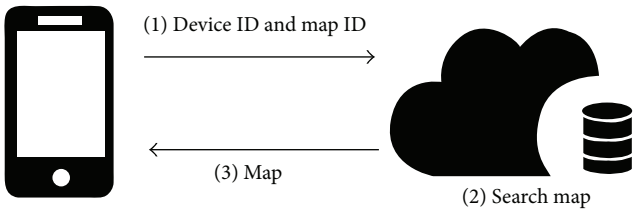


Figure 12: Flow chart showing map download process.

(Figure 12), and hence the need for repeated download events is avoided. Having downloaded the map, the user then selects the target destination (i.e., node) for which they require routing information. To ensure smooth node selection, the app stores an acceptable touch range error in addition to the coordinates of each node center. The corresponding JSON message exchange between the app and the server thus has the form shown in Algorithm 3.

When the user touches the screen to select a particular destination node, the app determines the intended node in accordance with (3), in which X_1 and Y_1 are the stored coordinates of the node center, X_2 and Y_2 are the coordinates of the point at which the user touches the screen, and r is the allowable touch range error:

$$|X_1 - X_2|^2 + |Y_1 - Y_2|^2 < r^2. \quad (3)$$

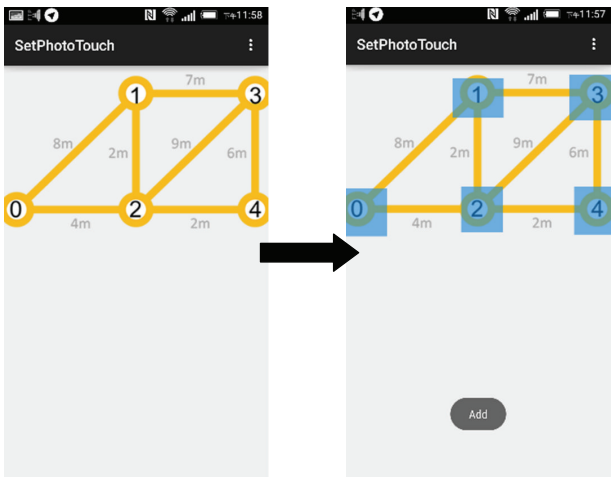
To support the node selection process described above, the center coordinates of each node must first be established and stored in the device. Estimating the node centers one by one and inputting them into the database are a huge task if performed manually. Accordingly, this study provides a tool for automating the node input process by enabling the map manager to simply mark the scope of each node on the map (see Figure 13). Having marked the scope, the tool estimates

```
{
"information": {
  "name": "Example",
  "photo": {
    "width": 1503,
    "height": 947,
    "radius": 100
  }
},
"algorithm": [
  {
    "touch": {
      "X": 110,
      "Y": 775
    },
    "this": 0
  },
  {
    "touch": {
      "X": 750,
      "Y": 150
    },
    "this": 1
  }
 ]
}
```

Algorithm 3: JSON format for map output information.

the coverage scope between the center of the nodes and the nodes on the map and uploads the related information to the database.

Having established the user's present location and his or her intended destination, the app invokes Dijkstra's shortest-path algorithm and marks the suggested route pictorially on the node map, as shown in Figure 14.



(a) (b)

Figure 13: Node coordinate definition process.

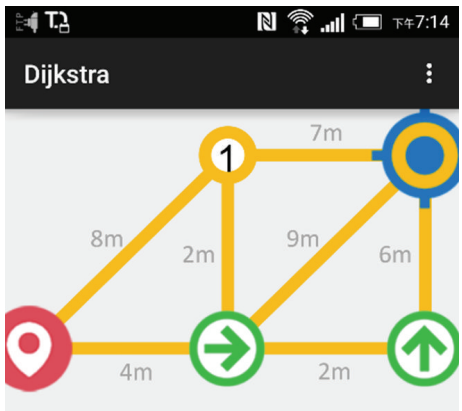


Figure 14: Suggested route interface.

4. Implementation Results

4.1. Beacon Signal Measurement. Localization experiments were performed using three commercial mobile phones, namely, HTC One, Sony Xperia Z1 Compact, and Samsung Galaxy S4. Each device was placed at three different distances from a beacon (i.e., 1, 2, and 4 m), and the distance measurements obtained by the device were then sampled 100 times over a 30-minute interval.

The corresponding results are presented in Figures 15–17. For a device-to-beacon distance of 1 m, the distance measurements obtained by the Samsung and HTC devices deviate quite significantly from the actual distance and vary markedly over the sampling period (see Figure 15). By contrast, the measurement results obtained using the Sony device are both closer to the true value and more stable. For a device-to-beacon distance of 2 m, the measurements obtained using the HTC and Samsung devices vary significantly over the 30-minute interval. However, the Sony device again achieves an accurate and relatively stable measurement performance (see Figure 16). Finally, for a measurement distance of 4 m, all of the devices achieve a poor localization performance (see Figure 17).

In general, the results presented in Figures 15–17 show that none of the devices provides an absolutely precise localization performance. Hence, in implementing the proposed app, the measurement distances obtained by the devices were

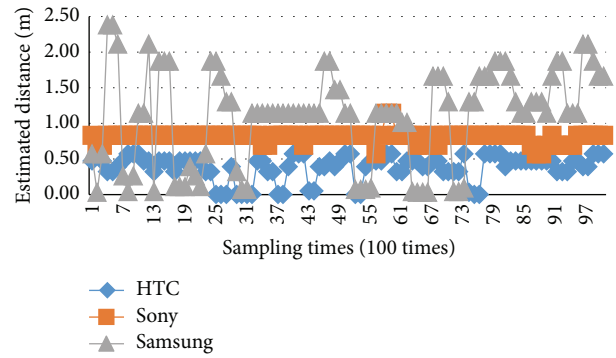


Figure 15: Signal measurement results at distance of 1 m from beacon.

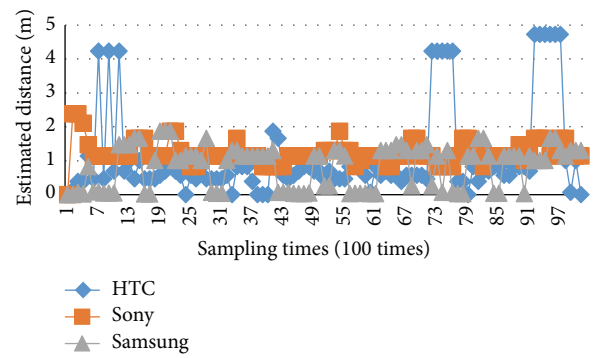


Figure 16: Signal measurement results at distance of 2 m from beacon.

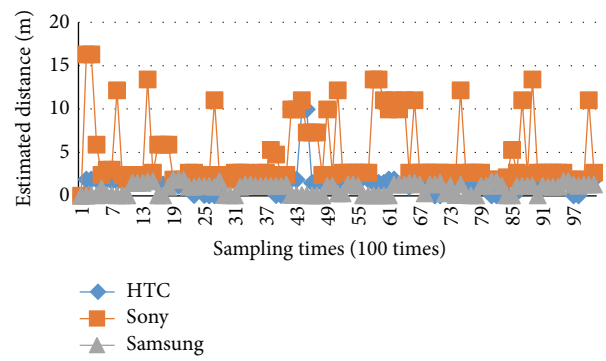


Figure 17: Signal measurement results at distance of 4 m from beacon.

quantified using the proximity value provided by the beacon, namely, a numerical value in the range of 1~3. A value of 1 was taken to indicate a close proximity of the device to the beacon, and hence the beacon location was used as a reference in estimating the current location measurement. By contrast, a value of 3 was taken to indicate a greater distance of the device from the beacon, and hence the location measurement obtained using the device was taken as an approximate value only.

4.2. System Implementation and Function Display. Performance evaluation trials were performed on a university campus in Taiwan, with the aim being to guide the user to a specific target in the building shown in Figure 18. As shown, 7 beacons and 31 NFC tags were placed at appropriate points throughout the experimental field, for example, on the doors of the main rooms in the building, at the entrances to staircases, and at forks or corners in the corridors. Due to their lower cost, the NFC tags greatly outnumbered the beacons (Figure 19) and were placed with an approximate spacing of 2~15 m.

When the app is launched, the system first checks the status of the communication services of the user device and issues a notification message if required (see Figure 20, e.g.). The system then performs a localization routine to establish the user position. If the app senses a GPS signal, it loads the outdoor map and marks the user location accordingly. By contrast, if the app detects a beacon or NFC signal, it loads the appropriate indoor map and again marks the user location as appropriate. Having received the map (indoor or outdoor), the user selects the destination node (e.g., a campus building or a room within the present building), and the app launches the shortest-path route discovery routine and marks the identified route on the map accordingly.

As shown in Figure 21(a), if the user is outdoors, the app automatically lists the main destinations within the closest building (Figure 21(b)) and indicates the user's location on the map (Figure 21(c)). The app then asks the user if a guidance function is required (Figure 22). If the user requests guidance, the app searches for the shortest-path to the selected destination and then plots the route on the map (Figure 23).

4.3. Experimental Design and Results. The performance of the proposed app was evaluated by comparing the time spent by four users in finding their way from the main campus entrance to a particular classroom within a certain building with and without the assistance of the localization/guidance system, respectively. In performing the experiments, the process of navigating to the classroom was separated into six steps, namely, (1) looking for campus map; (2) finding the target building; (3) searching for indoor floor layout; (4) reaching the floor; and (5) reaching the destination. Thus, we record the spent time in various steps as reference. As shown in Figures 24 and 25, the total time spent by each user in reaching the target was divided into four separate times, namely, the time taken in moving from the main campus gate to the first destination sign at Spot A; the time spent in moving from Spot A to the target building (Spot B); the time spent in walking from Spot B to the destination sign located in the building at Spot C; and the time taken in moving from Spot C to the target classroom at number I4401.

Tables 7(a) and 7(b) show the timing results obtained for the four users. Note that Users 1 and 2 performed the search

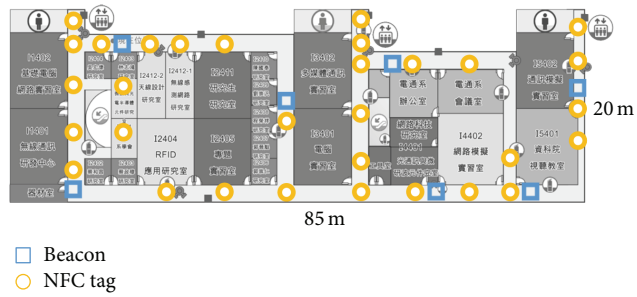
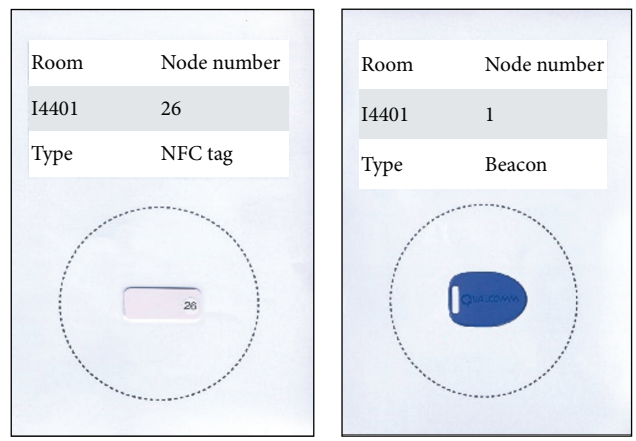


Figure 18: IoT building in the Department of Computing and Communication at Kun Shan University, Taiwan.



(a) Place NFC tag (b) Place beacon device

Figure 19: NFC tag and beacon device.

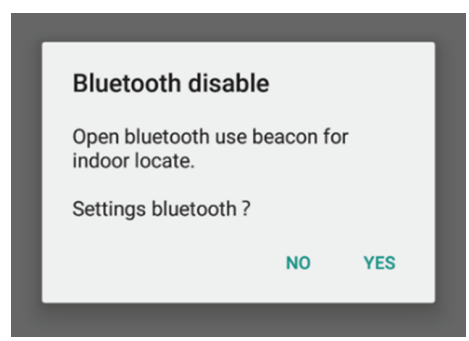
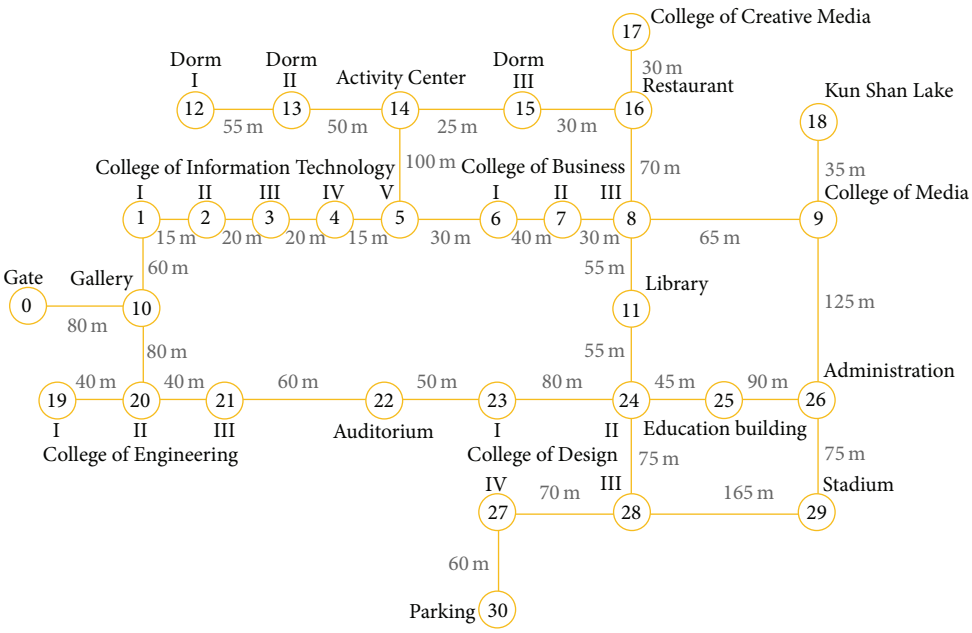
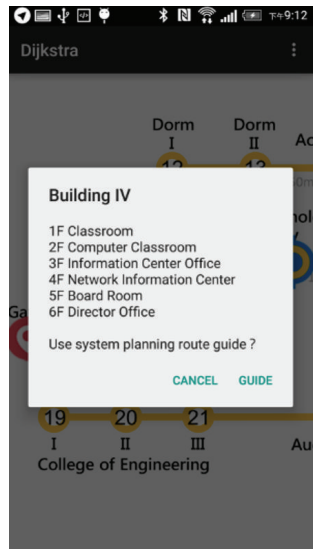


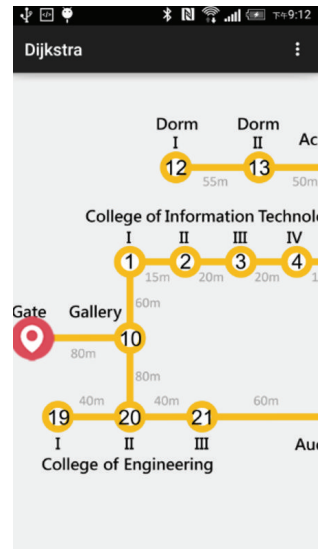
Figure 20: Open Bluetooth notification.



(a) Outdoor map overview



(b) Nearby building information



(c) Position of user

Figure 21: User located outside.

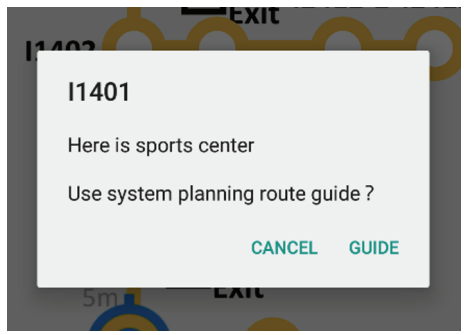
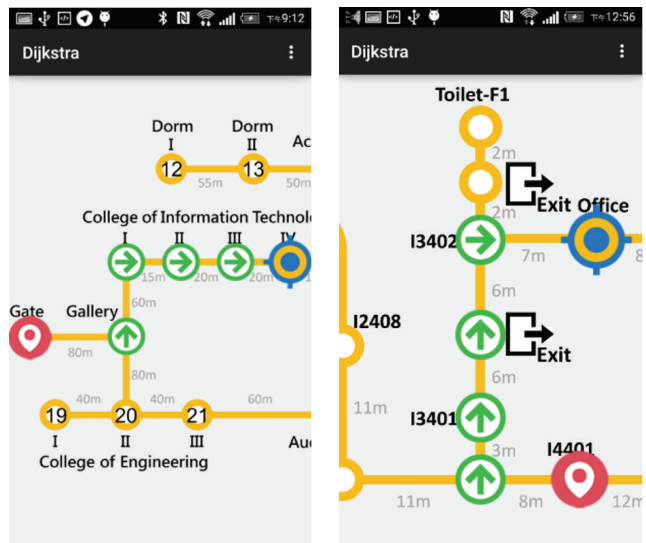


Figure 22: Guidance notification.

process in a nonassisted manner, while Users 3 and 4 both used the app. As shown, Users 1 and 2 completed the search process in 281.7 s and 333.9 s, respectively (i.e., an average search time of 307.8 s). Since Users 3 and 4 used the guidance app, they did not need to locate the direction signs at Spots A and C, respectively. Consequently, the total search times for the two users were just 206.48 s and 202.3 s, respectively (i.e., an average search time of 204.39 seconds). In other words, the mean time of the assisted users was 33.59% shorter than that of the two nonassisted users. Moreover, taking the search time of User 1 as a reference, the search times of Users 3 and 4 were reduced by 26.8% and 28.18%, respectively.



(a) Outdoor navigation aid (b) Indoor navigation aid

Figure 23: Navigation aid.

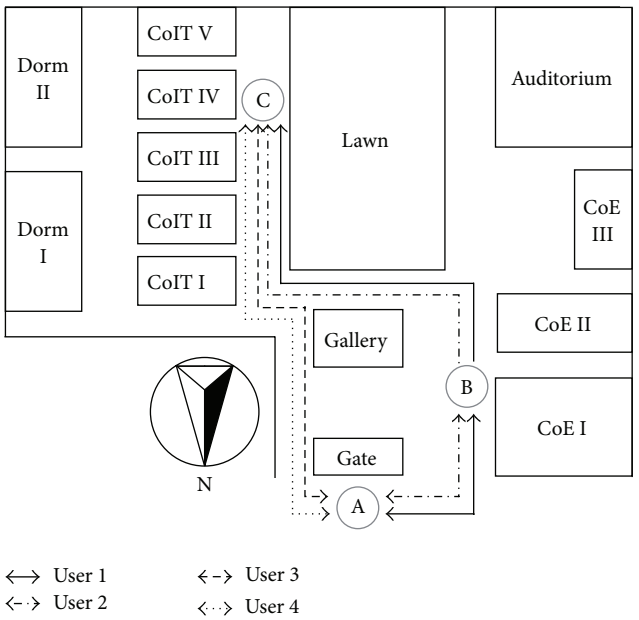


Figure 24: Experimental path (outdoor).

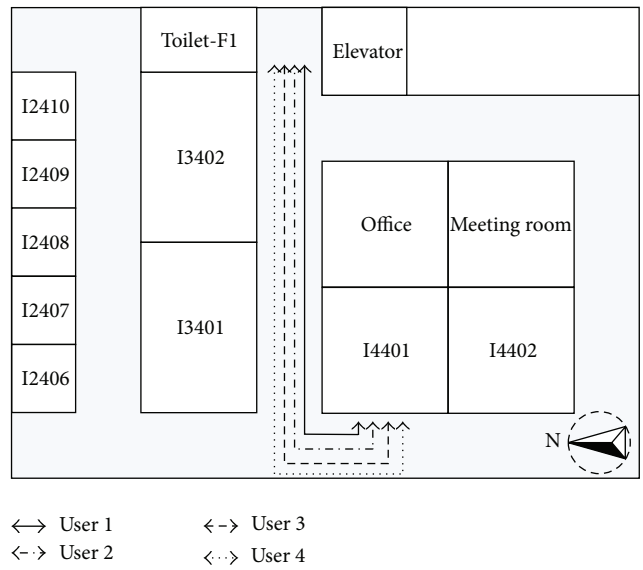


Figure 25: Experimental path (indoor).

5. Conclusion

With the emergence of Location-Based Services (LBS), the need to locate the position of the user with a high degree of accuracy has emerged as an important concern. Accordingly, this study has proposed an app based on GPS, Bluetooth beacon, and NFC technology for providing both a user localization service and an automatic guidance function. Importantly, the app functions in both outdoor and indoor environments and thus provides a seamless localization/guidance function as the user moves from one environment to the other. The feasibility of the proposed system has been demonstrated by means of localization and guidance tests on a typical student campus building.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Table 7: Times spent by four users in completing each stage of search process.

(a) Search times of four users (unit: s)

Name	Spot A	Spot B	Spot C	Spot D
Subject 1	43.87	113.93	118.97	4.93
Subject 2	23.4	153.6	128	28
Subject 3	N/A	109.43	N/A	97.05
Subject 4	N/A	129.9	N/A	72.4

(b) Total search times of four users and average saved time (unit: s)

Name	Total	Percentage of time saving
Subject 1	281.7	0%
Subject 2	333.9	−18.53%
Subject 3	206.48	26.8%
Subject 4	202.3	28.18%

Acknowledgment

The authors would like to thank the National Science Council, Taiwan, for the financial support of this study under Contract no. MOST 103-2627-E-168-001.

References

[1] S. Ray, R. Blanco, and A. K. Goel, "Supporting location-based services in a main-memory database," in *Proceedings of the 15th IEEE International Conference on Mobile Data Management (MDM '14)*, pp. 3–12, Brisbane, Australia, July 2014.

[2] A. Chandra, S. Jain, and M. A. Qadeer, "GPS locator: an application for location tracking and sharing using GPS for Java enabled handhelds," in *Proceedings of the International Conference on Computational Intelligence and Communication Networks (CICN '11)*, pp. 406–410, IEEE, Gwalior, India, October 2011.

[3] F. Liu and Z. Yang, "Study on applications of LBS based on electronic compass," in *Proceedings of the 5th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '09)*, pp. 1–4, Beijing, China, September 2009.

[4] N. E. Klepeis, W. C. Nelson, W. R. Ott et al., "The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants," *Journal of Exposure Analysis and Environmental Epidemiology*, vol. 11, no. 3, pp. 231–252, 2001.

[5] I. Yamada, T. Ohtsuki, T. Hisanaga, and L. Zheng, "An indoor position estimation method by maximum likelihood algorithm using RSS," in *Proceedings of the Annual Conference SICE*, pp. 2927–2930, IEEE, Takamatsu, Japan, 2007.

[6] J.-S. Leu and H.-J. Tzeng, "Received signal strength fingerprint and footprint assisted indoor positioning based on ambient Wi-Fi signals," in *Proceedings of the IEEE 75th Vehicular Technology Conference (VTC Spring '12)*, pp. 1–5, IEEE, Yokohama, Japan, June 2012.

[7] M. G. Moon and S. I. Choi, "Indoor position estimation using image sensor based on VLC," in *Proceedings of the International Conference on Advanced Technologies for Communications (ATC '14)*, pp. 11–14, Hanoi, Vietnam, October 2014.

[8] E. Nakamori, D. Tsukuda, M. Fujimoto et al., "A new indoor position estimation method of RFID tags for continuous moving navigation systems," in *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN '12)*, pp. 1–8, Sydney, Australia, November 2012.

[9] S. Khalifa and M. Hassan, "Evaluating mismatch probability of activity-based map matching in indoor positioning," in *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN '12)*, pp. 1–9, IEEE, Sydney, Australia, November 2012.

[10] A. Baniukevic, D. Sabonis, C. S. Jensen, and H. Lu, "Improving Wi-Fi based indoor positioning using bluetooth add-ons," in *Proceedings of the 12th IEEE International Conference on Mobile Data Management (MDM '11)*, pp. 246–255, Lulea, Sweden, June 2011.

[11] M. Ji, J. Kim, J. Jeon, and Y. Cho, "Analysis of positioning accuracy corresponding to the number of BLE beacons in indoor positioning system," in *Proceedings of the 17th International Conference on Advanced Communication Technology (ICACT '15)*, pp. 92–95, IEEE, Seoul, The Republic of Korea, July 2015.

[12] Global Positioning System—Wikipedia, the free encyclopedia, September 2015, https://en.wikipedia.org/wiki/Global_Positioning_System.

[13] iBeacon, 2015, https://en.wikipedia.org/wiki/IBeacon.

[14] Proximity Overview, September 2015, https://docs.gimbal.com/proximity_overview.html.

[15] AltBeacon/spec·GitHub, Deptember 2015, https://github.com/AltBeacon/spec.

[16] "Near field communication—Wikipedia, the free encyclopedia," September 2015, https://en.wikipedia.org/wiki/Near_field_communication.

[17] D. Eppstein, "Finding the k shortest paths," *SIAM Journal on Computing*, vol. 28, no. 2, pp. 652–673, 1999.

[18] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, "Section 24.3: Dijkstra's algorithm," in *Introduction to Algorithms*, pp. 595–601, MIT Press, McGraw-Hill, 2nd edition, 2001.

Hindawi Publishing Corporation
Mobile Information Systems
Volume 2016, Article ID 4313475, 10 pages
http://dx.doi.org/10.1155/2016/4313475

Research Article

An Analytical Study on Eavesdropping Attacks in Wireless Nets of Things

Xuran Li,[1] Hao Wang,[2] Hong-Ning Dai,[1] Yuanyuan Wang,[1] and Qinglin Zhao[1]

[1]*Faculty of Information Technology, Macau University of Science and Technology, Avenida Wai Long, Room A208, Taipa, Macau*
[2]*Big Data Lab, Faculty of Engineering and Natural Sciences, Norwegian University of Science & Technology, Postboks 1517, 6025 Ålesund, Norway*

Correspondence should be addressed to Hong-Ning Dai; hndai@ieee.org

Received 28 July 2015; Accepted 7 December 2015

Academic Editor: Jong-Hyouk Lee

The security of Internet of Things (IoT) has received extensive attention recently. This paper presents a novel analytical model to investigate the eavesdropping attacks in Wireless Net of Things (WNoT). Our model considers various channel conditions, including the *path loss*, the *shadow fading effect*, and *Rayleigh fading effect*. Besides, we also consider the eavesdroppers in WNoT equipped with either omnidirectional antennas or directional antennas. Extensive simulation results show that our model is accurate and effective to model the eavesdropping attacks in WNoT. Besides, our results also indicate that the probability of eavesdropping attacks heavily depends on the shadow fading effect, the path loss effect, Rayleigh fading effect, and the antenna models. In particular, we find that the shadow fading effect is beneficial to the eavesdropping attacks while both the path loss effect and Rayleigh fading effect are detrimental. Besides, using directional antennas at eavesdroppers can also increase the eavesdropping probability. Our results offer some useful implications on designing antieavesdropping schemes in WNoT.

1. Introduction

As one of the most promising information and communication technologies (ICT), IoT has received extensive attentions from both academia and industry recently. The basic idea of IoT is to integrate "smart" objects, the *things* into the Internet with provision of various services to users [1, 2]. The typical killer applications of IoT include the logistic management with RFID technology [3], environmental monitoring with wireless sensor networks [4], smart homes [5], e-health [6], smart grids [7], Maritime Industry [8], and so forth. There are a number of diverse smart objects ranging from small Radiofrequency Identification (RFID) tags to sensors, actuators, mobile phones, smart appliances, smart meters, and so forth. Due to the device heterogeneity, various wireless communication technologies (such as ISO/IEC 18000 [3], IEEE 802.15.4 [9], and Bluetooth [10]) are also exploited to interconnect the smart devices to form a Wireless Net of Things (WNoT). Note that the conventional wired communication technologies (Ethernets, fiber-optic communication, etc.) are also mandatory to connect the WNoT with the rest of the Internet.

Security is one of the fundamental issues in IoT since it is the prerequisite for most IoT applications [11–14]. There raise a number of security threats in IoT, especially in WNoT, where the conventional security countermeasures used in wired networks may not work well in WNoT due to the following inherent constraints of WNoT: (i) the wireless medium is open for any nodes [15]; (ii) it is extremely difficult to deploy centralized control mechanisms in such distributed WNoT [2, 16, 17]. Eavesdropping attack, as one of typical security threats in wireless communication systems, has attracted considerable attention recently [18–24] since many adversary attacks often follow the eavesdropping activity, for example, the man-in-the-middle attack [25] and the hear-and-fire attack [19].

Figure 1 shows a typical example of eavesdropping attacks in a warehouse environment, where each product is attached with an RFID tag, which can passively communicate with

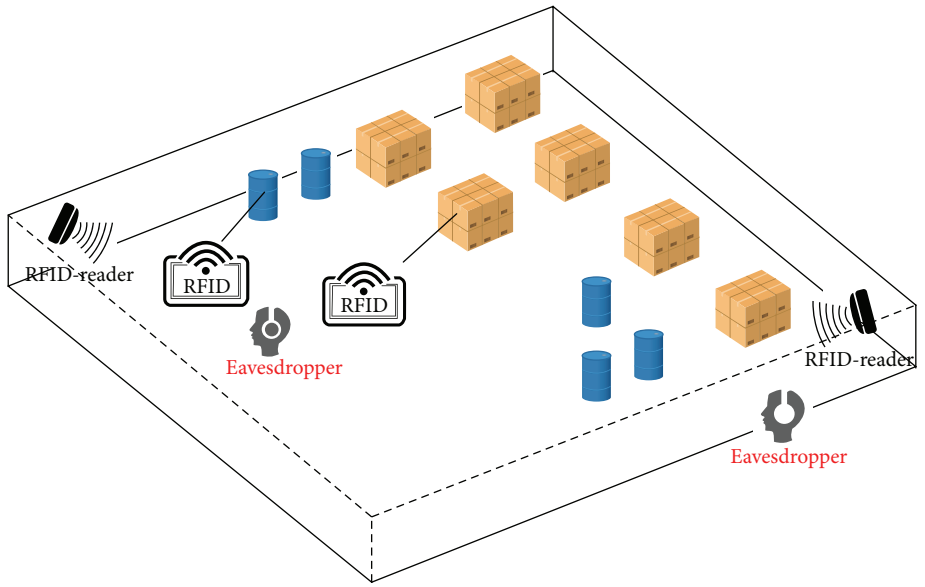


Figure 1: An example of eavesdropping activities in WNoT, where there are several eavesdroppers who are wiretapping the confidential ongoing communications between RFID tags and RFID-readers.

RFID-readers. In this environment, the confidential communications between RFID-readers and RFID tags can be easily wiretapped by eavesdroppers since it is difficult to apply antieavesdropping countermeasures (e.g., encryptions) in this scenario due to the limited computational capability and the energy-constraint of RFID tags. Note that we consider the far-field wireless communications in this scenario [26].

1.1. Related Works. Most of current studies have been concentrated on protecting the confidential communications of smart objects in WNoT, which are also named as *good* nodes in this paper. Encryption is one of the most commonly used techniques to protect the confidential communications in wireless personal area networks [11], wireless local area networks (e.g., WEP [27], WPA, and WPA2 [28]), wireless cellular networks (e.g., Cellular Message Encryption Algorithm [29]), and encryption algorithms for wireless sensor networks [30]. However, it is infeasible to apply cryptography-based techniques in WNoT due to the following reasons: (a) the inferior computational capability of smart objects [2], (b) the limited battery power of smart objects (e.g., the passive RFIDs can only harvest the energy from the readers) [1, 31], and (c) the difficulty of managing the widely distributed smart objects in centralized manner, which is the necessity for the encryption algorithms [11, 32, 33].

An alternate approach is either to design light-weighted encryption schemes [34] or to generate noise to limit the amount of information that can be extracted by an *eavesdropper* [35, 36]. However, one of the most important premises of the above schemes is that we shall have enough knowledge of the channel condition of eavesdroppers as indicated in [37–42], which nevertheless has received little attention. Besides, the wireless channel in WNoT fluctuates from time to time and is affected by various fading effects including the path loss, the shadowing effect, and the multipath effect [43]. Furthermore, most of current studies in WNoT only consider the nodes equipped with omnidirectional antennas, which radiate/receive RF signals in all directions (i.e., a less efficient way to propagate RF signals). As shown in some of the most recent studies [44, 45], directional antennas can be used at readers. Compared with omnidirectional antennas, directional antennas can concentrate the transmissions to some desired directions so that the performance can be further improved.

However, little attention has been paid to *investigating the eavesdropping behaviors conducted by the eavesdroppers in WNoT*, which is nevertheless important for us to offer better protection on the confidential communications since we can design antieavesdropping schemes with clearer targets if we have a better knowledge on the eavesdroppers, although we conducted a preliminary study on the eavesdropping probability of wireless ad hoc networks in [46]. But this paper is significantly different from our previous work [46] in the following aspects: (1) we are concerned with the eavesdropping activities in WNoT in this paper while the previous paper investigated the eavesdropping attacks in wireless ad hoc networks; (2) we propose a novel analytical model on the eavesdropping probability in this paper, where the channel randomness (including Rayleigh fading effect and the shadowing effects) is considered while the previous paper only considered a simplified geometric model; (3) we conduct extensive simulations to verify the accuracy of our proposed model in this paper while the previous paper only presented the numerical results.

1.2. Contributions. The aforementioned issues motivate us to conduct an investigation on the eavesdropping attacks in WNoT. In this paper, we analyze the eavesdropping activities

Table 1: Summary of effects on eavesdropping attacks.

Factors	Effects on eavesdropping attacks
Directional antenna	Positive
Shadow fading	Positive
Path loss	Negative
Rayleigh fading	Negative

conducted by eavesdroppers with consideration of various channel conditions and different types of antennas. *To the best of our knowledge, this is the first study on analyzing the eavesdropping attacks in WNoT from the viewpoints of eavesdroppers*. Our major research contributions in this paper can be summarized as follows:

(i) We formally establish an analytical framework to investigate the probability of eavesdropping attacks in WNoT with consideration of channel randomness. In particular, we consider the path loss effect, the shadow fading effect, and Rayleigh fading effect in our model. Besides, we also take both omnidirectional antennas and directional antennas into account of our analytical framework.

(ii) Extensive simulations show that the simulation results match the analytical results, indicating that our analytical model is accurate and effective. Our results also show that both the path loss effect and Rayleigh fading effect are *detrimental* to the probability of eavesdropping attacks while the shadow fading effect is beneficial to the eavesdropping attacks in WNoT. Besides, our results also indicate that using directional antennas at eavesdroppers can significantly improve the probability of eavesdropping attacks. We summarize our major findings in Table 1.

(iii) Our results can provide many useful implications on designing antieavesdropping schemes in WNoT. This is because we can provide the better protection on the confidential communications if we have the better knowledge about the eavesdroppers as implied in the previous studies [37–42]. For example, we can design light-weight encryption algorithms by exploiting the known channel features [47, 48]. Besides, we only need to encrypt the communications in the area or the direction that is vulnerable to eavesdropping attacks so that the security cost due to the computational complexity can be greatly saved.

The rest of this paper is organized as follows. Section 2 presents the models used in this paper. We then give the analysis on the eavesdropping attacks in Section 3. The impacts of channel randomness with consideration of the shadow fading effect and Rayleigh fading effect are discussed in Section 4. Finally, we conclude the paper in Section 5.

2. Models

In this section, we present the models used in this paper. (See Notations and Symbols section.)

2.1. Node Distribution. In this paper, we assume that all the smart objects (or nodes) are randomly distributed in a 2D area $\mathscr{A}$ according to a homogeneous Poisson point process with density ρ. We denote the number of nodes in an area $\mathscr{A}$ by a random variable N. Then, the probability mass function of N is given as follows:

$$f_N(n) = \frac{(\rho\mathscr{A})^n}{n!} e^{-\rho\mathscr{A}}, \tag{1}$$

where $\rho\mathscr{A}$ is the expected number of nodes in area $\mathscr{A}$.

2.2. Channel Model. We assume that all nodes use the common transmission power $\mathscr{P}_t$ similar to [49]. The channel gain from a node i to an eavesdropper j at a distance r is denoted by $\gamma_{ij}(r)$. Thus, the received power at the eavesdropper is $\mathscr{P}_t \cdot \gamma_{ij}(r)$. The signal-to-interference-plus-noise ratio (SINR) at the eavesdropper denoted by Λ is defined to be

$$\Lambda = \frac{\mathscr{P}_t \cdot \gamma_{ij}(r)}{\eta + \sum_{k\neq i}^{N} \mathscr{P}_t \cdot \gamma_{kj}(r)}, \tag{2}$$

where η is the power of the white noise and N denoted the number of good nodes.

The transmission from node i can be successfully eavesdropped by an eavesdropper if and only if

$$\Lambda \geq \beta, \tag{3}$$

where β is the minimum signal to interference and noise ratio.

In our analysis of eavesdropping activities, we ignore the impact of interference due to the following reasons. First, the passive eavesdroppers in WNoT do not transmit actively and therefore contribute nothing to the interference. Second, the interference is proved to converge when efficient MAC schemes are exploited and the traffic is low in a large-scale network [50, 51]. Thus, our analytical results in this paper can be regarded as the upper bound of the eavesdropping probability. We then have

$$\Lambda = \frac{\mathscr{P}_t \cdot \gamma_{ij}(r)}{\eta} \geq \beta. \tag{4}$$

2.3. Antennas. There are different types of antennas used in wireless communication systems: *omnidirectional* antennas (named *Omni* in short) and *directional antennas* (named *Dir* in short). Most of conventional smart objects are typically equipped with omnidirectional antennas, which radiate/collect radio signals into/from all directions equally. Different from an omnidirectional antenna, a directional antenna can concentrate transmitting or receiving capability on some desired directions consequently leading to the improved network performance. To model the transmitting or receiving capability of an antenna, we denote the *antenna gain* by G. It is obvious that an omnidirectional antenna has a constant antenna gain; that is, $G_o = 1$ in all directions.

We next give the antenna gain of a directional antenna. Since it is difficult to model a realistic directional antenna with precise values of antenna gain in each direction [52], we

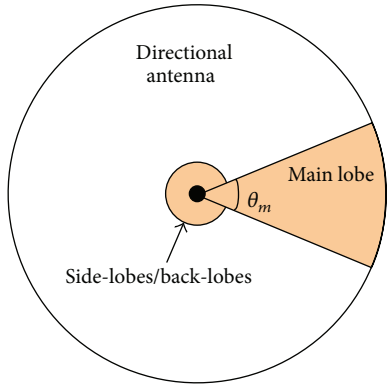


Figure 2: Directional antenna model.

use an approximate antenna model, which was first proposed in [53]. This model is also named as *Keyhole* due to the geometrical analogy to the archaic keyhole in 2D plane, as shown in Figure 2. In this model, the sector with angle θ_m represents the main *lobe* of the antenna, which has the maximum gain denoted by G_m (where θ_m is also called the antenna *beamwidth*), and the circular part represents the side-lobes and back-lobes with lower antenna gain denoted by G_s. In particular, when G_m and θ_m are given [53, 54], we can calculate G_s as follows:

$$G_s = \frac{2 - G_m\left(1 - \cos\left(\theta_m/2\right)\right)}{1 + \cos\left(\theta_m/2\right)}. \tag{5}$$

3. Analysis on Eavesdropping Attacks

This section presents our analytical framework to model the eavesdropping activities in WNoT. In particular, we first analyze *effective eavesdropping area* in Section 3.1 which is then used to derive the *probability of eavesdropping attacks* in Section 3.2. Section 3.3 presents the empirical results.

3.1. Deterministic Path Loss Model. We first consider that the channel gain is mainly determined by the large-scale path loss effect [43]. Thus, the channel gain is given by

$$\gamma_{ij}(r) = C \cdot G_g \cdot G_e \cdot \frac{1}{r^{\alpha}}, \tag{6}$$

where C is a constant, r is the distance between the good node and the eavesdropper, G_g and G_e are the antenna gains for the good node and the eavesdropper, respectively, and α is the path loss exponent ranging from 2 to 4 [43].

As shown in Section 2.2, an eavesdropper can successfully wiretap a transmission if and only if its $\Lambda \geq \beta$. In other words, the probability of no transmission eavesdropped is given by $P(\Lambda < \beta)$. Substituting (6) into inequality (4) and rearranging $P(\Lambda < \beta)$, we have

$$\begin{aligned} P\left(\Lambda < \beta\right) &= P\left(\frac{\mathscr{P}_t \cdot C \cdot G_g \cdot G_e}{\eta \cdot r^{\alpha}} < \beta\right) \\ &= P\left(r > \left(\frac{\mathscr{P}_t \cdot C \cdot G_g \cdot G_e}{\eta \cdot \beta}\right)^{1/\alpha}\right). \end{aligned} \tag{7}$$

We then define a random variable R as

$$R = \left(\frac{\mathscr{P}_t \cdot C \cdot G_g \cdot G_e}{\eta \cdot \beta}\right)^{1/\alpha}, \tag{8}$$

which is referred to the *eavesdropping range* of an eavesdropper. After substituting (8) into inequality (7), we have $P(\Lambda < \beta) = P(r > R)$, which implies that a transmission cannot be eavesdropped by an eavesdropper if and only if the transmitter falls outside the eavesdropping range R of the eavesdropper.

We then analyze the *effective eavesdropping area* of an eavesdropper, which is defined as $E[\pi R^2] = \pi E[R^2]$, where $E[R^2]$ is the second moment of the eavesdropping range R. The effective eavesdropping area is a *critical* region that only when the good node falls in this region, its transmission can be eavesdropped by eavesdroppers. We then have

$$E\left[\pi R^2\right] = \pi E\left[\left(\frac{C \cdot \mathscr{P}_t \cdot G_g \cdot G_e}{\eta \cdot \beta}\right)^{2/\alpha}\right]. \tag{9}$$

3.2. Probability of Eavesdropping Attacks. We model the successful chance of eavesdropping attacks by the *probability of eavesdropping attacks*, denoted by $P(E)$. To derive $P(E)$, we need to analyze the probability of no good node being eavesdropped first. We denote the number of good nodes falling in the eavesdropping area by a random variable Y. Since good nodes are randomly distributed according to a homogeneous Poisson point process (as shown in Section 2.1), we then have the probability of no good node falling in the eavesdropping area, which is given by the following equation:

$$P(Y = 0) = e^{-\rho \cdot E[\pi R^2]}. \tag{10}$$

We then can calculate $P(E)$ as follows:

$$P(E) = 1 - P(Y = 0) = 1 - e^{-\rho \cdot E[\pi R^2]}. \tag{11}$$

After substituting $E[\pi R^2]$ in (11) by Right-Hand Side (RHS) of (9), we have

$$P(E) = 1 - \exp\left(-\rho \cdot \pi E\left[\left(\frac{C \cdot \mathscr{P}_t \cdot G_g \cdot G_e}{\eta \cdot \beta}\right)^{2/\alpha}\right]\right). \tag{12}$$

The physical meaning of $P(E)$ is the probability that an eavesdropper can successfully eavesdrop at least one transmission in WNoT. Besides, as shown in (12), the probability of eavesdropping attacks heavily depends on the path loss effect. Note that this model can be extended to a more general case with consideration of the shadow fading effect and the Rayleigh fading effect, which will be analyzed in Section 4.

3.3. Empirical Results. We conduct extensive simulations to verify the effectiveness and the accuracy of our proposed model. In our simulations, the probability of eavesdropping attacks in a WNoT is calculated by

$$P'(E) = \frac{\Psi}{\Omega}, \tag{13}$$

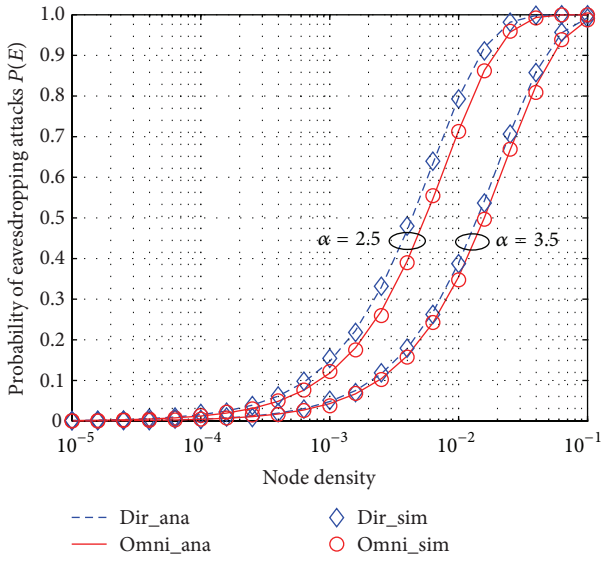


FIGURE 3: Probability of eavesdropping attacks $P(E)$ with path loss effect only when $\alpha = 2.5, 3.5$ and SINR threshold $\beta = 10$ dB.

where Ω and Ψ denote the number of total WNoT topologies and the number of WNoT topologies that have been eavesdropped, respectively. We say that a WNoT topology is eavesdropped when any smart object (node) in this topology is eavesdropped. Note that we denote the simulation results by $P'(E)$ in order to differentiate it from the analytical value $P(E)$. To minimize the impacts of the border effects, we conduct the simulations within an $l \times l$ area with the exclusion of the nodes falling in the outer box $l' \times l'$, where l' shall be significantly larger than l [55]. Note that l is chosen as 3000 m in our simulations. We fix the number of eavesdroppers and choose the node density ρ for the good nodes ranging from 10^{-5} to 10^{-1}. The other system parameters are selected as follows: $C = 10$, $\mathscr{P}_t = 1$ mWatt, $\eta = 0.01$ mWatt, and $\beta = 10$ dB. We consider eavesdroppers equipped with either omnidirectional antenna (Omni) or directional antenna (Dir) while the good nodes are equipped with omnidirectional antennas only.

Figure 3 shows both the analytical results and the simulation results of the probability of eavesdropping attacks with the path loss effect only. The curves and the markers represent the analytical results and simulation results, respectively. It is shown in Figure 3 that the simulation results have a good agreement with the analytical results, implying that our model is quite accurate.

As shown in Figure 3, we also find that the probability of eavesdropping attacks decreases with the increased path loss exponent α, implying that the path loss effect has the negative impact on eavesdropping attacks. Besides, we also find that using directional antennas at eavesdroppers can increase the probability of eavesdropping attacks although this effect is not that significant when the path loss effect is increased (e.g., $\alpha = 3.5$).

4. Impacts of Channel Randomness on Eavesdropping Attacks

In this section, we extend our analytical model in Section 3 to more general cases in consideration of two different effects of channel randomness: (1) shadow fading effect and (2) Rayleigh fading effect, which will be presented in Sections 4.1 and 4.2, respectively. We then give the empirical results in Section 4.3.

In order to model the two random effects, we introduce the *packet eavesdropping probability* denoted by $P_{E|\Lambda}(y)$, which is defined as the probability that a packet is successfully eavesdropped by an eavesdropper when the average signal-to-interference-noise ratio $\overline{\Lambda} = y$.

We then extend the analysis of eavesdropping range in Section 3.1 with consideration of the packet eavesdropping probability $P_{E|\Lambda}(y)$. We first consider the case that the packet eavesdropping probability $P_{E|\Lambda}(y)$ tends to approach a step function if good long code is used [56]. In particular, we have the cumulative distribution function (CDF) of eavesdropping range R, which is defined as follows:

$$F_R(r) = P[\Lambda(r) < \beta] = F_R\left(\frac{\eta \cdot \beta}{\mathscr{P}_t \cdot G_g \cdot G_e}\right). \quad (14)$$

In a more general case when $P_{E|\Lambda}(y)$ is not a step function, the cumulative distribution function is

$$F_R(r) = 1 - \int_0^{+\infty} f_R\left(\frac{\eta \cdot x}{\mathscr{P}_t \cdot G_g \cdot G_e} \mid r\right) P_{E|\Lambda}(x)\, dx, \quad (15)$$

where f_R is the probability density function (PDF) of R.

4.1. Shadow Fading Effect. Following the similar approach [51], we can derive the probability density function of R with consideration of the shadow fading effect as follows:

$$f_R(x) = \frac{1}{\sqrt{2\pi}\sigma x} \cdot \exp\left(-\frac{1}{2}\left(\frac{\ln x - \ln\left(C \cdot r^{-\alpha}\right)}{\sigma}\right)^2\right), \quad (16)$$

where r is the distance between a good node and an eavesdropper and σ is the standard deviation of the Gaussian distribution describing the shadow fading effect.

We then have the second moment of random variable R given as follows:

$$E\left[R^2\right] = \int_0^{+\infty} 2r\left[1 - F_l\left(\frac{\eta \cdot \beta}{\mathscr{P}_t \cdot A_G}\right)\right] dr. \quad (17)$$

After substituting $[1 - F_l(\eta\beta/\mathscr{P}_t A_G)]$ in (17) with RHS of (15) and RHS of (16) (note that $P_{E|\Lambda}(a) = 1$), we finally have

$$E\left[R^2\right] = \int_0^{+\infty} 2r \int_{\eta\beta/\mathscr{P}_t A_G}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma a} \cdot e^{-(1/2)((\ln a - \ln(C \cdot r^{-\alpha}))/\sigma)^2} da\, dr, \quad (18)$$

where $A_G = E[(G_g G_e)^{2/\alpha}]$, which is defined as the *effective antenna gain factor*. It is obvious that the effective antenna gain factor depends on both the antenna gains and the path loss effect.

Let $x = (\ln a - \ln Cr^{-\alpha})/\sigma = \ln(ar^{\alpha}/C)/\sigma$; we then have

$$E\left[R^2\right] = \int_0^{+\infty} 2r \int_{\ln(\eta\beta r^{-\alpha}/\mathscr{P}_t A_G C)/\sigma}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \mathrm{d}x\, \mathrm{d}r. \quad (19)$$

Since the integrals converge absolutely, applying Fubini's theorem [57], we next get

$$E\left[R^2\right] = \left(\frac{\mathscr{P}_t A_G C}{\eta\beta}\right)^{2/\alpha} \exp\left(\left(\frac{\sqrt{2}\sigma}{\alpha}\right)^2\right). \quad (20)$$

Finally, we have the probability of eavesdropping attacks, which is given as the following equation:

$$P(E) = 1 - \exp\left(-\rho\pi\left(\frac{\mathscr{P}_t A_G C}{\eta\beta}\right)^{2/\alpha} \exp\left(\frac{\sqrt{2}\sigma}{\alpha}\right)^2\right). \quad (21)$$

The probability of eavesdropping attacks in (21) is more general than that in (12). This is because (21) becomes (12) when σ becomes 0, implying that there is no shadow fading effect and SINR is completely determined by the path loss effect.

4.2. Rayleigh Fading Effect. Rayleigh fading effect is a stochastic model for wireless propagation when there are a large number of statistically independent reflected and scattered paths from the transmitters to the receivers (or the eavesdroppers).

In the following procedure, we consider the channel condition with superimposed shadow fading and Rayleigh fading effects. We then derive the second moment of random variable R. Since (17) still holds, we have

$$\begin{aligned} E\left[R^2\right] &= \int_0^{+\infty} 2r \left[1 - F_l\left(\frac{\eta\cdot\beta}{\mathscr{P}_t\cdot A_G}\right)\right] \mathrm{d}r \\ &= \int_0^{+\infty} 2r \int_0^{+\infty} f_R\left(\frac{\eta\cdot x}{\mathscr{P}_t\cdot G_g\cdot G_e} \mid r\right) \\ &\quad \cdot P_{E|\Lambda}(x)\, \mathrm{d}x\, \mathrm{d}r, \end{aligned} \quad (22)$$

where $f_R((\eta x/\mathscr{P}_t G_g G_e) \mid r)$, which can be calculated by (16).

We next derive $P_{E|\Lambda}(x)$. Since the instantaneous SINR is exponentially distributed with mean $\Lambda = y$ [51], with the given average SINR value $\overline{\Lambda}$ and the given SINR threshold β, the packet eavesdropping probability $P_{E|\Lambda}(y)$ can be calculated by

$$\begin{aligned} P_{E|\Lambda}(y) &= \int_\beta^{+\infty} f_\Lambda(y)\, \mathrm{d}x = \int_\beta^{+\infty} \frac{1}{y}\cdot e^{-x/y} \mathrm{d}x \\ &= e^{-\beta/y}. \end{aligned} \quad (23)$$

After substituting the corresponding parts in (22) by (16) and (23), we finally have the effective eavesdropping range as follows:

$$\begin{aligned} E\left[R^2\right] &= \int_0^{+\infty} \int_0^{+\infty} e^{-(\eta\cdot\beta)/(x\cdot\mathscr{P}_t\cdot A_G)} \cdot 2r \frac{1}{\sqrt{2\pi}\sigma x} \\ &\quad \cdot e^{-(1/2)((\ln x - \ln(C\cdot r^{-\alpha}))/\sigma)^2} \mathrm{d}r\, \mathrm{d}x \\ &= \int_{-\infty}^{+\infty} \int_0^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot 2r \\ &\quad \cdot e^{-(\eta\cdot\beta\cdot r^{\alpha}\cdot e^{-\sigma x})/(C\cdot\mathscr{P}_t\cdot A_G)} \mathrm{d}r\, \mathrm{d}x, \end{aligned} \quad (24)$$

where $A_G = E[(G_g G_e)^{2/\alpha}]$ is the effective antenna gain factor.

The integral in (24) can be calculated by the following equation [58]:

$$\begin{aligned} &\int_0^{+\infty} 2r \cdot e^{-(\eta\cdot\beta\cdot r^{\alpha}\cdot e^{-\sigma x})/(C\cdot\mathscr{P}_t\cdot A_G)} \mathrm{d}r \\ &\quad = \frac{2}{\alpha}\Gamma\left(\frac{2}{\alpha}\right) \cdot \left(\frac{\eta\cdot\beta\cdot e^{-\sigma x}}{C\cdot\mathscr{P}_t\cdot A_G}\right)^{-2/\alpha}, \end{aligned} \quad (25)$$

where $\Gamma(\cdot)$ represents the general Gamma function.

Substituting (25) into (24) and applying it to (11), we finally have

$$P(E) = 1 - e^{-\rho\pi(2/\alpha)\Gamma(2/\alpha)\cdot((\eta\cdot\beta)/(C\cdot\mathscr{P}_t\cdot A_G))^{-2/\alpha}\cdot e^{(\sqrt{2}\sigma/\alpha)^2}}. \quad (26)$$

4.3. Empirical Results. We have conducted extensive simulations to evaluate the accuracy of our extended model. In order to compare the new results with those under the case without shadowing effects in Section 3.3, we choose the same system parameters as those in Section 3.3. Note that in order to eliminate the impacts of the border effect, the border area of the simulation area shall be slightly increased. Similarly, we also consider eavesdroppers equipped with either omnidirectional antennas or directional antennas.

Figure 4 shows the empirical results of the probability of eavesdropping attacks with shadow fading effects, where the shadow fading deviation $\sigma = 3$. Note that the curves and the markers represent the analytical results and simulation results, respectively. Figure 3 also indicates that the simulation results match the analytical results, implying the accuracy of our model.

As shown in Figure 4, we find that the probability of eavesdropping attacks is affected by both the path loss effect and the shadow fading effect. In particular, $P(E)$ decreases with the increased path loss exponent α, implying that the path loss effect is *detrimental*. In other words, the path loss effect will decrease the probability of eavesdropping attacks, which agrees with the previous results without the shadowing effect (see Figure 3). On the contrary, the shadow fading effect is *beneficial*. More specifically, if we compare Figure 4 with Figure 3, we can find that $P(E)$ increases with the increased values of the shadow fading deviation σ (e.g., σ is increased from 0 to 3). This effect is remarkable when the path loss effect is less notable (e.g., $\alpha = 2.5$). However, $P(E)$ does not increase

Table 2: Comparison between the results under the channel with shadow fading effect only and the results under the channel with superimposed shadowing and Rayleigh fading effects when $\alpha = 3$, $\sigma = 3$, and SINR threshold $\beta = 10$ dB.

Node density	Shadow fading effect only (Figure 4)		Superimposed shadow fading and Rayleigh fading effects (Figure 5)	
ρ	Omni	Dir	Omni	Dir
1×10^{-5}	0.0050	***0.0059***	0.0045 (−10.00%)	***0.0053 (−10.17%)***
1×10^{-4}	0.0489	***0.0572***	0.0443 (−9.41%)	***0.0518 (−9.44%)***
1×10^{-3}	0.3945	***0.4453***	0.3642 (−7.68%)	***0.4126 (−7.34%)***
1×10^{-2}	0.9934	***0.9972***	0.9892 (−4.20%)	***0.9951 (−2.10%)***

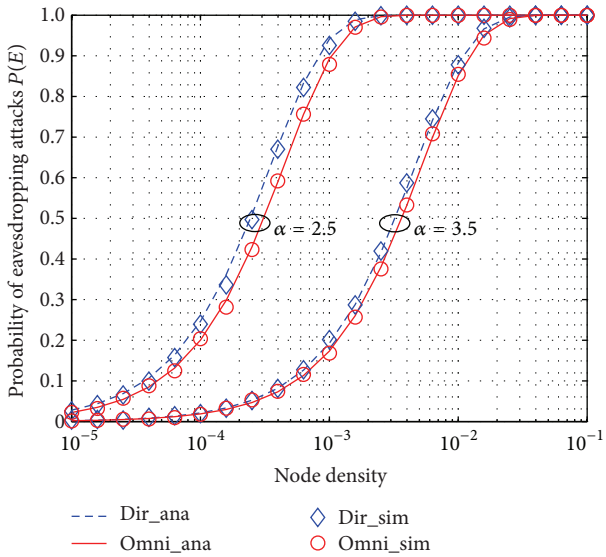


Figure 4: Probability of eavesdropping attacks $P(E)$ with shadowing effect ($\sigma = 3$) only when $\alpha = 2.5, 3.5$ and SINR threshold $\beta = 10$ dB.

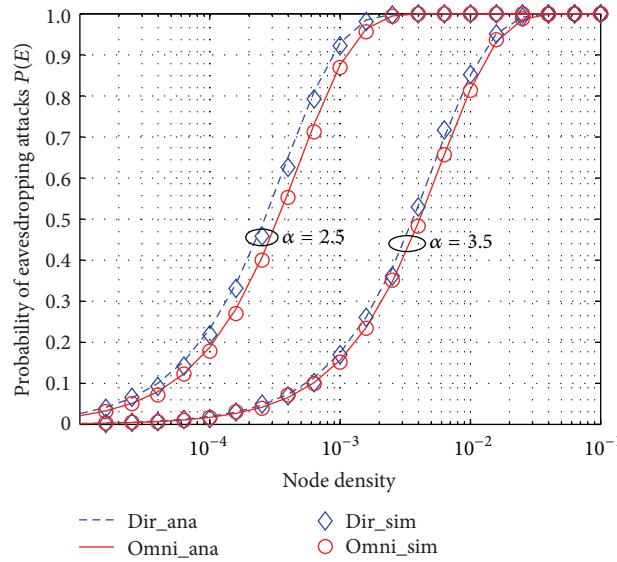


Figure 5: Probability of eavesdropping attacks $P(E)$ with superimposed shadowing effect and Rayleigh fading effect when $\sigma = 3$ and SINR threshold $\beta = 10$ dB.

significantly with the increased values of σ when $\alpha = 3.5$. Furthermore, we also find that using directional antennas at eavesdroppers can increase the probability of eavesdropping attacks with consideration of the shadowing effect.

We then investigate the probability of eavesdropping attacks under the channel with the superimposed shadow fading and Rayleigh fading effects. Figure 5 shows the results with the presence of both shadow fading and Rayleigh fading effects, where the shadow fading deviation $\sigma = 3$. As shown in Figure 5, we find that the probability of eavesdropping attacks is affected by both the shadow fading effect and the Rayleigh fading effect. Moreover, Figure 5 also indicates that Rayleigh fading effect has a *negative* impact on the probability of eavesdropping attacks even though it is not that noticeable compared with the path loss effect.

To illustrate the detrimental effect of Rayleigh fading effect, we conduct comparative study on the numerical results of the probability of eavesdropping attacks $P(E)$. In particular, Table 2 illustrates the comparison between the results of $P(E)$ under the channel with shadow fading effect only and the results under the channel with the superimposed shadow fading effect and Rayleigh fading effect when $\alpha = 3$ and $\sigma = 3$ corresponding to Figures 4 and 5, respectively. To make it clearer, we italicize the results with directional antennas in Table 2. It is shown in Table 2 that Rayleigh fading effect will decrease the probability of eavesdropping attacks compared with the results under the channel with the shadow fading effect only. For example, Rayleigh fading effect leads to the decrement of nearly 10% in terms of the probability of eavesdropping attacks when the node density $\rho = 10^{-5}$. Besides, Table 2 also indicates that using directional antennas at eavesdroppers can increase the probability of eavesdropping attacks, which is similar to the previous findings.

We also give the results under the scenario of eavesdropping attacks with Rayleigh fading effect only. Figure 6 shows the empirical results of the probability of eavesdropping attacks under the channel with Rayleigh fading effect only, where $\sigma = 0$ indicating no shadow fading effect. Similar to the previous results, we also denote the analytical results by the curves and the simulation results by the markers, as shown in Figure 6. It is shown in Figure 6 that the simulation results have a good agreement with the analytical results, implying that our analytical model is quite accurate.

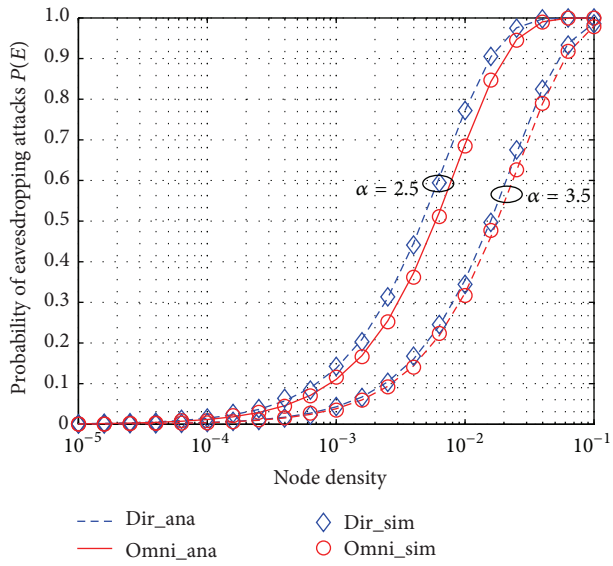


FIGURE 6: Probability of eavesdropping attacks $P(E)$ with Rayleigh fading effect only (without shadowing effect) where SINR threshold $\beta = 10$ dB and $\sigma = 0$.

As shown in Figure 6, we can see that the probability of eavesdropping attacks also depends on both the path loss effect and Rayleigh fading effect. In particular, $P(E)$ drops significantly when the path loss effect becomes more notable (e.g., $\alpha = 3.5$), as shown in Figure 6. Besides, under the wireless channel with Rayleigh fading effect, $P(E)$ in Figure 6 is even lower than that without Rayleigh fading effect in Figure 3, implying that Rayleigh fading effect is also detrimental to the eavesdropping attacks. The reason may owe to the counteracting effect of the multipath scattering signals under the channel with Rayleigh fading effect [43].

4.4. Discussions and Implications of Our Results. Our simulation results imply that using directional antennas at eavesdroppers in WNoT can significantly increase the probability of eavesdropping. Thus, directional antennas are *beneficial* to eavesdroppers. The improvement mainly owes to the effect that a directional antenna can accumulate the receiving capability of desired directions. However, we can not ignore another effect that a directional antenna can also narrow the angle of the receiving directions. More specifically, with the increased path loss (i.e., the larger α), the second effect can even counteract the first effect. Take Figure 6 as an example. The gap between the results of omnidirectional eavesdroppers and the results of directional eavesdroppers with $\alpha = 2.5$ is significantly bigger than that with $\alpha = 3.5$.

Secondly, as shown in our results, both the path loss effect and Rayleigh fading are always detrimental to the eavesdropping probability while shadowing effect and directional antennas are beneficial to the eavesdropping probability. Our findings are useful to help to design more effective antieavesdropping schemes in WNoT. This is because we need the knowledge of eavesdroppers (such as the channel characteristics) so that we can design the light-weight encryption algorithms as indicated in the previous studies [37–42]. Besides, we only need to take antieavesdropping measures in the area or the direction that is vulnerable to eavesdropping attacks so that the security cost due to the computational complexity can be greatly saved. For example, we can generate the noise only in the direction of eavesdroppers when the eavesdroppers are equipped with directional antennas while there is no noise in other directions. This new scheme may have a better performance than the existing one [35].

5. Conclusion

In this paper, we propose an analytical model to investigate the eavesdropping probability in Wireless Net of Things (WNoT) with consideration of channel randomness including the path loss effect, the shadow fading effect, and Rayleigh fading effect. After conducting extensive simulations, we show that our model is quite accurate. Besides, we have also shown that the eavesdropping probability heavily depends on the path loss effect, the shadow fading effect, and Rayleigh fading effect. More specifically, we find that the eavesdropping probability increases when the shadow fading factor σ increases and decreases when the path loss effect increases, implying that the path loss effect is detrimental to the eavesdropping attacks while the shadow fading is beneficial to the eavesdropping attacks. Moreover, similar to the path loss effect, Rayleigh fading is also destructive to the eavesdropping attacks. Furthermore, our results also indicate that using directional antennas at eavesdroppers can significantly improve the probability of eavesdropping attacks.

Notation and Symbols

$\mathcal{A}$: 2D area that nodes are randomly distributed
ρ: Density of the homogeneous Poisson point process
$\mathcal{P}_t$: Transmission power of nodes
r: Distance between the good node and the eavesdropper
$\gamma_{ij}(r)$: Channel gain from a good node i to an eavesdropper j at a distance r
Λ: SINR at an eavesdropper
β: Threshold value of SINR for eavesdropping a node successfully
η: Power of the white noise
N: Number of good nodes
α: Path loss exponent
G_m, G_s: Antenna gain of main lobe, antenna gain of side-lobe
θ_m: Main lobe beam-width of the keyhole antenna
G_g, G_e: Antenna gain of good node, antenna gain of eavesdropper
$P(E)$: Probability of eavesdropping attacks
l: Side length of topology area
R: Eavesdropping range of an eavesdropper

Ω: Number of total WNoT topologies
Ψ: Number of WNoT topologies that have been eavesdropped
$\overline{\Lambda}$: Average SINR value
$P_{E|\Lambda}(y)$: Packet eavesdropping probability when the average SINR is y
σ: Standard deviation of the Gaussian distribution describing the shadow fading effect
A_G: Effective antenna gain factor.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The work described in this paper was partially supported by Macao Science and Technology Development Fund under Grant no. 096/2013/A3 and Grant no. 104/2014/A3 and supported by Innovation Norway through the project "GCE BLUE Maritime Big Data." The authors would like to thank Gordon K.-T. Hon for his helpful comments that greatly improve the quality of this paper.

References

[1] L. Atzori, A. Iera, and G. Morabito, "The internet of things: a survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[2] D. Miorandi, S. Sicari, F. De Pellegrini, and I. Chlamtac, "Internet of things: vision, applications and research challenges," *Ad Hoc Networks*, vol. 10, no. 7, pp. 1497–1516, 2012.

[3] ISO/IEC 18000, 2013, http://en.wikipedia.org/wiki/ISO/IEC_18000.

[4] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Computer Networks*, vol. 52, no. 12, pp. 2292–2330, 2008.

[5] C. Dixon, R. Mahajan, S. Agarwal et al., "An operating system for the home," in *Proceedings of the 9th USENIX Conference on Networked Systems Design and Implementation (NSDI '12)*, p. 25, USENIX Association, San Jose, Calif, USA, April 2012.

[6] K. Habib, A. Torjusen, and W. Leister, "Security analysis of a patient monitoring system for the Internet of Things in eHealth," in *Proceedings of the International Conference on eHealth, Telemedicine, and Social Medicine (eTELEMED '15)*, Lisbon, Portugal, February 2015.

[7] Z. Fan, P. Kulkarni, S. Gormus et al., "Smart grid communications: overview of research challenges, solutions, and standardization activities," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 1, pp. 21–38, 2013.

[8] H. Wang, O. Osen, G. Li, W. Li, H.-N. Dai, and W. Zeng, "Big data and industrial internet of things for the maritime industry in Northwestern Norway," in *Proceedings of the IEEE Region 10 Conference (TENCON '15)*, Macau, China, November 2015.

[9] IEEE 802.15.4, 2011, http://standards.ieee.org/getieee802/download/802.15.4-2011.pdf.

[10] Bluetooth Core Specification 4.2, 2014, http://www.bluetooth.org/.

[11] J. Granjal, E. Monteiro, and J. Sa Silva, "Security for the internet of things: a survey of existing protocols and open research issues," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 3, pp. 1294–1312, 2015.

[12] A. Grau, *How to Build a Safer Internet of Things*, IEEE Spectrum, 2015.

[13] S. Sicari, A. Rizzardi, L. A. Grieco, and A. Coen-Porisini, "Security, privacy and trust in Internet of Things: the road ahead," *Computer Networks*, vol. 76, pp. 146–164, 2015.

[14] G. Strazdins and H. Wang, "Open security and privacy challenges for the internet of things," in *Proceedings of the 10th International Conference on Information, Communications and Signal Processing (ICICS '15)*, 2015.

[15] C. Cai, Y. Cai, X. Zhou, W. Yang, and W. Yang, "When does relay transmission give a more secure connection in wireless ad hoc networks?" *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 4, pp. 624–632, 2014.

[16] N. A. Alrajeh, S. Khan, and B. Shams, "Intrusion detection systems in wireless sensor networks: a review," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 167575, 7 pages, 2013.

[17] N. Meghanathan, "A survey on the communication protocols and security in cognitive radio networks," *International Journal of Communication Networks and Information Security*, vol. 5, no. 1, pp. 19–38, 2013.

[18] M. Anand, Z. G. Ives, and I. Lee, "Quantifying eavesdropping vulnerability in sensor networks," in *Proceedings of the 2nd International Workshop on Data Management for Sensor Networks (DMSN '05)*, pp. 3–9, August 2005.

[19] J.-C. Kao and R. Marculescu, "Eavesdropping minimization via transmission power control in ad-hoc wireless networks," in *Proceedings of the 3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks (SECON '06)*, vol. 2, pp. 707–714, IEEE, Reston, Va, USA, September 2006.

[20] H.-N. Dai, D. Li, and R. C.-W. Wong, "Exploring security improvement of wireless networks with directional antennas," in *Proceedings of the IEEE 36th Conference on Local Computer Networks (LCN '11)*, pp. 191–194, Bonn, Germany, October 2011.

[21] X. Lu, F. Wicker, P. Lio, and D. Towsley, "Security estimation model with directional antennas," in *Proceedings of the IEEE Military Communications Conference (MILCOM '08)*, pp. 1–6, IEEE, San Diego, Calif, USA, November 2008.

[22] Q. Wang, H.-N. Dai, and Q. Zhao, "Eavesdropping security in wireless Ad Hoc networks with directional antennas," in *Proceedings of the 22nd Wireless and Optical Communications Conference (WOCC '13)*, pp. 687–692, May 2013.

[23] H.-N. Dai, Q. Wang, D. Li, and R. C.-W. Wong, "On eavesdropping attacks in wireless sensor networks with directional antennas," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 760834, 13 pages, 2013.

[24] E. Alsaadi and A. Tubaishat, "Internet of things: features, challenges, and vulnerabilities," *International Journal of Advanced Computer Science and Information Technology*, vol. 4, no. 1, pp. 1–13, 2015.

[25] F. Anjum and P. Mouchtaris, *Security for Wireless Ad Hoc Networks*, Wiley-Interscience, 1st edition, 2007.

[26] R. Want, "An introduction to RFID technology," *IEEE Pervasive Computing*, vol. 5, no. 1, pp. 25–33, 2006.

[27] IEEE 802.11a-1999, http://standards.ieee.org/getieee802/download/802.11a-1999.pdf.

[28] IEEE 802.11i-2004, http://standards.ieee.org/getieee802/download/802.11i-2004.pdf.

[29] D. Wagner, B. Schneier, and J. Kelsey, "Cryptanalysis of the cellular message encryption algorithm," in *Advances in Cryptology—CRYPTO '97*, vol. 1294 of *Lecture Notes in Computer Science*, pp. 526–537, Springer, Berlin, Germany, 1997.

[30] M. Turkanović, B. Brumen, and M. Hölbl, "A novel user authentication and key agreement scheme for heterogeneous ad hoc wireless sensor networks, based on the Internet of Things notion," *Ad Hoc Networks*, vol. 20, pp. 96–112, 2014.

[31] A.-R. Sadeghi, C. Wachsmann, and M. Waidner, "Security and privacy challenges in industrial internet of things," in *Proceedings of the 52nd Annual Design Automation Conference (DAC '15)*, San Francisco, Calif, USA, June 2015.

[32] S. L. Keoh, S. S. Kumar, and H. Tschofenig, "Securing the internet of things: a standardization perspective," *IEEE Internet of Things Journal*, vol. 1, no. 3, pp. 265–275, 2014.

[33] Z. Yan, P. Zhang, and A. V. Vasilakos, "A survey on trust management for internet of things," *Journal of Network and Computer Applications*, vol. 42, pp. 120–134, 2014.

[34] B. Azimi-Sadjadi, A. Kiayias, A. Mercado, and B. Yener, "Robust key generation from signal envelopes in wireless networks," in *Proceedings of the 14th ACM Conference on Computer and Communications Security (CCS '07)*, pp. 401–410, Denver, Colo, USA, November 2007.

[35] O. Savry, F. Pebay-Peyroula, F. Dehmas, G. Robert, and J. Reverdy, "RFID noisy reader how to prevent from eavesdropping on the communication?" in *Cryptographic Hardware and Embedded Systems—CHES 2007*, vol. 4727 of *Lecture Notes in Computer Science*, pp. 334–345, Springer, Berlin, Germany, 2007.

[36] A. Mukherjee, S. A. A. Fakoorian, J. Huang, and A. L. Swindlehurst, "Principles of physical layer security in multiuser wireless networks: a survey," *IEEE Communications Surveys and Tutorials*, vol. 16, no. 3, pp. 1550–1573, 2014.

[37] G. P. Hancke, "Practical eavesdropping and skimming attacks on high-frequency RFID tokens," *Journal of Computer Security*, vol. 19, no. 2, pp. 259–288, 2011.

[38] F. Oggier and B. Hassibi, "The secrecy capacity of the MIMO wiretap channel," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 4961–4972, 2011.

[39] R. Liu, T. Liu, H. V. Poor, and S. Shamai, "Multiple-input multiple-output gaussian broadcast channels with confidential messages," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4215–4227, 2010.

[40] X. He, A. Khisti, and A. Yener, "MIMO multiple access channel with an arbitrarily varying eavesdropper: secrecy degrees of freedom," *IEEE Transactions on Information Theory*, vol. 59, no. 8, pp. 4733–4745, 2013.

[41] I. Hero, "Secure space-time communication," *IEEE Transactions on Information Theory*, vol. 49, no. 12, pp. 3235–3249, 2003.

[42] Y. Zou, B. Champagne, W.-P. Zhu, and L. Hanzo, "Relay-selection improves the security-reliability trade-off in cognitive radio systems," *IEEE Transactions on Communications*, vol. 63, no. 1, pp. 215–228, 2015.

[43] T. S. Rappaport, *Wireless Communications: Principles and Practice*, Prentice Hall, Upper Saddle River, NJ, USA, 2nd edition, 2002.

[44] A. Sawadi, *An RFID directional antenna for location positioning [Ph.D. dissertation]*, University of Windsor, 2012.

[45] D. M. Dobkin, *The RF in RFID: Passive UHF RFID in Practice*, Newnes, 2nd edition, 2012.

[46] X. Li, H.-N. Dai, and Q. Zhao, "An analytical model on eavesdropping attacks in wireless networks," in *Proceedings of the IEEE International Conference on Communication Systems (ICCS '14)*, pp. 538–542, IEEE, Macau, China, November 2014.

[47] S. Mathur, W. Trappe, N. Mandayam, C. Ye, and A. Reznik, "Radio-telepathy: extracting a secret key from an unauthenticated wireless channel," in *Proceedings of the ACM 14th Annual International Conference on Mobile Computing and Networking (MobiCom '08)*, pp. 128–139, ACM, San Francisco, Calif, USA, September 2008.

[48] F. Huo and G. Gong, "A new efficient physical layer OFDM encryption scheme," in *Proceedings of the 33rd IEEE Conference on Computer Communications (INFOCOM '14)*, pp. 1024–1032, Toronto, Canada, May 2014.

[49] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, 2000.

[50] M. Franceschetti, O. Dousse, D. N. Tse, and P. Thiran, "Closing the gap in the capacity of wireless networks via percolation theory," *IEEE Transactions on Information Theory*, vol. 53, no. 3, pp. 1009–1018, 2007.

[51] D. Miorandi, E. Altman, and G. Alfano, "The impact of channel randomness on coverage and connectivity of ad hoc and sensor networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 3, pp. 1062–1072, 2008.

[52] C. A. Balanis, *Antenna Theory: Analysis and Design*, John Wiley & Sons, New York, NY, USA, 2nd edition, 1997.

[53] R. Ramanathan, "On the performance of ad hoc networks with beamforming antennas," in *Proceedings of the 2nd ACM International Symposium on Mobile Ad Hoc Networking & Computing (MobiHoc '01)*, pp. 95–105, ACM, Long Beach, Calif, USA, October 2001.

[54] Q. Wang, H.-N. Dai, and Q. Zhao, "Connectivity of wireless Ad Hoc networks: impacts of antenna models," in *Proceedings of the 14th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT '13)*, pp. 298–303, Taipei, Taiwan, December 2013.

[55] C. Bettstetter, "On the connectivity of ad hoc networks," *The Computer Journal*, vol. 47, no. 4, pp. 432–447, 2004.

[56] M. Zorzi and S. Pupolin, "Outage probability in multiple access packet radio networks in the presence of fading," *IEEE Transactions on Vehicular Technology*, vol. 43, no. 3, pp. 604–610, 2002.

[57] J. Borwein, D. Bailey, and R. Girgensohn, *Experimentation in Mathematics: Computational Paths to Discovery*, Wellesley, 2004.

[58] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, New York, NY, USA, 7th edition, 2007.

Hindawi Publishing Corporation
Mobile Information Systems
Volume 2016, Article ID 9817374, 10 pages
http://dx.doi.org/10.1155/2016/9817374

Research Article

A Geo-Aware Taxi Carrying Management System by Using Location Based Services and Zone Queuing Techniques on Internet of Things

Chyi-Ren Dow, Duc-Binh Nguyen, Shr-Chen Wang, Shiow-Fen Hwang, and Ming Fong Tsai

Department of Information Engineering and Computer Science, Feng Chia University, No. 100, Wenhua Road, Seatwen, Taichung 40724, Taiwan

Correspondence should be addressed to Chyi-Ren Dow; crdow@fcu.edu.tw

Received 29 September 2015; Accepted 16 November 2015

Academic Editor: Jong-Hyouk Lee

Taxi plays a crucial role in the transportation system because of the characteristic that can be hailed conveniently. Most of the taxi drivers obtain passengers by hunting on the road or waiting in a fixed taxi queuing point; however these methods have poor performance, high vacancy rate, and several critical problems such as air pollution and foul up traffic. This study proposed a taxi carrying management system by using location based services and zone queuing techniques on Internet of things. The proposed system allows drivers to both hunt on the road and wait in a queuing zone. A queuing table is used in the control center and neighbor tables are used in RSUs for zone queuing establishment. Joining and leaving mechanisms are developed for zone queuing management. To enhance service efficiency and quality, we present a scheme to prevent the ping-pong effect which is based on the location based services, a hunting rate calculation scheme, and a path planning service for taxi drivers according to the history carrying record. PRISM is used to simulate the proposed system, and the results indicated that our scheme outperforms the waiting and hunting models in terms of number of customers, vacancy rate, and profit.

1. Introduction

Internet of vehicles (IoV) is an emerging research problem in recent years. It is a complex integrated network system which converges the mobile Internet and the Internet of Things (IoT) by comprising of various vehicles. It is a converged technology that encompasses information communication, environmental protection, energy conservation, and safety. Numerous applications use vehicle-to-vehicle [1], vehicle-to-sensor [2], and vehicle-to-infrastructure [3] communications in IoV, such as traffic event alarms, advertisement broadcasting, and entertainment services. Because of the characteristics of IoV networks, vehicles can conveniently transmit their traveling information and receive information from other vehicles with supported hardware and software.

Taxis play a crucial role for the convenience of traffic transportation, which reflects the levels of economic development and civilization of a city. Internet of Taxis is an important part of a smart city. The characteristics of IoV help taxi drivers to easily hail customers and reduce the vacancy rate. Presently, there are three common taxi carrying models. The first model is hunting, and in this model, a taxi driver finds passengers on the street through luck and experience. This consumes considerable time and gasoline. The second model is fixed queue in which taxis wait and queue for passengers at specific places. Most of these places are around city hot spots such as a train station with high demand for taxi services. In this model, taxi drivers can pick up a passenger without consuming considerable gasoline. However, the waiting time increases when there are a high number of taxis in the queue. The third model is dispatching, and this model comprises drivers who join a fleet, such as "Taiwan Taxi" or "M-Taxi." A passenger can call the fleet for a taxi by phone or through the Internet, and the fleet dispatches a taxi to a specific place. This increases the opportunity for drivers to obtain passengers. However, the drivers must pay for each call of dispatching, even if the passenger does not show at the specified place.

Each of these traditional taxi carrying models has its own pros and cons. The major disadvantages of the waiting model are long waiting time and blocked traffic when the number of taxis in the queue is higher than the capacity of the waiting area. It is vital to demarcate a fixed point of the waiting model to a larger area for reducing the waiting time and blocked traffic. However, ping-pong effect [4] is a potentially undesirable phenomenon, in which the system needs to perform handovers frequently between the same pair of demarcated areas back and forth, when drivers roam around the boundary of the areas in a short time period. Ping-pong effect can cause inefficiency, dropping from the queue of queuing zone, and degrading of the system performance. Coverage parameters, vehicle's location area and its movement, are the main considerations that can cause the ping-pong effect. Furthermore, the hunting model is more favorable than the waiting model in terms of average number of customers and vacancy rate because drivers actively hunt for customers on the street. The waiting model is more favorable than the hunting model in terms of gas consumption that directly affects a taxi driver's profit. In order to obtain advantages of both the hunting and waiting models, a hybrid model should be investigated and developed to allow drivers to both hunt on the road and wait in a queue.

To resolve the aforementioned problems, this study proposed and developed a geo-aware taxi carrying management system by using location based service (LBS) and zone queuing techniques on Internet of vehicles. This is a distributed system, and in this system, the location based services were used in zone queuing establishment and service areas are demarcated based on a geogrid [5]. Furthermore, in the proposed system, each service area can function independently as a queuing zone. This study also implemented a system prototype with wireless access in vehicular environments/dedicated short-range communications (WAVE/DSRC) to provide communications and user interfaces for both the control center and vehicles to verify the feasibility of the proposed system. In the experimental results section, we simulated and proved that the efficiency of our proposed scheme is better than the waiting and hunting methods via PRISM [6] which is a module checker supporting many probabilistic models. PRISM uses a simple and state-based language to express probabilistic models and automatically analyze data patterns. It also includes a discrete-event simulation engine, providing support for approximate/statistical model checking, and implementations of various analysis techniques, such as quantitative abstraction refinement and symmetry reduction.

The rest of this paper is organized as follows. Related work is reviewed in Section 2. In Section 3, the zone queuing scheme is summarized. Section 4 presents the prototype of the implemented user interface and communication setting. Section 5 presents the experimental results, and Section 6 presents the conclusion and future work.

2. Related Work

This section presents some related work, including geo-aware service's research for taxi and queuing techniques. The geo-aware service is well-developed presently and uses the GPS to observe and collect information. Several studies used sensors and GPS to monitor environments in specific places for preventing and promptly notifying natural disaster [7], to propose more convenient traveling methods in transportation [8]. Because of the fast growing of intelligent transportation systems (ITS), numerous applications were proposed such as traffic control [9], road safety [10], and routing path planning [11, 12]. A traffic information system is proposed [13] which used RSUs to collect vehicle information in each road segment and provided traffic flow state to drivers for safe and shortest distance to destination. Tornell et al. [14] proposed an application for smart phones based on the enhanced Message Dissemination based on Roadmaps protocol (eMDR). The application integrates a navigation system and wireless communication device and uses ITS information to avoid critical-mission collision and achieve safety driving. IoV also is considered as ITS combined with Vehicular Ad Hoc Network (VANet) techniques to provide communications between vehicles and infrastructure by using road side units (RSUs) and on-board units (OBUs) via Internet environment [15].

There are three common research problems regarding taxi. The first problem involves real-taxi environment modeling. Most of researches used an amount of real-taxi GPS traced records to analyze statistics and emulate the taxi module, such as behavior of taxi drivers [16] and passenger generators [17]. The next problem involves designing a protocol to achieve the goal of curtailing waiting time or saving gas. Chen et al. [18] proposed a dynamic taxi-sharing protocol which utilized traffic information and ITS technology to avoid traffic congestion. This work, though getting fuel-saving and pollution reducing when the number of passengers willing to share taxis increases dramatically, is still not well enough to reach high business efficiency. Sheu et al. [19] proposed a distributed taxi hailing protocol which aims to hail a taxi with shortest distance under the rules of road signature and reduce vacancy rate as much as possible. However, this method planned and recommended path without the highest possibility to hail customers which may help to increase income for a taxi driver.

The final problem involves implementing application systems for taxi environments. Hosni et al. [20] developed a shared taxi service system which benefits both the taxi drivers and the passengers. In these studies, a passenger can make a reservation on the Internet or call the service center to book a taxi. The communication system between taxi and passenger has been implemented by Liu et al. [21] in which passengers can join an occupied taxi on road if they have the same destination. The system also provides trip history model to analyze behavior of passengers in a specific place, and this model can reduce vacancy rate of taxis.

Regarding scheduling schemes, Yuan et al. [22] proposed a method to guarantee strict fairness and utilize prediction better in parallel job scheduling. McKeown [23] showed scheduling algorithm called iSLIP. An iterative, round-robin algorithm, iSLIP, can achieve 100% throughput for uniform traffic, simple to implement in hardware and extensively used in various applications. Gabale et al. [24] classify scheduling

algorithms with problem setting, problem goal, type of inputs, and solution techniques in wireless mesh networks. Fan and Quan [25] proposed harmonic-aware multicore scheduling based on the rate monotonic scheduling (RMS) policy to guarantee the schedulability of real-time tasks. A dynamic scheduling algorithm proposed by Ren and van der Schaar [26] can be used in wireless cloud computing.

3. Zone Queuing Scheme

This section presents the proposed zone queuing scheme, including zone queuing establishment, zone queuing management, and path planning protocol.

3.1. Zone Queuing Establishment. The first part of the proposed zone queuing scheme involves zone queuing establishment, which can be divided into area demarcation, center selection, and weight calculation formula. The area demarcation depends on the number of passengers in the whole area; this is because most of the existing queuing zones are located in high service demand areas, such as Taipei city. The center is the brain of the queuing zone and controls all the actions of the queuing zone and provides planning path service for drivers. The weight calculation formula is used to calculate the weight for each grid, and paths are planned based on this weight.

3.1.1. Area Demarcation and Center Selection. For the area demarcating, a queuing zone is established in specific places, such as department stores, hospitals, stations, night markets, or other hot spots that have a high service demand for taxis. Thus, the first step in establishing the queue involves defining the size of a queuing zone based on the service requirement. The boundary was determined using GPS technology and this is important to deal with the ping-pong effect (see Section 3.2 for more details). In order to obtain queuing information from the control center, this study assumed that each taxi is equipped with wireless communication devices. The taxi also requires a GPS sensor and a digital map to determine its location.

To ensure the communication between a taxi and the center, the second step, this study used geogrids that cut the queuing area into several grids; the grid size was based on the transmission range R of the devices. In each grid, a RSU was developed as the grid center. Normally, each RSU can communicate with the service center and other RSUs through wired networks. This model is used to collect information from taxis and serves as a hailing stand for passengers who can use wireless communications, such as Wi-Fi, 3G, or 4G, to hail a taxi.

The third step involves determining the control center of the queuing zone. The control center plays a vital role in this scheme because it manages queuing processes, such as joining the queue, leaving the queue, and calculating the average wait time. To ensure efficient data collection, the node located at the center of the queuing zone was used as the control center because the distance from this node to each of the other nodes is slightly similar.

3.1.2. Queuing Table in the Control Center. The control center manages all processes in a queuing zone. Thus, it must maintain a queuing table to manage such processes. Equation (1) shows the queuing table. Queuing number is a serial number, and lower value means higher priority. Vehicular ID is a unique plate number. RSU number records the RSU that enables a taxi to transmit messages. Timestamp is used to determine which messages must be updated or deleted. Because only the newest message is recorded in the queuing table, the control center can communicate with the taxi via the RSU number.

Queuing Table in the Control Center

Queuing Number	Vehicular ID	RSU Number	Timestamp	TTL	(1)

Neighbor Table in RSU

Vehicular ID	State	Timestamp	TTL	(2)

Basically, a taxi essentially sends "hello" messages periodically to the RSU in its grid for updating its location if it is in the queue. When the RSU receives a hello message, it compares message content with information in the neighboring table. If the taxi information is not in the table, the RSU transmits an updated message to the control center and updates its neighbor table. Otherwise, it updates only its neighboring table. In addition to periodically updating the state of queuing, the RSU sends an advertisement message when it receives hailing requests from passengers. Each taxi can send a response when it receives the message, even if it is not in the queue. Equation (2) illustrates the neighbor table format for the RSU; in this table, state denotes whether a taxi is in the queue. Ping-pong time to live (TTL) denotes the time used for the taxi to prevent the ping-pong effect.

3.2. Zone Queuing Management. The second part of the proposed scheme is zone queuing management, which involves joining and leaving mechanisms and provides a scheme for preventing the ping-pong effect. In the joining and leaving mechanisms, the workflow for a taxi to join or leave the queue is presented in detail. The ping-pong effect is a prevalent problem in wireless communications and is discussed in Section 3.2.2.

3.2.1. Joining and Leaving Mechanisms. When a taxi enters a queuing zone, it transmits a joining message to the control center via the RSU in its grid, and the control center returns

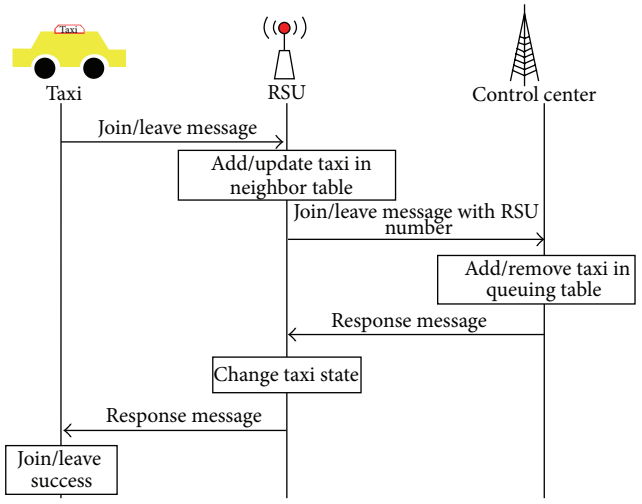


Figure 1: Workflow of joining or leaving the queue.

the queuing number and planning path to the taxi after receiving a joining message. The taxi joins a queue successfully when it receives the response message, and it starts transmitting update message periodically. Figure 1 shows the workflow for joining or leaving a queue.

When a taxi driver intends to obtain passengers by hunting or to leave the queue for personal reasons, the driver can transmit a leaving message by using the console queuing function. When the control center receives a leaving message, it deletes the information of the taxi from the queuing table and responds with a message confirming the completion of the leaving process. If a taxi leaves without transmitting a leaving message, the control center deletes it from the queuing table when the timestamp is not updated in a specific period.

Because the taxi periodically transmits updating messages to the RSU after successfully joining the queue, the control center can accurately locate the taxi according to the RSU number. When the control center receives a leaving message from a taxi, it examines the RSU number to ensure whether the message is legal. If someone attempts to transmit a fake leaving message, the control center ignores the message if its RSU number does not match.

3.2.2. Prevention of the Ping-Pong Effect. The ping-pong effect is a prevalent problem in cellular network systems. It occurs when users are located on the boundary of two base stations, and it requires frequent handover to keep service in operation, thus leading to a long delay time and unfavorable performance. In the proposed system, this problem is encountered when a taxi is roaming around the boundary of a queuing zone. If the boundary problem is not handled, the taxi is deleted from the queue if the control center cannot locate it, and the taxi must rejoin the queue and wait for a longer period. To ensure connection between a taxi and the control center, the taxi must determine whether it is prepared to leave the queue; the first step involves defining a boundary area.

Therefore, the boundary area comprises two domains: warning domain and switching domain.

The formulas of these two domains are defined as follows.

Warning domain: taxi in the grid of boundary RSU and its location (x, y)

$$\left[(x, y) \mid \begin{Bmatrix} x_{\min} + \alpha < x \cup x \le x_{\max} - \alpha \\ y_{\min} + \alpha < y \cup y \le y_{\max} - \alpha \end{Bmatrix}\right]. \tag{3}$$

Switching domain:

$$\left[(x, y) \mid \begin{Bmatrix} x_{\min} < x \le x_{\min} + \alpha \cup x_{\max} - \alpha \le x < x_{b\max} \\ y_{\min} < y \le y_{\min} + \alpha \cup y_{\max} - \alpha \le y < y_{b\max} \end{Bmatrix}\right]. \tag{4}$$

$x_{\min}$ is the minimum and $x_{\max}$ is the maximum longitude, $y_{\min}$ is the minimum and $y_{\max}$ is the maximum latitude of queuing zone, and α is a threshold for flexible boundary area. In this study, we suggest α in a range from 1/3 to 2/3 of d, whereas d is an offset between $x_{\min}$ and longitude value of nearest boundary RSU.

If an in-queue taxi is in the warning domain of the boundary area, the driver is warned about leaving the queuing zone via a graphical user interface. The warning is continually shown until the taxi is out of the boundary area. If the taxi continues moving to the switching domain from the warning domain, a hello message is sent to RSU with nonnull TTL value, and the TTL information is updated into the queuing table of the control center. If the control center does not receive any update from the taxi after a certain period of time or the difference between the new timestamp and the old timestamp is greater than the TTL value, the control center deletes the taxi from the queuing table. Figure 2 illustrates the boundary area diagram.

3.3. Path Planning Protocol. The final part of the zone queuing scheme is path planning. To achieve path planning, the information of the taxi carrying history must be recorded and the recorded information must be classified regularly. The path can then be planned according to the records at any time and any place. In this section, the data structure of the carrying records is presented first. Next, the methods involved in collecting and integrating the carrying records are introduced. Finally, the path planning mechanism is discussed.

3.3.1. Collection and Management of Carrying Records. For collecting carrying records, each RSU maintains a carrying record table as shown in Table 1. When a taxi obtains a customer, it transmits a carrying message to the RSU, which is in its grid to notify the information of the location and time about the passenger. Because the number of the carrying records is extremely large, the records must be managed effectively to derive useful information. Thus, the carrying records were divided into seven sets according to the days of the week, and each set was divided into 24 slots.

Table 1: Carrying records table in RSU.

Date	Week	Time interval	Vehicular ID	Location
2014/1/12	Sat.	(17-18)	AX-15	(24.186861, 120.646382)

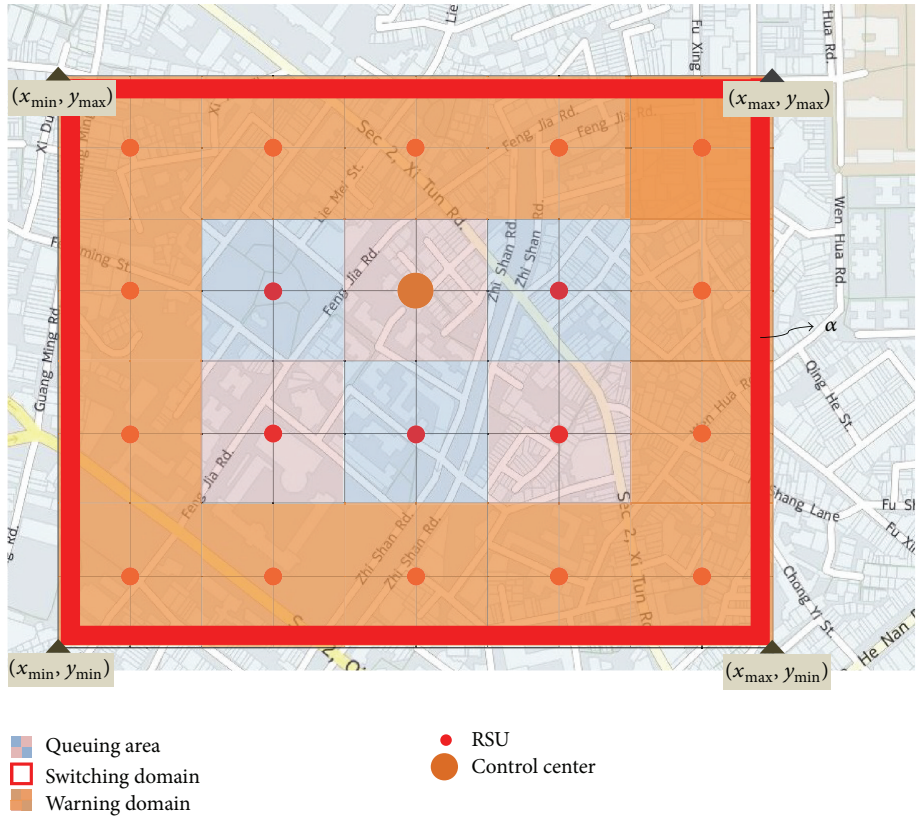


Figure 2: Queuing and boundary area diagram.

When the carrying records are classified, the hunting rate of passengers in each time slot of each day of the week must be calculated. Each RSU then transmits the set of the calculated values to the control center periodically, and the control center plans the paths for taxi drivers according to the transmitted set. The following assumptions were defined to explain the calculation efficiently:

(1) $W = \{\text{Mon, Tue, Wed, Thu, Fri, Sat, Sun}\}$ indicates the carrying day of the week;

(2) $T = \{(0\text{-}1), (1\text{-}2), \ldots, (23\text{-}24)\}$ indicates the time interval for the carrying record;

(3) $R_j^i = \{r_1, r_2, \ldots, r_m\}$ indicates a set of carrying records in day of week $i \in W$ and time slot $j \in T$ received from t axis S_j^i indicates the number of hirings in day of week $i \in W$ and time slot $j \in T$;

(4) H_j^i indicates the hunting rate for day of week $i \in W$ and time slot $j \in T$;

(5) N_j^i indicates historical hunting rate in day of week $i \in W$ and time slot $j \in T$.

The following algorithm will be performed at the end of day of week i and time slot j for calculating the hunting rate, which is used in the path planning phase and operates as follows.

Hunting Rate Calculation Algorithm

(1) $S_j^i = 0$

(2) Read (H_j^i)

(3) Read (N_j^i)

(4) **Do** $S_j^i = \text{Count}\,(R_j^i)$

(5) $H_j^i = (H_j^i * N_j^i + S_j^i)/(N_j^i + 1)$

(6) $N_j^i = N_j^i + 1$

(7) Write (N_j^i)

(8) Write (H_j^i)

(9) return H_j^i

(1) (Line (1)) First, variable S_j^i is initialized. It is used to store the number of carrying records of each time slot i and

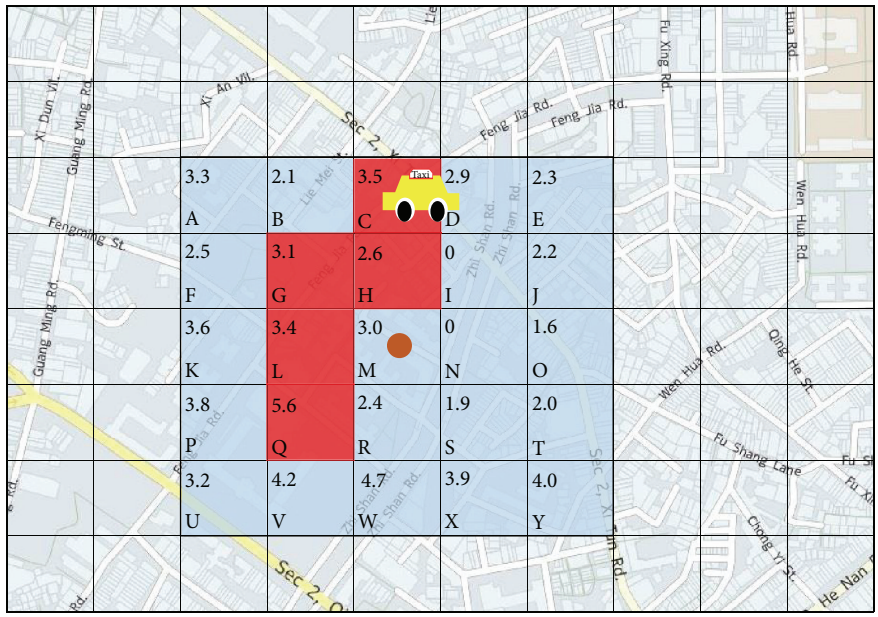


Figure 3: Example of path planning.

day of the week j. At the beginning, the value of each S_j^i is assigned by zero.

(2) (Lines (2)-(3)) Second, reading the historical data of hunting rate in previous session is done.

(3) (Line (4)) Third, we count the total number of carrying records in the database. At the end of the process, the total number of taxis hailed by passengers for time slot j and day of the week i is derived.

(4) (Lines (5)–(9)) Finally, the algorithm is used to calculate the hunting rate H_j^i of the grid in time slot j and day of the week i. H_j^i is the average number of customers for time slot j and day of the week i of current session and historical session. At the end of this process, the value of hunting rate and historical hunting rate will be stored into the database for next session hunting rate calculation.

3.3.2. Path Planning Mechanism. When a taxi driver requests a path planning service, the control center determines all H_j^i from each RSU according to the request date and time. Then, the control center plans a path according to the priority of H_j^i. For example, as shown in Figure 3, the queuing zone area comprises Grids A to Y, and the control center is located in Grid M.

When the control center receives a path planning request from a taxi in Grid C, it broadcasts collection record messages to collect all H_j^i in this area. Next, the highest value of H_j^i is used to choose the destination of the path with the highest opportunity of having passengers. In this case, the highest grid of this area is Q, indicating that the path is from C to Q and the direction is down and left. Therefore, the path is always selected as down or left grid. For example, the grids neighboring to Grid C are B, D, and H, despite Grid D having a higher value than Grids B and H. The system does not select Grid D as the path because this grid is on the right side.

4. System Implementation and Prototype

To provide services for taxi drivers, this study designed and implemented a simple prototype system. We describe the framework of this prototype in this section.

4.1. System Overview. Figure 4 is our system overview in which taxi is equipped with an on-board smart device and communication device (IWCU). The RSU is equipped with an embedded platform and communication device that provides two types of communication interface. One is Wi-Fi communication for a passenger who is near the RSU, and the other is 802.11p wireless communication for taxi drivers. A passenger can use a smartphone to hail a taxi via Wi-Fi communication with the RSU.

4.2. System Implementation. This prototype can be divided into three parts. In the first part, an application is designed for an on-board device for drivers, and simple functions of zone queuing are provided. In the second part, a RSU is implemented on embedded devices. This RSU performs two types of functions, center and normal RSUs. In the final part, the communication device is configured. The device must be set up and some configuration must be performed on it.

For the on-board device, iOS devices are used to implement this service because such devices are popular and extensively used presently. Furthermore, the smart device provides base units of GPS and wireless network communications that can be easily activated in the proposed application; it also has a good user interface that can enable the proposed application

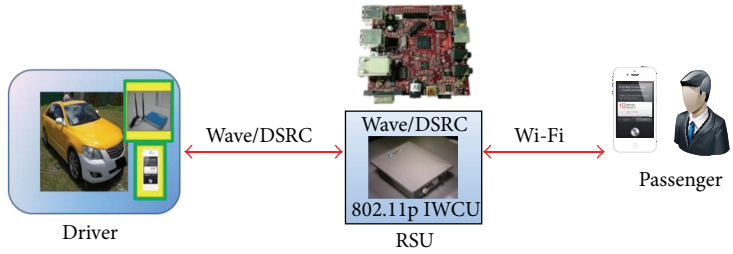


Figure 4: System overview.

Figure 5: Dispatching message from queuing zone on the user interface.

to be more user-friendly. To design and implement the proposed system, this study used Xcode Version 4.4 as the integrated development environment (IDE) and Objective-C as the programming language. Notably, Xcode and Objective-C are the original IDE and programming language in iOS platform.

To handle communications between taxis and passengers, this study used IWCU, a WAVE/DSRC communication device developed by ITRI of Taiwan to provide end-to-end communications. This device currently can transmit only UDP packet to a UDP port number that is in the same subnetwork. Because it is a prototype, it is like a forwarder, not a router; therefore, two ports must be implemented in different devices as a pair before initiating the packet transmission. The prototype provides a Web interface for users to set the relation of each device.

4.3. System Prototype. The monitoring system can show all on-queue taxis on a digital map, and it can also show the queuing sequence on the left-hand side. Although the control center is operated automatically, an administrator can manage and dispatch taxis using this monitoring system if required.

Figure 5 illustrates the user interface designed for taxi drivers. When a taxi goes into a queuing zone, it can join the queue if it is in service. Drivers can use navigation buttons to request the control center to plan paths. Moreover, drivers can see the boundary line on the digital map, and a warning message is provided if they are too close to the switching domain. This can prevent taxis from being deleted because they cross over the domain. Drivers can also adjust some personal settings such as the priority of modes or start the navigation in the setting window. When the taxi receives dispatching messages from the control center (Figure 5), the location of the passenger, distance between the taxi and the passenger, and estimated arrival time are provided in a conforming window. The driver can decide whether to pick up the passenger. The information about the destination is not provided to prevent drivers from refusing to accept short-distance rides. If the driver accepts the mission, a navigation service is also provided to the drivers if it is required.

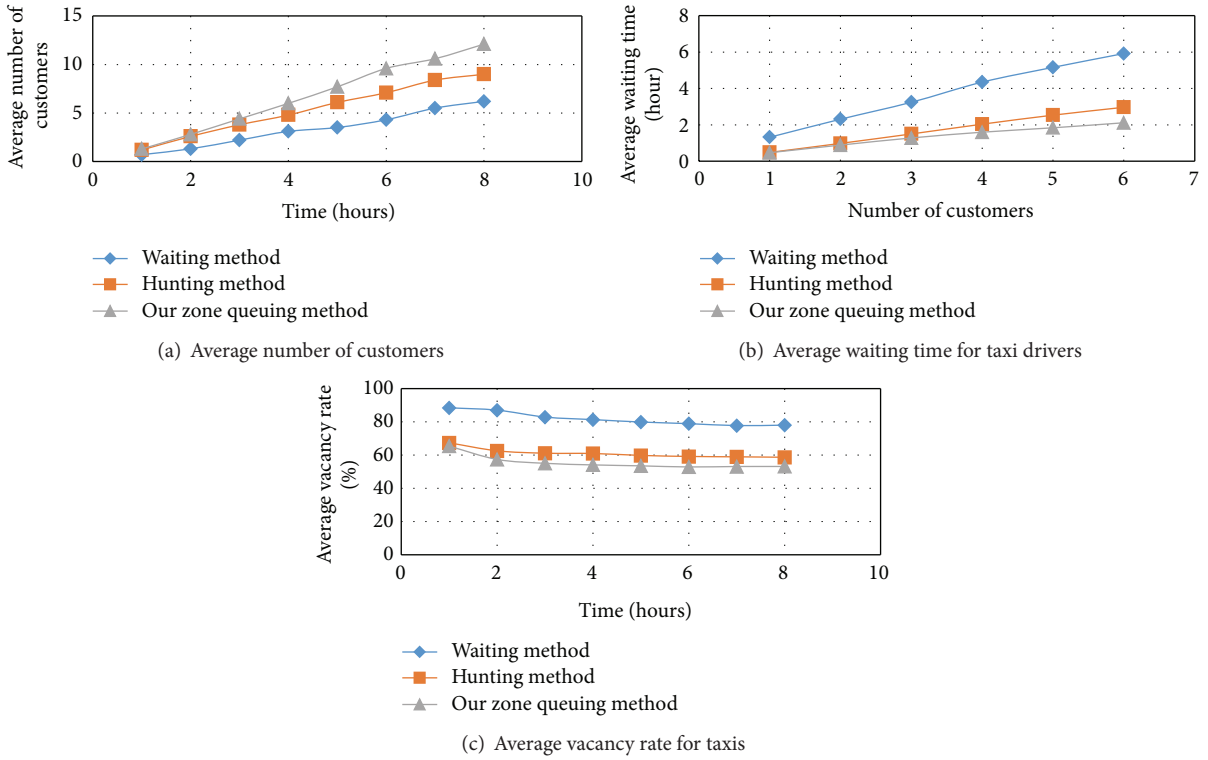
Table 2: Statistical data table for the fixed taxi queuing point at Wuri station.

Item	Value
Average waiting time for taxi	60 (minutes)
Average number of customers per day	8 (sets)
Average working time for a driver	10 (hours)
Average number of taxis in a queuing point	76 (units)
Average income for a driver	1520 (NT dollars)

5. Experimental Results

This section presents the experimental results. The results include field investigation about real-taxi behavior and simulation results with investigation statistics. On this part, we use PRISM to simulate and prove that the efficiency of this model is better than waiting model in which taxi drivers wait in a fixed taxi queuing point and hunting model in which taxi drivers obtain passengers by hunting on the road.

5.1. Field Investigation. To obtain rational experimental data, this study designed a questionnaire for taxi drivers and investigated the Wuri station of the Taiwan High-Speed Rail, Taichung. Approximately 30 effective questionnaires were retrieved. As shown in Table 2, the average waiting time for a driver at the Wuri station is approximately 60 minutes, average number of customers is eight sets, and average working time is nearly 10 hours. The average income can be calculated by multiplying the average number of customers



(a) Average number of customers

(b) Average waiting time for taxi drivers

(c) Average vacancy rate for taxis

FIGURE 6: Results of number of customers and vacancy rate.

by the income generated in the average distance. It was established according to charging standard in Taichung, and the statistical data of the average distance were provided by the Ministry of Transportation Communications.

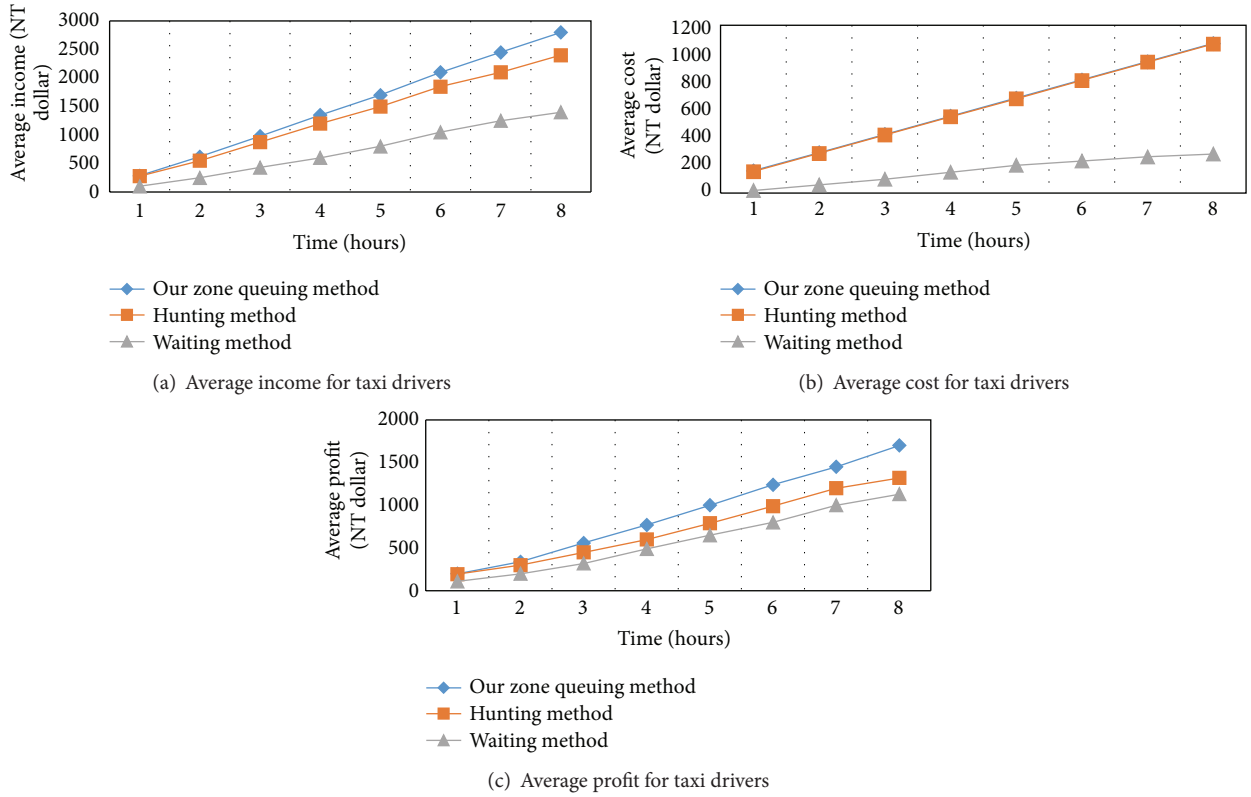
5.2. Simulation Results of Average Number of Customers, Waiting Time, and Vacancy Rate. In this section, we simulate waiting, hunting, and our proposed zone queuing models via PRISM to evaluate models' performance. To make the simulation environment more practical, we set the experimental environment and parameter values of simulation equal to the values obtained by field investigation in Table 2.

Figure 6(a) depicts the result of the average number of customers. The zone queuing model can obtain more customers than the waiting and hunting models do because this model combines the waiting and hunting model. When drivers drive in a queuing zone, they can obtain customers from the street and can also obtain customers from the dispatching center if they wait in a queue. The hunting model is more favorable than the waiting model because taxis consume very little gas when they wait at a fixed queuing point. Therefore, most taxi drivers go to fixed queuing points, even if they must spend a longer time waiting for customers; this is because drivers perceive this model as a stable business approach.

Figure 6(b) shows the results of average waiting time for a taxi driver. Drivers obtained more customers in zone queuing mode than they did in waiting and hunting modes. The average waiting time for each taxi driver in the zone queuing mode is lower than that in the hunting and waiting modes. They may obtain a customer in approximately 20 to 25 minutes. However, the hunting mode demonstrates more favorable performance than the waiting mode because drivers can roam in the streets according to their experience and they may have support by their companies to help them obtain customers more easily. By contrast, in the waiting mode, taxi drivers hold on positions and wait for customers; their only option is choosing a fixed queuing point that has more service requests and fewer taxis waiting. Drivers obtain customers in the hunting mode on average every 30 minutes, whereas they obtain customers in the waiting mode on average every 1 hour.

The vacancy rate is a vital indicator of the performance of each mode. Figure 6(c) shows the average vacancy rate for taxis. The vacancy rate of the zone queuing model is approximately 53%. Although it is greater than half, according to the investigation, the average vacancy rate of taxis in Taiwan is nearly 65%; therefore, the zone queuing model can reduce more than 10% vacancy for taxi drivers. Regarding the hunting mode, the vacancy rate is approximately 60%, and this rate is more consistent with the investigation; however, the waiting model has a vacancy rate of approximately 80%. This thus indicates that a driver wasted considerable time in a fixed queuing point to wait for customers again.

5.3. Simulation Results of Average Profit. The average profit was derived by analyzing the average waiting time and carrying time. The popular brand, Toyota, a taxi business brand, was used as the simulation vehicle. The average income and average cost values for taxi drivers were calculated



(a) Average income for taxi drivers

(b) Average cost for taxi drivers

(c) Average profit for taxi drivers

Figure 7: Results of average profit.

according to the average distance for each customer and average fuel consumption. The average profit by income and cost can thus be derived.

Figure 7(a) shows the average income for taxi drivers. As expected, the waiting model has lower income than the hunting and zone queuing modes because income is directly proportional to the number of customers. Figure 7(a) illustrates the relationship between time and the average income.

Figure 7(b) shows the average cost for taxi drivers. The values of the hunting model and zone queuing model are similar because this study assumed in the simulation that drivers are always roaming for customers either in the hunting model or the zone queuing mode. However, the greatest advantage of the waiting model is that drivers are not required to roam for customers because they are already waiting at a hot spot. Thus, they can reduce gas consumption.

Figure 7(c) shows the evaluated average profit for drivers. The average profit represents the amount of money a driver earns in a day. The zone queuing model demonstrates the highest profit among all modes because drivers in this model can hunt on the road or wait in a queue.

6. Conclusions

This study designed a geo-aware taxi carrying management system by using LBS and zone queuing techniques on Internet of vehicles, including zone queuing establishment, zone queuing management, and path planning protocol. The system can provide taxi drivers a new style to hail customers. In this system, taxi drivers can either hunt or wait for passengers in a queuing zone, which provides an LBS in a specific place that has high demand for taxi service. This study also presents a path planning service for taxi drivers according to the history carrying records, which enhance hunting efficiency and improve service quality. PRISM was used to perform our simulation, and the results show that the proposed system has more favorable performance than existing systems. In the future, zone queuing and cruising mechanisms will be developed according to hot spot information, grid weight values, and crowdsourcing information and a prototype of the zone queuing and cruising system will be implemented to provide advice and guidance to taxi drivers. Also, we will use Big Data resources provided by Taiwan Taxi Company, the biggest taxi company in Taiwan, to analyze LBS data and calculate average waiting time of each queuing zone and then adaptively guide taxis to different hot spots.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

The authors would like to thank the Ministry of Science and Technology of the Republic of China for financially supporting this research under Contract nos. MOST103-2221-E-035-057 and MOST104-2221-E-035-021.

References

[1] R. He, A. F. Molisch, F. Tufvesson, Z. Zhong, B. Ai, and T. Zhang, "Cooperative positioning for vehicular networks: facts and future," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2237–2248, 2014.

[2] P. Papadimitratos, A. La Fortelle, K. Evenssen, R. Brignolo, and S. Cosenza, "Vehicular communication systems: enabling technologies, applications, and future outlook on intelligent transportation," *IEEE Communications Magazine*, vol. 47, no. 11, pp. 84–95, 2009.

[3] J. Li, T. Luo, and L. Ding, "A location aware based handoff algorithm in V2I system of railway environment," in *Proceedings of the Global Information Infrastructure and Networking Symposium (GIIS '14)*, pp. 1–6, IEEE, Québec, Canada, September 2014.

[4] A. Singhrova and N. Prakash, "Vertical handoff decision algorithm for improved quality of service in heterogeneous wireless networks," *IET Communications*, vol. 6, no. 2, pp. 211–223, 2012.

[5] K.-H. Chen, C.-R. Dow, S.-C. Chen, Y.-S. Lee, and S.-F. Hwang, "HarpiaGrid: a geography-aware grid-based routing protocol for vehicular Ad Hoc networks," *Journal of Information Science and Engineering*, vol. 26, no. 3, pp. 817–832, 2010.

[6] PRISM, http://www.prismmodelchecker.org.

[7] Y. J. Lee, J. Trevathan, I. Atkinson, W. Read, T. Myers, and R. Johnstone, "The evolution of the SEMAT sensor network management system," in *Proceedings of the 7th International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, pp. 229–234, IEEE, Adelaide, Australia, December 2011.

[8] N. Alam and A. G. Dempster, "Cooperative positioning for vehicular networks: facts and future," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 4, pp. 1708–1717, 2013.

[9] P. J. Tseng, C. C. Hung, T. H. Chang, and Y. H. Chung, "Real-time urban traffic sensing with GPS equipped probe vehicles," in *Proceedings of the 12th International Conference on ITS Telecommunications (ITST '12)*, pp. 306–310, Taipei, Taiwan, November 2012.

[10] C. T. Barba, M. Á. Mateos, P. R. Soto, A. M. Mezher, and M. A. Igartua, "Smart city for VANETs using warning messages, traffic statistics and intelligent traffic lights," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV '12)*, pp. 902–907, IEEE, Alcalá de Henares, Spain, June 2012.

[11] A. Dua, N. Kumar, and S. Bawa, "A systematic review on routing protocols for vehicular Ad-Hoc networks," *Vehicular Communications*, vol. 1, no. 1, pp. 33–52, 2014.

[12] D. Lee, Y. Kim, and H. Lee, "Route prediction based vehicular mobility management scheme for VANET," *International Journal of Distributed Sensor Networks*, vol. 2014, Article ID 679780, 9 pages, 2014.

[13] X.-Y. Bai, X.-M. Ye, H. Jiang, and J. Li, "A novel traffic information system for VANET based on location service," in *Proceedings of the 16th International Conference on Networks (ICON '08)*, pp. 1–6, New Delhi, India, December 2008.

[14] S. M. Tornell, C. T. Calafate, J. C. Cano, and P. Manzoni, "Implementing and testing a driving safety application for smartphones based on the eMDR protocol," in *Proceedings of the IFIP Wireless Days (WD '12)*, pp. 1–3, Dublin, Ireland, November 2012.

[15] R. Tiwari and N. Kumar, "Minimizing query delay using cooperation in IVANET," *Procedia Computer Science*, vol. 57, pp. 84–90, 2015.

[16] C. Chen, D. Zhang, N. Li, L. Sun, S. Li, and Z. Wang, "iBOAT: isolation-based online anomalous trajectory detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 806–818, 2013.

[17] S.-F. Cheng and T. D. Nguyen, "TaxiSim: a multiagent simulation platform for evaluating taxi fleet operations," in *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT '11)*, pp. 14–21, IEEE, Lyon, France, August 2011.

[18] P.-Y. Chen, J.-W. Liu, and W.-T. Chen, "A fuel-saving and pollution-reducing dynamic taxi-sharing protocol in VANETs," in *Proceedings of the 72nd Vehicular Technology Conference Fall (VTC '10)*, pp. 1–5, IEEE, Ottawa, Canada, September 2010.

[19] J.-P. Sheu, G.-Y. Chang, and C.-H. Chen, "A distributed taxi hailing protocol in vehicular ad-hoc networks," in *Proceedings of the IEEE 71st Vehicular Technology Conference (VTC '10)*, pp. 1–5, IEEE, Taipei, Taiwan, May 2010.

[20] H. E. Hosni, N. Farhat, R. Nimer et al., "An optimization-based approach for passenger to shared taxi allocation," in *Proceedings of the 20th International Conference on Software, Telecommunications and Computer Networks (SoftCOM '12)*, pp. 1–7, IEEE, Split, Croatia, September 2012.

[21] N. Liu, M. Liu, J. Cao, G. Chen, and W. Lou, "When transportation meets communication: V2P over VANETs," in *Proceedings of the 30th IEEE International Conference on Distributed Computing Systems (ICDCS '10)*, pp. 567–576, Genova, Italy, June 2010.

[22] Y. Yuan, Y. Wu, W. Zheng, and K. Li, "Guarantee strict fairness and utilizeprediction better in parallel job scheduling," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 4, pp. 971–981, 2014.

[23] N. McKeown, "The iSLIP scheduling algorithm for input-queued switches," *IEEE/ACM Transactions on Networking*, vol. 7, no. 2, pp. 188–201, 1999.

[24] V. Gabale, B. Raman, P. Dutta, and S. Kalyanraman, "A classification framework for scheduling algorithms in wireless mesh networks," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 1, pp. 199–222, 2013.

[25] M. Fan and G. Quan, "Harmonic-aware multi-core scheduling for fixed-priority real-time systems," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 6, pp. 1476–1488, 2014.

[26] S. Ren and M. van der Schaar, "Dynamic scheduling and pricing in wireless cloud computing," *IEEE Transactions on Mobile Computing*, vol. 13, no. 10, pp. 2283–2292, 2014.

Hindawi Publishing Corporation
Mobile Information Systems
Volume 2015, Article ID 841341, 10 pages
http://dx.doi.org/10.1155/2015/841341

Research Article

2D-DOA and Mutual Coupling Estimation in Vehicle Communication System via Conformal Array

Yan Zou,[1,2] Hong Xie,[1] Liangtian Wan,[3] Guangjie Han,[3] and Wei Li[3]

[1]*College of Information & Communication Engineering, Harbin Engineering University, Harbin 150001, China*
[2]*No. 91404 Army, Qinhuangdao 066000, China*
[3]*Department of Information and Communication Systems, Hohai University, Changzhou 213022, China*

Correspondence should be addressed to Guangjie Han; hanguangjie@gmail.com

Received 29 September 2015; Accepted 1 December 2015

Academic Editor: Jong-Hyouk Lee

Many direction-of-arrival (DOA) estimation algorithms have been proposed recently. However, the effect of mutual coupling among antenna elements has not been taken into consideration. In this paper, a novel DOA and mutual coupling coefficient estimation algorithm is proposed in intelligent transportation systems (ITS) via conformal array. By constructing the spectial mutual coupling matrix (MCM), the effect of mutual coupling can be eliminated via instrumental element method. Then the DOA of incident signals can be estimated based on parallel factor (PARAFAC) theory. The PARAFAC model is constructed in cumulant domain using covariance matrices. The mutual coupling coefficients are estimated based on the former DOA estimation and the matrix transformation between MCM and the steering vector. Finally, due to the drawback of the parameter pairing method in Wan et al., 2014, a novel method is given to improve the performance of parameter pairing. The computer simulation verifies the effectiveness of the proposed algorithm.

1. Introduction

Intelligent transportation systems (ITS) have emerged as an effective way of improving the performance of transportation systems and enhancing travel security. The term connected vehicles refers to applications, services, and technologies that connect a vehicle to its surroundings. With the rapid development of 5G, the peak data rate will likely be in the range of tens of Gbps, which is suitable for real-time communication between vehicles and BS (base station). In order to minimize the cost of the vehicles and reduce the friction resistance between the surface of the vehicles and atmosphere, the conformal array antennas installed in the vehicles should be a good choice.

Direction-of-arrival (DOA) estimation plays an important role in array signal processing, which has been widely used in radar, sonar, and smart antenna [1–3]. Multiple signal classification- (MUSIC-) based algorithm [4] and estimation of signal parameters via rotational invariance techniques- (ESPRIT-) based algorithm [5] are two types of DOA estimation algorithms which can achieve superresolution. However, the effect of mutual coupling will reduce the performance of DOA estimation severely.

In order to solve this problem, many algorithms have been proposed to deal with it. An effective compensation method for the effect of mutual coupling in uniform circular arrays (UCAs) employed for two-dimensional (2D) DOA estimations was proposed [6]. A DOA estimation algorithm for mixed signals with unknown mutual coupling was proposed in [7]. The noncoherent (uncorrelated or partially correlated) signals were firstly estimated with unknown mutual coupling; then the mutual coupling coefficients can be obtained by these estimation. Finally, the noncoherent signals were eliminated and the effect of mutual coupling was compensated. The bias in uniform linear array (ULA) and general linear array caused by effect of mutual coupling was studied in [8]. The DOA estimation for noncoherent and coherent signals has been proposed in [9], which is used for patient's localization.

Parallel factor (PARAFAC) analysis attracted the attention of researchers when it was originally introduced in array signal processing in 2000 [10, 11]. It has been widely used for low-rank decomposition of three-way and higher way array.

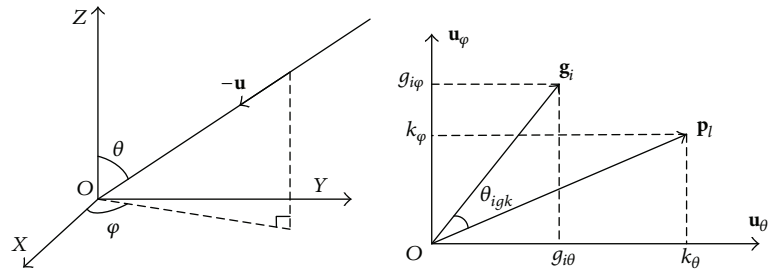


Figure 1: (a) The incident signal **u**. (b) The ith element's response.

Based on PARAFAC and cumulant, a new array structure was designed for 2D-DOA estimation [12]. The DOA and polarization estimation for a single electromagnetic vector-sensor were acquired by using PARAFAC [13], and the proposed algorithm performs better than the ESPRIT algorithm.

Conformal array is the array integrated on surface of object [14, 15]. Some DOA estimation algorithms have been proposed for conformal array. The propagator method and three rotational invariance relationships have been used for fast DOA estimation for cylindrical conformal array [16]. A ultrawideband DOA estimation algorithm has been proposed for DOA estimation based on spatial baseline method and multiple subarray technique [17]. The manifold separation technique has been used for DOA estimation of wireless sensor network with arbitrary array configuration [18, 19]. A real-time DOA estimation algorithm has been proposed in [20]. The DOA estimation is accomplished via ESPRIT and the elimination of the effect of mutual coupling [21]. Based on spatial baseline technique, the DOA can be estimated with low computational complexity. However, only one source can be estimated [22]. In [23], the authors have extended the PARAFAC theory for frequency and DOA estimation. By using space-time matrix technique, the frequency and DOA estimation have been completed as well. However, the positions of four guiding elements have to be known as a prior [24].

In this paper, a novel DOA and mutual coupling coefficient estimation algorithm for conformal array is proposed in vehicle communication system. In realistic application, the effect of mutual coupling should not be ignored. Based on the selection matrix, the spectial mutual coupling matrix (MCM) is constructed. The effect of mutual coupling can be eliminated by using the spectial MCM. Then two rotational invariant matrices are constructed in order to estimate the DOA of incident signals based on conformal array. Based on PARAFAC theory, the model is constructed in cumulant domain using covariance matrices. The mutual coupling coefficient is estimated finally via matrix transformation.

2. The Snapshot Data Model

The algorithm performance would deteriorate significantly because of the mutual coupling among different elements. In order to estimate DOA estimation accurately, the effect of mutual coupling should be considered in the snapshot data model of the conformal array. Due to the varying curvature of carriers, each element possesses different patterns [18]. The incident signals are shown in Figure 1(a) with elevation θ and azimuth φ. We can give the form of steering vector as follows:

$$\mathbf{a}(\theta,\varphi) = \left[h_1 e^{-j2\pi((\mathbf{p}_1\cdot\mathbf{u})/\lambda)}, h_2 e^{-j2\pi((\mathbf{p}_2\cdot\mathbf{u})/\lambda)}, \ldots, h_{2M} e^{-j2\pi((\mathbf{p}_{2M}\cdot\mathbf{u})/\lambda)}\right]^T, \tag{1}$$

$$h_m = \left(g_{m\theta}^2 + g_{m\varphi}^2\right)^{1/2}\left(k_\theta^2 + k_\varphi^2\right)^{1/2}\cos\left(\theta_{mgk}\right) = |g_m| \cdot |p_l|\cos\left(\theta_{mgk}\right) = \mathbf{g}_m\cdot\mathbf{p}_l = g_{m\theta}k_\theta + g_{m\varphi}k_\varphi. \tag{2}$$

The notation of alphabet mentioned above can be found in [25]. Taking (2) into (1), we have

$$\mathbf{a}(\theta,\varphi) = \mathbf{a}_\theta(\theta,\varphi) + \mathbf{a}_\varphi(\theta,\varphi), \tag{3}$$

where

$$\mathbf{a}_\theta(\theta,\varphi) = \left[g_{1\theta}k_\theta e^{-j2\pi((\mathbf{p}_1\cdot\mathbf{u})/\lambda)}, g_{2\theta}k_\theta e^{-j2\pi((\mathbf{p}_2\cdot\mathbf{u})/\lambda)}, \ldots, g_{2M\theta}k_\theta e^{-j2\pi((\mathbf{p}_{2M}\cdot\mathbf{u})/\lambda)}\right]^T,$$
$$\mathbf{a}_\varphi(\theta,\varphi) = \left[g_{1\varphi}k_\varphi e^{-j2\pi((\mathbf{p}_1\cdot\mathbf{u})/\lambda)}, g_{2\varphi}k_\varphi e^{-j2\pi((\mathbf{p}_2\cdot\mathbf{u})/\lambda)}, \ldots, g_{2M\varphi}k_\varphi e^{-j2\pi((\mathbf{p}_{2M}\cdot\mathbf{u})/\lambda)}\right]^T. \tag{4}$$

When P incident signals arrive in the array, the data matrix of array output can be expressed as

$$\mathbf{X}(n) = \mathbf{G}\odot\mathbf{CAS}(n) + \mathbf{N}(n) = \mathbf{C}\left(\mathbf{G}_\theta\odot\mathbf{A}_\theta\mathbf{K}_\theta + \mathbf{G}_\varphi\odot\mathbf{A}_\varphi\mathbf{K}_\varphi\right)\mathbf{S}(n) + \mathbf{N}(n) = \mathbf{CBS}(n) + \mathbf{N}(n), \tag{5}$$

$$\mathbf{G}_\theta = \left[\mathbf{g}_\theta(\theta_1,\varphi_1), \mathbf{g}_\theta(\theta_2,\varphi_2), \ldots, \mathbf{g}_\theta(\theta_P,\varphi_P)\right], \tag{6}$$

$$\mathbf{G}_\varphi = \left[\mathbf{g}_\varphi(\theta_1,\varphi_1), \mathbf{g}_\varphi(\theta_2,\varphi_2), \ldots, \mathbf{g}_\varphi(\theta_P,\varphi_P)\right], \tag{7}$$

$$\mathbf{A}_\theta = \left[\mathbf{a}_\theta(\theta_1,\varphi_1), \mathbf{a}_\theta(\theta_2,\varphi_2), \ldots, \mathbf{a}_\theta(\theta_P,\varphi_P)\right], \quad (8)$$

$$\mathbf{A}_\varphi = \left[\mathbf{a}_\varphi(\theta_1,\varphi_1), \mathbf{a}_\varphi(\theta_2,\varphi_2), \ldots, \mathbf{a}_\varphi(\theta_P,\varphi_P)\right], \quad (9)$$

$$\mathbf{K}_\theta = \operatorname{diag}\left(k_{1\theta}, k_{2\theta}, \ldots, k_{P\theta}\right), \quad (10)$$

$$\mathbf{K}_\varphi = \operatorname{diag}\left(k_{1\varphi}, k_{2\varphi}, \ldots, k_{P\varphi}\right), \quad (11)$$

$$\mathbf{S}(n) = \left[s_1(n), s_2(n), \ldots, s_P(n)\right]^T, \quad (12)$$

$$\mathbf{W}(n) = \left[w_1(n), w_2(n), \ldots, w_{2M}(n)\right]^T, \quad (13)$$

where $\mathbf{G}$ denotes the pattern matrix and $\mathbf{A}$ denotes the manifold matrix. $\odot$ stands for the Kronecker product. $k_{1\theta}, k_{2\theta}, \ldots, k_{r\theta}$ and $k_{1\varphi}, k_{2\varphi}, \ldots, k_{r\varphi}$ construct $\mathbf{K}_\theta$ and $\mathbf{K}_\varphi$, respectively. The ith signal's polarization parameters are $k_{i\theta}$ and $k_{i\varphi}$, respectively. The signal vector is $\mathbf{S}(n)$. The covariance matrix of noise $\mathbf{W}(n)$ is

$$E\left\{\mathbf{W}(n)\,\mathbf{W}(n)^H\right\} = \mathbf{Q} = \sigma^2\mathbf{I}. \quad (14)$$

$(\cdot)^H$ denotes conjugate transpose of matrix $(\cdot)$. $\mathbf{I}$ is the identical matrix.

For simplicity, only the effect of mutual coupling among the same uniform linear array (ULA) is taken into consideration. The mutual coupling matrix (MCM) $\mathbf{C}$ is modeled as a banded symmetric Toeplitz matrix. $\mathbf{c} = [c_1, c_2, \ldots, c_q, 0, \ldots, 0]$ is the first row of $\mathbf{C}$. The element of vector $\mathbf{c}$ satisfying $0 < |c_q| < |c_{q-1}| < \cdots < c_1 = 1$. This matrix is formulated as

$$\mathbf{C} = \text{Toeplitz}(\mathbf{c}) = \text{Toeplitz}\left\{\left[c_1, c_2, \ldots, c_q, 0, \ldots, 0\right]\right\} = \begin{bmatrix} 1 & c_1 & \cdots & c_q & \cdots & 0 \\ c_1 & 1 & c_1 & \cdots & \ddots & 0 \\ \vdots & c_1 & 1 & \ddots & \cdots & c_q \\ c_q & \cdots & \ddots & \ddots & c_1 & \vdots \\ 0 & \ddots & \cdots & c_1 & 1 & c_1 \\ 0 & \cdots & c_q & \cdots & c_1 & 1 \end{bmatrix}. \quad (15)$$

Collecting N snapshots, we have the matrix form of (5) as

$$\mathbf{X} = \mathbf{CBS} + \mathbf{W}. \quad (16)$$

There is a characteristic called "shadow effect" which most conformal arrays possess. It means that when the incident signal impinges on the array, not all elements could receive it. In order to solve this problem, the subarray divided technique (SDT) proposed in [18, 19] is adopted in this paper. The whole conformal array consists of several subarrays, and each subarray covers a certain range of angle. For the single curvature and symmetry of the cylinder, the parameter estimation mechanism and array design of each subarray are identical. Thus only one subarray is considered in this paper.

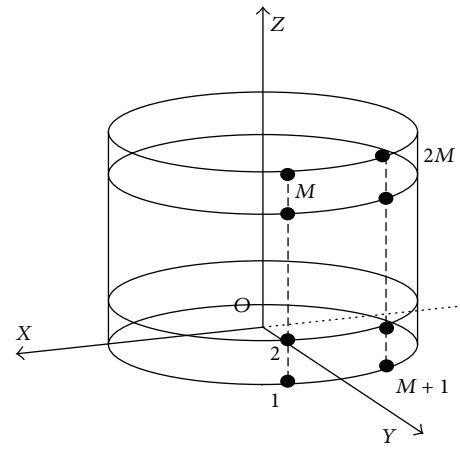


Figure 2: The structure of cylindrical conformal array.

3. The Mechanism of Mutual Coupling Elimination

The design of cylindrical conformal array is shown in Figure 2, the distance between two elements in the identical plane is $\lambda/2$, and the distance between two neighbouring planes is $\lambda/2$. The radius of the cylinder is 5λ.

According the form of the MCM $\mathbf{C}$, a mutual coupling elimination mechanism is introduced based on instrumental elements [26]. To show this mechanism in an intuitive way, we give an example based on $1 \sim M$ elements. The position vector $\mathbf{p}_m$ of the mth element can be expressed as

$$\mathbf{p}_m = x_m\vec{\mathbf{e}}_x + y_m\vec{\mathbf{e}}_y + z_m\vec{\mathbf{e}}_z. \quad (17)$$

Since M elements are arranged on the same generatrix, the x-coordinate has the relationship $x_1 = x_2 = \cdots = x_M = x$ as well as $y_1 = y_2 = \cdots = y_M = y$ and $h_1 = h_2 = \cdots = h_M$. The manifold matrix of this subarray is $\mathbf{B}_1 = [\mathbf{b}_1(\theta_1), \mathbf{b}_1(\theta_2), \ldots, \mathbf{b}_1(\theta_P)]$, and the steering vector is

$$\mathbf{b}_1\left(\theta_p\right) = \left[H_p, H_p v_p^1, \ldots, H_p v_p^{M-1}\right]^T, \quad (18)$$

where $v_p = \exp(-j2\pi d\cos(\theta_p)/\lambda)$ and

$$H_p = h_1 e^{-j2\pi((x\sin(\theta_p)\cos(\varphi_p) + y\sin(\theta_p)\sin(\varphi_p))/\lambda)}. \quad (19)$$

The instrumental elements are set as the first and last q elements. Although the array aperture is reduced, the effect of mutual coupling is eliminated as well. We construct a selection matrix that has the following form: $\mathbf{F} = [\mathbf{0}_{(M-2q)\times q}\mathbf{I}_{M-2q}\mathbf{0}_{(M-2q)\times q}]$, and the received data $\overline{\mathbf{X}}$ of center array can be given by

$$\overline{\mathbf{X}} = \mathbf{PCB}_1\mathbf{S} + \mathbf{PW} = \overline{\mathbf{C}}\mathbf{B}_1\mathbf{S} + \mathbf{PW}, \quad (20)$$

where the new $\overline{M} \times M$ $\overline{\mathbf{C}}$ is expressed as

$$\overline{\mathbf{C}} = \mathbf{PC} = \begin{bmatrix} c_q & \cdots & 1 & \cdots & c_q & 0 & \cdots & 0 \\ 0 & c_q & \cdots & 1 & \cdots & c_q & \cdots & 0 \\ \vdots & \ddots & \ddots & \cdots & \ddots & \cdots & \ddots & \vdots \\ 0 & \cdots & 0 & c_q & \cdots & 1 & \cdots & c_q \end{bmatrix}. \quad (21)$$

By using instrumental elements, we can get an important equation as follows:

$$\overline{\mathbf{C}}\mathbf{b}_1\left(\theta_p\right) = \begin{bmatrix} H_p\left(c_q v_p^{-q} + \cdots + c_1 v_p^{-q} + 1 + c_1 v_p + \cdots + c_q v_p^{q}\right) \\ H_p\left(c_q v_p^{1-q} + \cdots + c_1 + v_p + c_1 v_p^{2} + \cdots + c_q v_p^{q+1}\right) \\ \vdots \\ H_p\left(c_q v_p^{M-q-2} + \cdots + c_1 v_p^{M-2} + v_p^{M-1} + c_1 v_p^{M} + \cdots + c_q v_p^{M+q-2}\right) \\ H_p\left(c_q v_p^{M-q-1} + \cdots + c_1 v_p^{M-1} + v_p^{M} + c_1 v_p^{M+1} + \cdots + c_q v_p^{M+q-1}\right) \end{bmatrix}$$

$$= H_p\left(c_q v_p^{-q} + \cdots + c_1 v_p^{-1} + 1 + c_1 v_p + \cdots + c_q v_p^{q}\right) \begin{bmatrix} 1 \\ v_p \\ \vdots \\ v_p^{\overline{M}-1} \end{bmatrix} \tag{22}$$

$$= \left(2H_p \sum_{q'=1}^{q} c_{q'} \cos\left(\frac{2q'\pi \sin\left(\theta_p\right) d}{\lambda}\right) + 1\right) \overline{\mathbf{b}}_1\left(\theta_p\right) = c\left(\theta_p\right) \overline{\mathbf{b}}_1\left(\theta_p\right),$$

where $\overline{\mathbf{b}}_1(\theta_p)$ is the ideal steering vector of the center array ($q+1 \sim M-q$ elements), $\overline{M} = M - 2p$, and $c(\theta_p)$ is a scalar which can be written as

$$c\left(\theta_p\right) = \left(2\sum_{q'=1}^{q} c_{q'} \cos\left(\frac{2q'\pi \sin\left(\theta_p\right) d}{\lambda}\right) + 1\right). \tag{23}$$

Based on the (21), the new covariance matrix can be expressed as

$$\begin{aligned} \overline{\mathbf{R}}_x &= E\left\{\overline{\mathbf{X}}\overline{\mathbf{X}}^H\right\} = \overline{\mathbf{C}}\mathbf{B}_1 \overline{\mathbf{R}}_s \mathbf{B}_1^H \overline{\mathbf{C}}^H + \sigma^2 \mathbf{I} \\ &= \overline{\mathbf{B}}_1 \mathbf{D} \overline{\mathbf{R}}_s \mathbf{D}^H \overline{\mathbf{B}}_1^H + \sigma^2 \mathbf{I}, \end{aligned} \tag{24}$$

where

$$\begin{aligned} \overline{\mathbf{B}}_1 &= \left[\overline{\mathbf{b}}_1\left(\theta_1\right), \overline{\mathbf{b}}_1\left(\theta_2\right), \ldots, \overline{\mathbf{b}}_1\left(\theta_P\right)\right], \\ \mathbf{D} &= \operatorname{diag}\left\{c\left(\theta_1\right), c\left(\theta_2\right), \ldots, c\left(\theta_P\right)\right\}. \end{aligned} \tag{25}$$

Then (24) can be written as

$$\overline{\mathbf{R}}_x = \overline{\mathbf{B}}_1 \mathbf{R} \overline{\mathbf{B}}_1^H + \sigma^2 \mathbf{I}, \tag{26}$$

where

$$\mathbf{R} = \operatorname{diag}\left\{\sigma_1^2 \left|c\left(\theta_1\right)\right|^2, \sigma_2^2 \left|c\left(\theta_2\right)\right|^2, \ldots, \sigma_P^2 \left|c\left(\theta_P\right)\right|^2\right\} \tag{27}$$

and σ_p^2 stands for the power of pth incident signal. A very interesting result is shown in (27) where the mutual coupling coefficients are coming into the novel signal covariance matrix completely. When $c(\theta_p) \neq 0$, $p = 1, 2, \ldots, P$, $\operatorname{rank}(\mathbf{R}) = P$, the elimination of effect of mutual coupling can be accomplished. When $c(\theta_p) = 0$, $p = 1, 2, \ldots, P$, we have

$$c\left(\theta_p\right) \overline{\mathbf{b}}\left(\theta_p\right) = 0. \tag{28}$$

Equation (33) implies that the center array is blind at some particular angles. It could not receive incident signal from certain directions. Thus, these angles are called "blind angles." Fortunately, for given mutual coefficients, $c(\overline{\theta}_l)$ is a continuous function, the probability of $c(\overline{\theta}_l) = 0$ is approximate to zero, which means that "blind angles" phenomenon rarely happens in practice. More details can be found in [27].

As shown in Figure 3, there are two rotational invariance relations in the designed array. The ESPRIT algorithm which is similar to the algorithm described in [17] can be used for DOA estimation. The parameter pairing between elevation and azimuth has to be considered. However, the algorithm is not suitable for estimating conformal array. Based on PARAFAC theory, a novel high-accuracy 2D-DOA estimation algorithm is proposed in this paper for conformal array.

4. DOA Estimation Based on PARAFAC

When the elimination of mutual coupling is done, the decoupling between polarization and direction can be completed

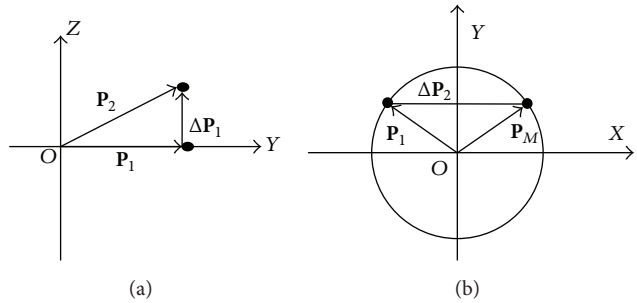


Figure 3: The distance vector. (a) The distance vector $\Delta\mathbf{P}_1$ between array 1 and array 2. (b) The distance vector $\Delta\mathbf{P}_2$ between array 3 and array 4.

based on well array design. Define two $(\overline{M}-1)\times\overline{M}$ selection matrices:

$$\mathbf{P}_1 = \begin{bmatrix} 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix}, \quad \mathbf{P}_2 = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}. \tag{29}$$

$q+1 \sim M-q-1$, $q+2 \sim M-q$, $q+1 \sim M-q$, and $M+q+1 \sim 2M-q$ elements constitute array 1, array 2, array 3, and array 4, respectively. The distance vector between arrays 1 and 2 is $\Delta\mathbf{P}_1$ and $d_1 = |\Delta\mathbf{P}_1| = \lambda/2$. The distance vector between arrays 1 and 3 is $\Delta\mathbf{P}_2$ and $d_2 = |\Delta\mathbf{P}_2| = \lambda/2$, which is shown in Figure 3. Since the direction pattern is identical through one generatrix, the effect of polarization parameter can be ignored.

The distance among different subarrays is shown in Figures 2 and 3. In the global coordinate, the elevation and azimuth of $\Delta\mathbf{P}_1$ and $\Delta\mathbf{P}_2$ can be represented as $\theta_{\Delta\mathbf{P}_1} = 0$, $\varphi_{\Delta\mathbf{P}_1} = \pi/2$, and $\theta_{\Delta\mathbf{P}_2} = \pi/2$, $\varphi_{\Delta\mathbf{P}_2} = 0$, respectively.

Assume that $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4$ represent the constructed data of array 1, array 2, array 3, and array 4, respectively. The origin of the coordinate is the reference point. So $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4$ can be expressed as

$$\mathbf{X}_1 = \mathbf{BS} + \mathbf{N}_1, \tag{30}$$

$$\mathbf{X}_2 = \mathbf{B\Psi}_1\mathbf{S} + \mathbf{N}_2, \tag{31}$$

$$\mathbf{X}_3 = \mathbf{B\Psi}_2\mathbf{S} + \mathbf{N}_3, \tag{32}$$

$$\mathbf{X}_4 = \mathbf{B\Psi}_1\mathbf{\Psi}_2\mathbf{S} + \mathbf{N}_4. \tag{33}$$

The covariance matrices among the received data are

$$\begin{aligned} \mathbf{R}_1 &= E\left\{\mathbf{X}_1\mathbf{X}_1^H\right\} = \mathbf{BR}_s\mathbf{B}^H + \mathbf{Q}_1, \\ \mathbf{R}_2 &= E\left\{\mathbf{X}_2\mathbf{X}_1^H\right\} = \mathbf{B\Psi}_1\mathbf{R}_s\mathbf{B}^H + \mathbf{Q}_2, \\ \mathbf{R}_3 &= E\left\{\mathbf{X}_3\mathbf{X}_1^H\right\} = \mathbf{B\Psi}_2\mathbf{R}_s\mathbf{B}^H + \mathbf{Q}_3, \\ \mathbf{R}_4 &= E\left\{\mathbf{X}_4\mathbf{X}_1^H\right\} = \mathbf{B\Psi}_1\mathbf{\Psi}_2\mathbf{R}_s\mathbf{B}^H + \mathbf{Q}_4, \end{aligned} \tag{34}$$

where $\mathbf{B} = \mathbf{P}_1\overline{\mathbf{B}}_1$, $\mathbf{R}_s = \mathrm{diag}\{s_1^2, \ldots, s_r^2\}$ represents the signal covariance matrix, and the noise covariance matrices are $\mathbf{Q}_1 \sim \mathbf{Q}_4$, respectively:

$$\begin{aligned} \mathbf{\Phi}_1 &= \mathrm{diag}\left[\exp\left(-j\omega_{11}\right), \exp\left(-j\omega_{12}\right), \ldots, \exp\left(-j\omega_{1r}\right)\right], \\ \mathbf{\Phi}_2 &= \mathrm{diag}\left[\frac{h_3\left(\theta_1,\varphi_1\right)}{h_1\left(\theta_1,\varphi_1\right)}\exp\left(-j\omega_{21}\right), \frac{h_3\left(\theta_2,\varphi_2\right)}{h_1\left(\theta_2,\varphi_2\right)}\right. \\ &\quad \left.\cdot\exp\left(-j\omega_{22}\right), \ldots, \frac{h_3\left(\theta_r,\varphi_r\right)}{h_1\left(\theta_r,\varphi_r\right)}\exp\left(-j\omega_{2r}\right)\right], \\ \omega_{1i} &= \left(\frac{2\pi}{\lambda}\right)\Delta\mathbf{P}_1\cdot\mathbf{u}_i = \left(\frac{2\pi d_1}{\lambda}\right)\left[\sin\left(\theta_{\Delta\mathbf{P}_1}\right)\cos\left(\varphi_{\Delta\mathbf{P}_1}\right)\right. \\ &\quad \cdot\sin\left(\theta_i\right)\cos\left(\varphi_i\right) + \sin\left(\theta_{\Delta\mathbf{P}_1}\right)\sin\left(\varphi_{\Delta\mathbf{P}_1}\right)\sin\left(\theta_i\right) \\ &\quad \left.\cdot\sin\left(\varphi_i\right) + \cos\left(\theta_{\Delta\mathbf{P}_1}\right)\cos\left(\theta_i\right)\right], \\ \omega_{2i} &= \left(\frac{2\pi}{\lambda}\right)\Delta\mathbf{P}_2\cdot\mathbf{u}_i = \left(\frac{2\pi d_2}{\lambda}\right)\left[\sin\left(\theta_{\Delta\mathbf{P}_2}\right)\cos\left(\varphi_{\Delta\mathbf{P}_2}\right)\right. \\ &\quad \cdot\sin\left(\theta_i\right)\cos\left(\varphi_i\right) + \sin\left(\theta_{\Delta\mathbf{P}_2}\right)\sin\left(\varphi_{\Delta\mathbf{P}_2}\right)\sin\left(\theta_i\right) \\ &\quad \left.\cdot\sin\left(\varphi_i\right) + \cos\left(\theta_{\Delta\mathbf{P}_2}\right)\cos\left(\theta_i\right)\right], \end{aligned} \tag{35}$$

where $\theta_{\Delta\mathbf{P}_i}$ and $\varphi_{\Delta\mathbf{P}_i}$ $(i = 1, 2)$ represent the elevation and azimuth of the distance vector in the global coordinate, respectively.

On the basis of the PARAFAC theory [11], the $m \times m \times 4$ three-way array of the cylindrical conformal array is constructed by (34):

$$\begin{vmatrix} \mathbf{R}(:,:,1) \\ \mathbf{R}(:,:,2) \\ \mathbf{R}(:,:,3) \\ \mathbf{R}(:,:,4) \end{vmatrix} = \begin{vmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \\ \mathbf{R}_3 \\ \mathbf{R}_4 \end{vmatrix} = \begin{vmatrix} \mathbf{BR}_s\mathbf{B}^H \\ \mathbf{B\Psi}_1\mathbf{R}_s\mathbf{B}^H \\ \mathbf{B\Psi}_2\mathbf{R}_s\mathbf{B}^H \\ \mathbf{B\Psi}_1\mathbf{\Psi}_2\mathbf{R}_s\mathbf{B}^H \end{vmatrix} + \widetilde{\mathbf{Q}}, \tag{36}$$

where $\mathbf{R}_i$ $(i = 1, \ldots, 4)$ are the sample covariance matrices, respectively, and $\widetilde{\mathbf{Q}}$ represents the noise in real observation.

Let $\mathbf{C} = \mathbf{B}^H$, and, based on the definition of Khatri-Rao product, (36) can be transformed as

$$\mathbf{R} = (\mathbf{D} \odot \mathbf{B})\,\mathbf{C} + \widetilde{\mathbf{Q}}, \tag{37}$$

$$\mathbf{R}_X = \left(\mathbf{C}^T \odot \mathbf{D}\right)\mathbf{B}^T + \widetilde{\mathbf{Q}}_X, \tag{38}$$

$$\mathbf{R}_Y = \left(\mathbf{B} \odot \mathbf{C}^T\right)\mathbf{D}^T + \widetilde{\mathbf{Q}}_Y, \tag{39}$$

$$\mathbf{D} = \begin{bmatrix} \Lambda^{-1}(\mathbf{R}_s) \\ \Lambda^{-1}(\boldsymbol{\Psi}_1\mathbf{R}_s) \\ \Lambda^{-1}(\boldsymbol{\Psi}_2\mathbf{R}_s) \\ \Lambda^{-1}(\boldsymbol{\Psi}_1\boldsymbol{\Psi}_2\mathbf{R}_s) \end{bmatrix}, \tag{40}$$

where $\Lambda^{-1}(\mathbf{R}_s)$ is the row vector constructed by the diagonal entries of the diagonal matrix $\mathbf{R}_s$.

On the basis of noisy observation, problem (37) can be transformed into solving a least square problem:

$$\min_{\mathbf{D},\mathbf{B},\mathbf{C}} \|\mathbf{R} - (\mathbf{D} \odot \mathbf{B})\,\mathbf{C}\|_F^2. \tag{41}$$

The principle of alternating least squares (ALS) can be used to fit the problem in (41). In the noiseless condition, ALS can be used to solve the matrices $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ which constructed the three-way array $\mathbf{R}$. The least square estimation of matrix $\mathbf{C}$ can be expressed as

$$\mathbf{C} = \arg\min_{\mathbf{C}} \|\mathbf{R} - (\mathbf{D} \odot \mathbf{B})\,\mathbf{C}\|_F^2. \tag{42}$$

Similarly, the matrices $\mathbf{B}$ and $\mathbf{D}$ can be expressed as

$$\begin{aligned} \mathbf{B}^T &= \arg\min_{\mathbf{B}} \left\|\mathbf{R}_X - \left(\mathbf{C}^T \odot \mathbf{D}\right)\mathbf{B}^T\right\|_F^2, \\ \mathbf{D}^T &= \arg\min_{\mathbf{D}} \left\|\mathbf{R}_Y - \left(\mathbf{B} \odot \mathbf{C}^T\right)\mathbf{D}^T\right\|_F^2. \end{aligned} \tag{43}$$

In the iterative procedure, given matrices $\mathbf{B}$ and $\mathbf{D}$, the matrix $\mathbf{C}$ can be represented as

$$\mathbf{C} = (\mathbf{D} \odot \mathbf{B})^{\dagger}\,\mathbf{R}. \tag{44}$$

The expression of matrices $\mathbf{B}^T$ and $\mathbf{D}^T$ is

$$\begin{aligned} \mathbf{B}^T &= \left(\mathbf{C}^T \odot \mathbf{D}\right)^{\dagger}\mathbf{R}_X, \\ \mathbf{D}^T &= \left(\mathbf{B} \odot \mathbf{C}^T\right)^{\dagger}\mathbf{R}_Y, \end{aligned} \tag{45}$$

where $(\cdot)^{\dagger}$ denotes the pseudoinverse of matrix $(\cdot)$.

The matrix $\mathbf{D}$ can be acquired by the TALS algorithm. ω_{1i} and ω_{2i} can be calculated by matrix $\mathbf{D}$:

$$\omega_{1i} = -\frac{1}{2}\left(\text{angle}\left\lceil\frac{\mathbf{D}_{2i}}{\mathbf{D}_{1i}}\right\rceil + \text{angle}\left\lceil\frac{\mathbf{D}_{4i}}{\mathbf{D}_{3i}}\right\rceil\right) \tag{46}$$

$\lceil\cdot\rceil$ stands for the matrix and $\lceil\cdot\rceil$ stands for the absolute value operation, where $\mathbf{D}_{ji}$ represents the jth row of $\mathbf{D}$. Because h_1, h_3, and h_4 are real numbers, $\mathbf{D}_{3i}/\mathbf{D}_{1i}$ and $\mathbf{D}_{4i}/\mathbf{D}_{2i}$ are squared to solve the ambiguity caused by the positive and negative values of h_1, h_3, and h_4:

$$\begin{aligned} \omega_{2i} &= -\frac{1}{2}\text{angle}\left(\left[\frac{h_3(\theta_i,\varphi_i)}{h_1(\theta_i,\varphi_i)}\exp(-j\omega_{2i})\right]^2\right) \\ &= -\frac{1}{2}\text{angle}\left(\exp(-j2\omega_{2i})\right) = -\frac{1}{2}\text{angle}\left(\left[\frac{\mathbf{D}_{3i}}{\mathbf{D}_{1i}}\right]^2\right) \\ &= -\frac{1}{2}\text{angle}\left(\left[\frac{\mathbf{D}_{4i}}{\mathbf{D}_{2i}}\right]^2\right), \end{aligned} \tag{47}$$

and then

$$\omega_{2i} = -\frac{1}{4}\left|\text{angle}\left(\left[\frac{\mathbf{D}_{3i}}{\mathbf{D}_{1i}}\right]^2\right) + \text{angle}\left(\left[\frac{\mathbf{D}_{4i}}{\mathbf{D}_{2i}}\right]^2\right)\right|. \tag{48}$$

The elevation θ_i and azimuth φ_i of ith incident signal can be expressed as

$$\theta_i = \arccos\left(\frac{\lambda\omega_{1i}}{2\pi d_2}\right) = \arccos\left(\frac{2\omega_{1i}}{\pi}\right), \tag{49}$$

$$\begin{aligned} \varphi_i &= \arccos\left(\frac{\lambda\omega_{2i}}{2\pi d_2\sin(\theta_i)}\right) \\ &= \arccos\left(\frac{2\omega_{2i}}{\pi\sin(\theta_i)}\right). \end{aligned} \tag{50}$$

5. Mutual Coupling Coefficient Estimation

Based on the 2D-DOA estimation of incident signals, the estimations of mutual coupling coefficients can be obtained. The covariance matrix of $\mathbf{X}$ can be expressed as

$$\mathbf{R}_X = E\left\{\mathbf{X}\mathbf{X}^H\right\} = \mathbf{C}\mathbf{A}\mathbf{R}_s\mathbf{A}^H\mathbf{C}^H + \sigma_n^2\mathbf{I}_N. \tag{51}$$

The rank of $\mathbf{C}\mathbf{A}\mathbf{R}_s\mathbf{A}^H\mathbf{C}^H$ is P. Take the eigendecomposition of $\mathbf{R}_X$; the P big eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_P$ and $M - P$ small eigenvalues $\lambda_{P+1} = \lambda_{P+2} = \cdots = \lambda_K = \sigma_n^2$ can be obtained, respectively. Their corresponding eigenvectors are $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_M$. The signal subspace is spanned by $\mathbf{CA}$, which is orthogonal to the noise subspace spanned by $\mathbf{u}_{K_u+P+1}, \mathbf{u}_{K_u+P+2}, \ldots, \mathbf{u}_M$:

$$\mathbf{U}_n^H\mathbf{C}\mathbf{a}(\theta_i) = 0, \quad i = 1, 2, \ldots, P. \tag{52}$$

By considering the complex symmetric Toeplitz form of $\mathbf{C}$, $\mathbf{Ca}(\theta)$ can be expressed as

$$\mathbf{Ca}(\theta) = \begin{bmatrix} 1 & c_1 & \cdots & c_{N-1} \\ c_1 & 1 & \cdots & c_{N-2} \\ \vdots & \vdots & \vdots & \vdots \\ c_{N-1} & c_{N-2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} 1 \\ v \\ \vdots \\ v^{M-1} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & v & \cdots & v^{M-2} & v^{M-1} \\ v & 1+v^2 & \cdots & v^{M-1} & 0 \\ v^2 & v+v^3 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ v^{M-2} & v^{M-3}+v^{M-1} & \cdots & 1 & 0 \\ v^{M-1} & v^{M-2} & \cdots & v & 1 \end{bmatrix} \begin{bmatrix} 1 \\ c_1 \\ \vdots \\ c_{N-1} \end{bmatrix} \quad (53)$$

$$= \mathbf{T}(\theta)\,\mathbf{c},$$

where $\mathbf{T}(\theta)$ is the sum of two $M \times (M_0 + 1)$ matrices [28].

Based on (52) and the estimations of $\theta_1, \theta_2, \ldots, \theta_P$, we have

$$\mathbf{U}_n^H \mathbf{T}(\theta_i)\,\mathbf{c} = 0, \quad i = 1, 2, \ldots, P. \quad (54)$$

It can be seen that (54) is the linear equations of mutual coupling coefficients $\mathbf{c}$. The coefficient matrix can be defined as

$$\mathbf{Q} = \begin{bmatrix} \mathbf{U}_n^H \mathbf{T}(\theta_1) \\ \vdots \\ \mathbf{U}_n^H \mathbf{T}(\theta_P) \end{bmatrix}. \quad (55)$$

Then (50) can be written as

$$\mathbf{Qc} = 0, \quad (56)$$

where $\mathbf{Q}$ is a $P(M - P) \times (M_0 + 1)$ matrix and $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_P]$. Due to $\mathbf{c}(1) = 1$, we have

$$\mathbf{Qc} = \left[\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_{M_0+1}\right] \begin{bmatrix} 1 \\ c_1 \\ \vdots \\ c_{M_0} \end{bmatrix} = 0. \quad (57)$$

When $P(M-P) \geq M_0 - 1$, the least square solution is obtained as

$$\left[c_1, c_2, \ldots, c_{M_0}\right] = -\left[\mathbf{q}_2, \ldots, \mathbf{q}_{M_0+1}\right]^{\dagger} \mathbf{q}_1. \quad (58)$$

The mutual coupling coefficients estimation is completed.

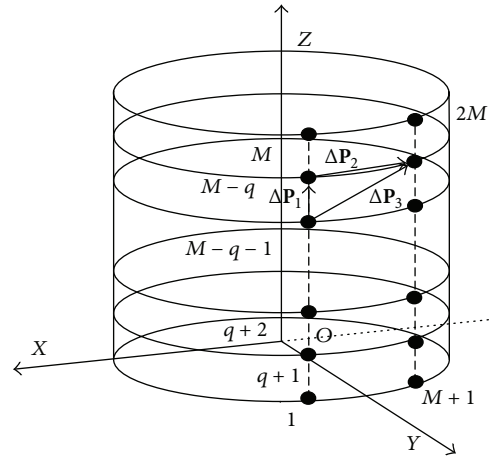


Figure 4: The distance vector between distinct subarrays.

6. Discussion

The estimators of $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ have the same column permutation matrix; that is, the ith column of the steering matrix $\mathbf{B}$ corresponds to the ith of the matrix $\mathbf{D}$. Thus, the elevation and azimuth pair with each other automatically.

When $P \geq 2$, the parameter pairing between ω_{1i} $(i = 1, 2, \ldots, P)$ and ω_{2i} $(i = 1, 2, \ldots, P)$ is a big problem in [21]. Here, we give another method to solve the parameter pairing problem in [21]. Theoretically, the eigenvector of $\mathbf{\Phi}_1$, $\mathbf{\Phi}_2$, and $\mathbf{\Phi}_3$ is identical. However, the EVD of them are done independently. The order of eigenvector may be distinct. In order to obtain accurate DOA estimation, this problem must be solved. We construct another subarray: array 1 and array 5 constructed by $M + q + 2 \sim 2M - q$ elements. The distance vector is $\Delta\mathbf{P}_3$. Based on the relationship between $\Delta\mathbf{P}_1$, $\Delta\mathbf{P}_2$, and ω_{1i}, ω_{2i}, the relationship between $\Delta\mathbf{P}_3$ and ω_{3k} $(k = 1, 2, \ldots, P)$ can be obtained similarly. As shown in Figure 4, $\Delta\mathbf{P}_3 = \Delta\mathbf{P}_1 + \Delta\mathbf{P}_2$.

For the identical source, we have

$$\begin{aligned} &\text{angle}\left(\left[\frac{h_2(\theta_p, \varphi_p)}{h_1(\theta_p, \varphi_p)} \exp\left(-j\omega_{3p}\right)\right]^2\right) \\ &= \text{angle}\left(\left(t_{3k}\right)^2\right) \\ &= \text{angle}\left(\exp\left(-j2\omega_{1i}\right)\right) + \text{angle}\left(\exp\left(-j2\omega_{2i}\right)\right) \\ &= \text{angle}\left(\left(t_{1i}\right)^2\right) + \text{angle}\left(\left(t_{2j}\right)^2\right), \end{aligned} \quad (59)$$

where t_{3k} is the eigenvalue of $\mathbf{\Phi}_3$. $\mathbf{\Phi}_3$ can be solved easily according to the solution procedure of $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$. Thus, the parameter pairing problem can be transformed as a problem which minimizes

$$\begin{aligned} &\min \Big| \text{angle}\left(\left(t_{3k}\right)^2\right) \\ &\quad - \left[\text{angle}\left(\left(t_{1i}\right)^2\right) + \text{angle}\left(\left(t_{2j}\right)^2\right)\right] \Big|. \end{aligned} \quad (60)$$

Based on (60), the parameter pairing is completed finally.

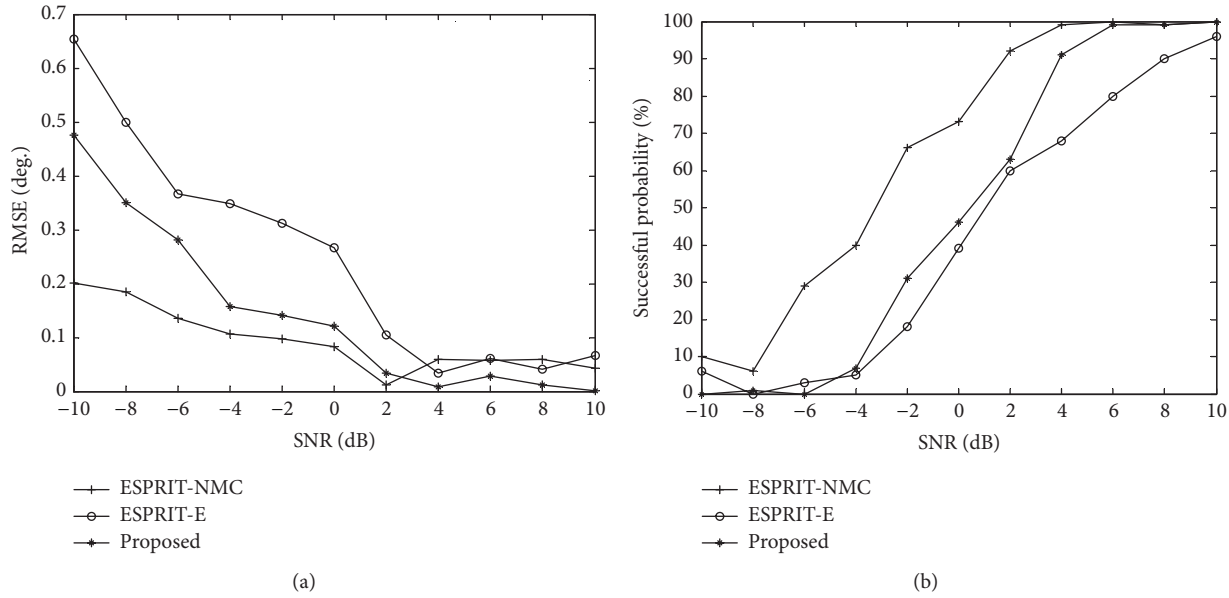


Figure 5: The performance against SNR (a). The RMSE against SNR (b). The successful probability against SNR.

The steps of DOA estimation for cylindrical conformal array in the presence of mutual coupling can be summarized as follows:

(1) Use the selection matrix $\mathbf{F} = [\mathbf{0}_{(M-2q)\times q}\mathbf{I}_{M-2q}\mathbf{0}_{(M-2q)\times q}]$ to eliminate the effect of mutual coupling.

(2) Set array 1~array 4 to construct three subarrays.

(3) Construct the PARAFAC model based on (36).

(4) Update (44), (45), and the matrix $\mathbf{D}$ can be obtained by TALS algorithm.

(5) Calculate ω_{1i} and ω_{2i} based on (46) and (47).

(6) Estimate the 2D-DOA for cylindrical conformal array based on (49) and (50).

(7) Calculate the covariance matrix of the data received by element $1 \sim M$ based on (51).

(8) Obtain the mutual coupling coefficients based on (58).

7. Simulation Results

In this section, we present numerical simulation results to illustrate the performance of the proposed algorithm. We compare the proposed algorithm with ESPRIT without mutual coupling (ESPRIT-NMC) [5] and the ESPRIT-E algorithm proposed in [21]. In this paper, root mean square error (RMSE) is calculated by the formula

$$\text{RMSE} = \sqrt{\frac{1}{200}\sum_{\eta=1}^{200}\left[\left(\widehat{\theta}_{k,\eta}-\theta\right)^2+\left(\widehat{\varphi}_{k,\eta}-\varphi\right)^2\right]}, \quad (61a)$$

$$\text{RMSE}_c = \sqrt{\frac{1}{200\,\|c\|}\sum_{\eta=1}^{200}\left\|\widehat{c}\,(\eta)-c\right\|}\times 100\%, \quad (61b)$$

where $\widehat{\theta}_{k,\eta}$ and $\widehat{\varphi}_{k,\eta}$ are the estimate of the elevation and azimuth of the kth incident signal in the ηth trial. For the sake of simplicity, only the cylindrical conformal array is utilized for simulation, and the array structure is shown in Figure 2. The number of elements is 16; that is, $m = 8$. Without loss of generality, $k_{1\theta} = 0.5$, $k_{1\varphi} = 0.5$; $k_{2\theta} = 0.3$, $k_{2\varphi} = 0.7$. The pattern transformation can be found in [25]. The azimuth and elevation of two incident signals are (100°, 60°) and (95°, 50°), respectively. The freedom of degree (FOD) of mutual coupling is 2, and $c_1 = 0.4500 + i0.5362$ and $c_2 = 0.2598 - i0.1500$. Since the instrumental elements are used, the elements actually used are 12; that is, $m = 6$. For ESPRIT-NMC, the elements actually used are 12; that is, $m = 6$. 200 Monte Carlo trials are considered in the simulations.

The RMSE and successful probability of different algorithms against SNR are shown in Figures 5(a) and 5(b), respectively. The snapshot number is 200. It can be seen from Figure 5(a) that the RMSE of the proposed algorithm is smaller than that of ESPRIT-E. Since the mutual coupling coefficient has been estimated, this result can be used to recover the original covariance matrix. The array aperture is extended compared with ESPRIT-E and ESPRIT-NMC. It can be seen from Figure 5(b) that the successful probability increase as SNR increases. The successful probability of ESPRIT-E is higher than that of ESPRIT-NMC. The proposed algorithm outperforms ESPRIT-E and ESPRIT-NMC.

The RMSE and successful probability of different algorithms against snapshot number are shown in Figures 5(a) and 5(b), respectively. SNR is fixed at 0 dB. It can be seen from Figure 6(a) that the RMSE of proposed algorithm is smaller than that of ESPRIT-E. When snapshot number is larger than 700, the RMSE of proposed algorithm approximates ESPRIT-NMC. The successful probability of proposed algorithm is higher than that of ESPRIT-E. However, the proposed algorithm does not reach 100% because of the low SNR (0 dB).

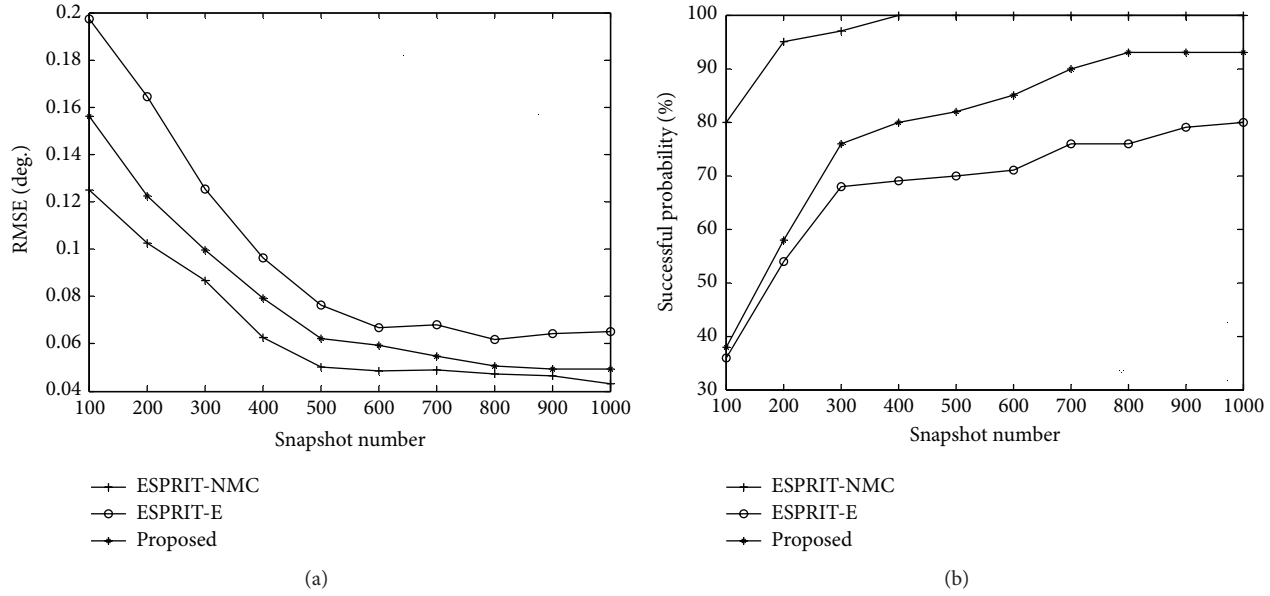


Figure 6: The performance against snapshot number (a). The RMSE against snapshot number (b). The successful probability against snapshot number.

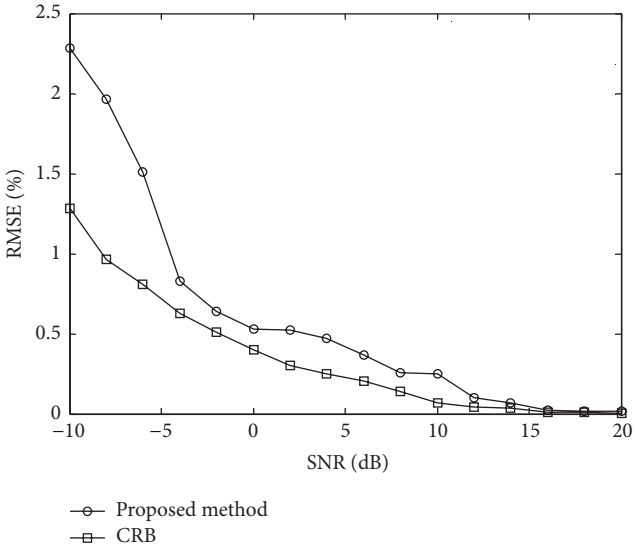


Figure 7: The RMSE against SNR.

Thus, the proposed algorithm has better performance compared with ESPRIT-E.

The mutual coupling estimation versus input SNR is shown in Figure 7. The results show that the mutual coupling has little influence on the proposed algorithm. Although some array apertures are used to eliminate the effect of mutual coupling, the proposed algorithm still has high estimation accuracy and approaches CRB at high SNR.

8. Conclusions

In this paper, a novel DOA and mutual coupling coefficient estimation algorithm for conformal array is proposed in ITS. By constructing the spectical mutual coupling matrix (MCM), the effect of mutual coupling can be eliminated via instrumental element method. Then the DOA of incident signals can be estimated based on parallel factor (PARAFAC) theory. The proposed algorithm can also be extended to other conformal arrays. However, only the mutual coupling in the identical linear array is considered in this paper. The mutual coupling between different generatrices should be studied. The Cramér-Rao bound is also an interesting problem to research, and we will focus on it in the future work.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported in part by Qing Lan Project, by the National Science Foundation of China under Grant nos. 61572172, 61401147, 61301095, and 61201237, and by the Natural Science Foundation of JiangSu Province of China, no. BK20140248. This work was supported in part by the Key Development Program of Basic Research of China (JCKY2013604B001), Nature Science Foundation of Heilongjiang Province of China (nos. QC2012C069 and F201408), and the Fundamental Research Funds for the Central Universities (no. HEUCF1508).

References

[1] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, 1996.

[2] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing-Concepts and Techniques*, Prentice Hall, Englewood Cliffs, NJ, USA, 1993.

[3] P. Stocia and R. Moses, *Introduction to Spectral Analysis*, Prentice Hall, Upper Saddle River, NJ, USA, 2004.

[4] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[5] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.

[6] T. T. Zhang, Y. L. Lu, and H. T. Hui, "Compensation for the mutual coupling effect in uniform circular arrays for 2D DOA estimations employing the maximum likelihood technique," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 3, pp. 1215–1221, 2008.

[7] L. Wan, G. Han, J. J. P. C. Rodrigues, W. Si, and N. Feng, "An energy efficient DOA estimation algorithm for uncorrelated and coherent signals in virtual MIMO systems," *Telecommunication Systems*, vol. 59, no. 1, pp. 93–110, 2015.

[8] Z. T. Huang, Z. M. Liu, J. Liu, and Y. Y. Zhou, "Performance analysis of MUSIC for non-circular signals in the presence of mutual coupling," *IET Radar, Sonar & Navigation*, vol. 4, no. 5, pp. 703–711, 2010.

[9] L. Wan, G. Han, L. Shu, and N. Feng, "The critical patients localization algorithm using sparse representation for mixed signals in emergency healthcare system," *IEEE Systems Journal*, 2015.

[10] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Blind PARAFAC receivers for DS-CDMA systems," *IEEE Transactions on Signal Processing*, vol. 48, no. 3, pp. 810–823, 2000.

[11] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Transactions on Signal Processing*, vol. 48, no. 8, pp. 2377–2388, 2000.

[12] J. L. Liang, "Joint azimuth and elevation direction finding using cumulant," *IEEE Sensors Journal*, vol. 9, no. 4, pp. 390–398, 2009.

[13] X.-F. Gong, Z.-W. Liu, and Y.-G. Xu, "Regularised parallel factor analysis for the estimation of direction-of-arrival and polarisation with a single electromagnetic vector-sensor," *IET Signal Processing*, vol. 5, no. 4, pp. 390–396, 2011.

[14] L. Josefsson and P. Persson, *Conformal Array Antenna Theory and Design*, IEEE Press, Piscataway, NJ, USA, 2006.

[15] W. T. Li, X. W. Shi, Y. Q. Hei, S. F. Liu, and J. Zhu, "A hybrid optimization algorithm and its application for conformal array pattern synthesis," *IEEE Transactions on Antennas and Propagation*, vol. 58, no. 10, pp. 3401–3406, 2010.

[16] W.-J. Si, L.-T. Wan, and Z.-X. Tian, "Fast DOA estimation based on cylindrical conformal array antenna," *Systems Engineering and Electronics*, vol. 35, no. 8, pp. 1589–1595, 2013.

[17] L. Wan, L. Liu, G. Han, and J. Rodrigues, "A low energy consumption DOA estimation approach for conformal array in ultra-wideband," *Future Internet*, vol. 5, no. 4, pp. 611–630, 2013.

[18] L. Wan, G. Han, J. Jiang, and L. Shu, "Distributed DOA estimation based on manifold separation technique in mobile wireless sensor networks," in *Proceedings of the 2nd Workshop on Mobile Sensing, Computing and Communication (MSCC '15)*, pp. 1–6, ACM, Hangzhou, China, June 2015.

[19] L. Wan, G. Han, L. Shu, N. Feng, C. Zhu, and J. Lloret, "Distributed parameter estimation for mobile wireless sensor network based on cloud computing in battlefield surveillance system," *IEEE Access*, vol. 3, pp. 1729–1739, 2015.

[20] W. Si, L. Wan, L. Liu, Z. Tian, and X. Lan, "Real-time ultra-wideband direction finding for the conformal array antenna," *Journal of Harbin Engineering University*, vol. 35, no. 7, pp. 913–918, 2014.

[21] H. Feng, L. Liu, and B. Wen, "2D-DOA estimation for cylindrical array with mutual coupling," *Mathematical Problems in Engineering*, vol. 2014, Article ID 716978, 8 pages, 2014.

[22] W. Si, L. Wan, L. Liu, Z. Tian, and L. Li, "Direction-of-arrival estimation for arbitrary array configurations in ultra-wideband," in *Proceedings of the 2nd International Conference on Instrumentation, Measurement, Computer, Communication and Control (IMCCC '12)*, pp. 234–237, IEEE, Harbin, China, December 2012.

[23] Y. Zou, H. Xie, L. Wan, and G. Han, "High accuracy frequency and 2D-DOAs estimation of conformal array based on PARAFAC," *Journal of Internet Technology*, vol. 16, no. 1, pp. 107–119, 2015.

[24] L.-T. Wan, L.-T. Liu, W.-J. Si, and Z.-X. Tian, "Joint estimation of 2D-DOA and frequency based on space-time matrix and conformal array," *The Scientific World Journal*, vol. 2013, Article ID 463828, 10 pages, 2013.

[25] L. Wan, W. Si, L. Liu, Z. Tian, and N. Feng, "High accuracy 2D-DOA estimation for conformal array using PARAFAC," *International Journal of Antennas and Propagation*, vol. 2014, Article ID 394707, 14 pages, 2014.

[26] X. Xu, Z. Ye, and Y. Zhang, "DOA estimation for mixed signals in the presence of mutual coupling," *IEEE Transactions on Signal Processing*, vol. 57, no. 9, pp. 3523–3532, 2009.

[27] J. Dai, W. Xu, and D. Zhao, "Real-valued DOA estimation for uniform linear array with unknown mutual coupling," *Signal Processing*, vol. 92, no. 9, pp. 2056–2065, 2012.

[28] Z. Ye and C. Liu, "On the resiliency of MUSIC direction finding against antenna sensor coupling," *IEEE Transactions on Antennas and Propagation*, vol. 56, no. 2, pp. 371–380, 2008.

Hindawi Publishing Corporation
Mobile Information Systems
Volume 2015, Article ID 369489, 12 pages
http://dx.doi.org/10.1155/2015/369489

Research Article

Analyzing User Awareness of Privacy Data Leak in Mobile Applications

Youngho Kim,[1] Tae Oh,[2] and Jeongnyeo Kim[1]

[1]*Electronics and Telecommunications Research Institute (ETRI), 161 Gaejong-Dong, Yuseong-Gu, Daejeon 34129, Republic of Korea*
[2]*Rochester Institute of Technology (RIT), 1 Lomb Memorial Drive, Rochester, NY 14623-5603, USA*

Correspondence should be addressed to Youngho Kim; wtowto@etri.re.kr

Received 7 August 2015; Revised 7 November 2015; Accepted 16 November 2015

Academic Editor: Kamal Deep Singh

To overcome the resource and computing power limitation of mobile devices in Internet of Things (IoT) era, a cloud computing provides an effective platform without human intervention to build a resource-oriented security solution. However, existing malware detection methods are constrained by a vague situation of information leaks. The main goal of this paper is to measure a degree of hiding intention for the mobile application (app) to keep its leaking activity invisible to the user. For real-world application test, we target Android applications, which unleash user privacy data. With the TaintDroid-ported emulator, we make experiments about the timing distance between user events and privacy leaks. Our experiments with Android apps downloaded from the Google Play show that most of leak cases are driven by user explicit events or implicit user involvement which make the user aware of the leakage. Those findings can assist a malware detection system in reducing the rate of false positive by considering malicious intentions. From the experiment, we understand better about app's internal operations as well. As a case study, we also presents a cloud-based dynamic analysis framework to perform a traffic monitor.

1. Introduction

Malicious code in the form of computer viruses and another malware is known to wreak havoc on IoT infrastructure as well as edge devices including mobile devices. Since mobile devices are in much wider use, the devices are more likely to be exposed to malicious code and environments similar to those devices that target enterprise systems. Antivirus and antimalware tools built for enterprise systems do not transfer well with mobile devices. A part of reasons is to the limited ability of mobile devices to efficiently run antivirus tools. Mobile device limitations include power issues due to reliance on batteries, fewer CPU cycles to dedicate to running protective software, and a smaller memory footprint to run the tools. Few literatures propose an architecture to perform mobile malware analysis in the cloud. The main purpose of the architecture is to identify the malware prior to activation on the mobile device. This can preempt the malicious code and mitigate the threat before it causes any harm to the user or the device.

However, it is quite challenging to detect an application as malware when information leakage really happens. In many cases, users are willing to send their private or sensitive data to a remote server in exchange for useful service(s) such as location-aware search services. Therefore, making a decision by a seemingly data exfiltration within a certain period of time can lead to false positives in identifying a malicious information leakage. A best way of getting the original intention of the outgoing data flow would be to ask the user if the data flow would be permissible with him/her. However, getting the intention from the user is not allowed in an automated cloud system. Instead, we choose to analyze the intention of the application causing outgoing data flow in terms of its hiding efforts in preventing the user from getting aware of the application's leaking activity.

The contributions of this paper are in three folds. First, we differentiate the data by a knowledgeable user's request and data exfiltration by malware. According to a recent study [1], Android malware tends to transmit private data without a user consent. Therefore, most of the malware samples listen

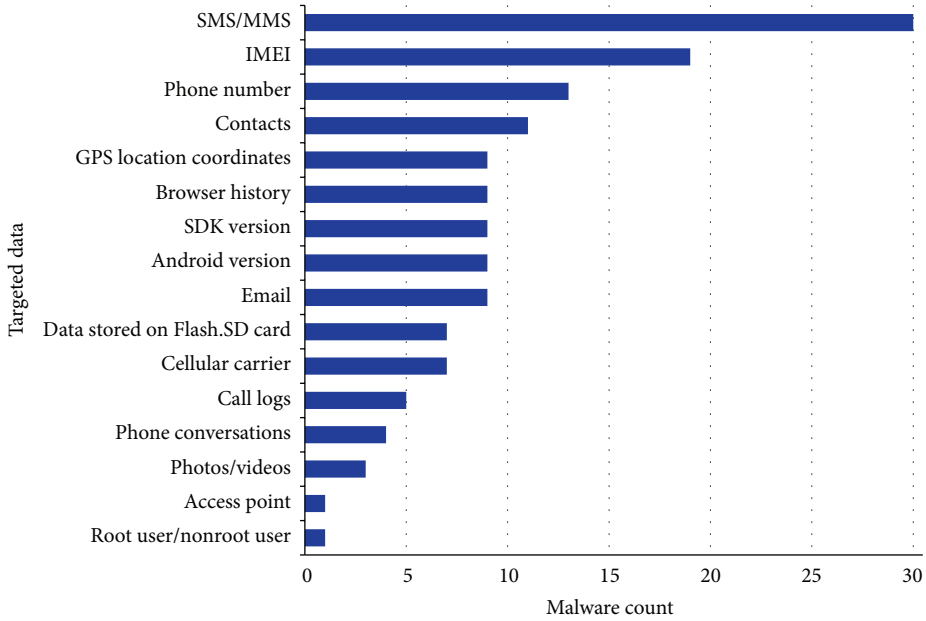


Figure 1: Private data targeted by Android malware.

to unnoticeable system events to start off their undercover services. To quantify this, the paper presents a methodology on how to measure user awareness on the data leaks. We produce this quantitative value in terms of the timing distance between a user event and data leaks activated by the event.

Second, we analyze a timing distance of real-world Android apps downloaded from the Google Play, based on the TaintDroid-ported emulator. The TaintDroid [2] traces internal data flow in order to detect a privacy leak. This instrumentation allows us to spot the leaked information and record a time stamp in milliseconds for each event. Combined with the Android's logging system, logcat [3], we can build a complete record table, featuring leaked information type, invoking interface, destination IP address, and time for the event. Our experiments show that the IMEI, accounting for 50 percent of privacy leaks, is frequently transmitted during the device identification process. In addition, the portion of IMEI leak, occurring within 30 seconds after the user's first input like a button, stands at 65.5%. The remaining cases take place during the automatic location update under a user's consciousness.

Third, to expand the concept of user awareness on the data leak, we build a cloud-based analysis system for detecting mobile malware [4, 5]. Our proposed system involves an amalgamation of both a dynamic application analysis and a network traffic analysis while the application is running. The finding that some leaked data were captured in a plain form while data are being transmitted calls for the need to secure the sensitive data with end-to-end encryption mechanism. The user awareness analysis and network-traffic monitoring can help to detect more sophisticated malware like botnet [6].

The rest of this paper is organized as follows: Section 2 gives the problem we are targeting and explanation of preliminary analysis result. Section 3 describes a design of the proposed measurement methodology and architecture. In Section 4 we present experiment results. In Section 5, we explain a cloud-based dynamic analysis as a case study. And in Section 6, we discuss challenging issues and problems. We present previous work in Section 7 and, lastly, the conclusion is shared in Section 8.

2. Background

Limited resources available to software tools deployed on mobile platforms indicate the value of a cloud-based solution. A cloud computing provides multiple instances of mobile platform with more powerful CPU and less memory constraints. In addition, the analysis and processing of potentially malicious code occur in a virtualized phone outside of the real mobile device [4, 7–9]. Aside from the limited resources, the vulnerabilities [10] of mobile devices and the hacking capabilities from the malware would make the problem more complex than enterprise computing systems.

In an effort to understand categories and behavioral characteristics of malware in obtaining the private data targeted by them, we performed a static analysis on 55 collected Android malware samples from different families by reverse engineering [11] and analyzing the codes. The analysis result of Figure 1 suggests that sensitive data including Short Message Service (hereinafter SMS) and contacts are mainly targeted by Android malware. Most malware usually has excessive access permissions to SMS and IP networking. Another aspect to note is that many malware samples imitate popular applications to deceive users into installing the malware apps without any skepticism. So we compare the permissions requested by the original application with the imitations. The chart (Figure 2) of sample applications shows three different types of characteristics. First, being the obvious one, is where

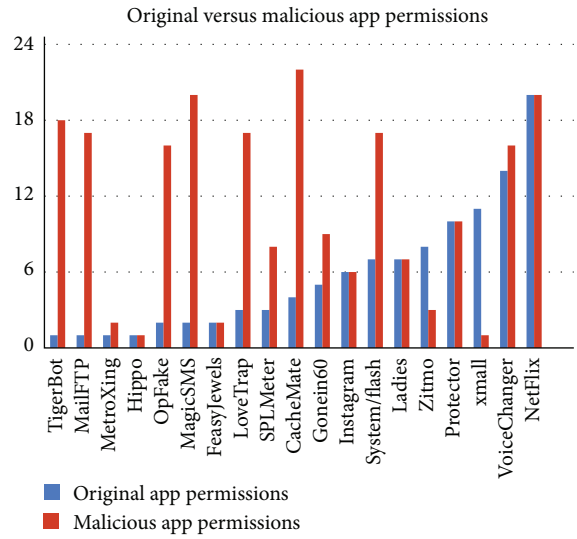


FIGURE 2: Permission analysis of original app versus imitation app (fake and malicious).

the malicious application requests permissions above and beyond the ones requested by the original application. For the second type, distinguishing malicious applications from the original applications based on the same number of permissions is challenging. One of the reasons is that permissions are broad in nature and encompass many features. For example, the permission called "INTERNET" allows the application complete network access. Finally the most interesting type of malicious application requires fewer number of permissions in comparison to the original application. This is the case where malware simply looks like the original application but has completely different code.

It is very challenging to draw a clear line between the malicious activity and a normal operation of mobile applications. In general, a large number of malware samples harvest the private data stored in the mobile device and send the data to remote servers with malicious intent. This form of information leak attack could have more impact on the users when financial credential is targeted. However, to make our goals more evident, this paper defines the malware as private data leaks without user's consciousness. A recent study [1, 12] shows that existing Android malware is mainly activated by BOOT, SMS, and NET related events. In particular, BOOT_COMPLETED event among system events is mostly targeted by the malware. Another interesting feature is that most malware is commanded through network and SMS messages from a remote Command and Control (hereinafter C&C) server. This trend shows a stealthy feature of malicious apps and makes them unnoticeable to the user. Also, this is a critical challenge for existing malware mitigation approaches [2, 9, 13]. Among the approaches, TaintDroid provides runtime analysis by tracking the sensitive data flow to identify the misuse of user's private information. In addition to Java-written portion of application, the DroidScope allows its internal analyzer to track the information flow of native library portion where malicious elements could reside [14].

However, automatic malware detection without human intervention tends to be erroneous due to lack of understanding of user intention. Many popular mobile applications demand user's geographic information to offer accurate locality-based services without any explicit input from the user. For example, Foursquare [15] helps people find places of interest nearby using their location. Therefore, it is really hard to tell whether the data exfiltration of the GPS location is malicious or not.

Therefore, a result of data transmission from a mobile device does not necessarily mean a data leak situation. Even benign activity could be considered as a data leakage from a simple decision based solely on information flow. As a result, a malware analysis model with rough threat definition could lead to a high false positive even in a situation when users purposely allow the internal information to leave their devices. In addition to the internal data flow itself, we should consider external contexts of applications (such as user intent) in making a decision on whether the leak activity is malicious or not.

Instead of incurring immediate information leak on the stored data, some malware aims to control the mobile device for future exfiltration. For instance, the Android.Stels [16] is one of infamous Trojans recently reported in Android platform. Once installed on the infected device, the Trojan opens a back door for handling requests from a C&C server. And later the Trojan initiates phone calls to premium numbers obtained through the C&C server. Making phone calls and sending SMS messages to premium numbers could result in a financial damage to the user. However, the Trojan never moves sensitive data out of the device or exfiltrates private user information to the remote server. Instead, this special malware activates the malicious operations without user consent. Developing a method to measure user awareness of the application's activity can assist in identifying the unintended sensitive API calls: sending SMS messages and making phone calls.

The architecture and design principles presented in this paper do not limit their applications to any specific mobile platform. The predominance of the Android platform in the mobile device market and continually expanding growth (approximately 78 percent of the mobile market in the first quarter of 2015 [17]) convince the authors to choose the Android platform as a reference model.

3. Design

3.1. Measurement Methodology. A recent user-friendly mobile application comprises multiple instances of view objects and layers. This type of application is running in an event-driven way, rather than a sequential execution flow from a program's entry point. For example, an Android app consists of several activities, representing independent execution element with corresponding user-interactive object. Each activity responds to user's inputs and displays results for visible information. For example, the Android app of Figure 3 consists of six independent elements, expressed as a set of $\{E_0, E_1, E_2, E_3, E_4, E_5\}$. Among them, the two elements, E_2 and E_5, have user-interactive operations in the blue

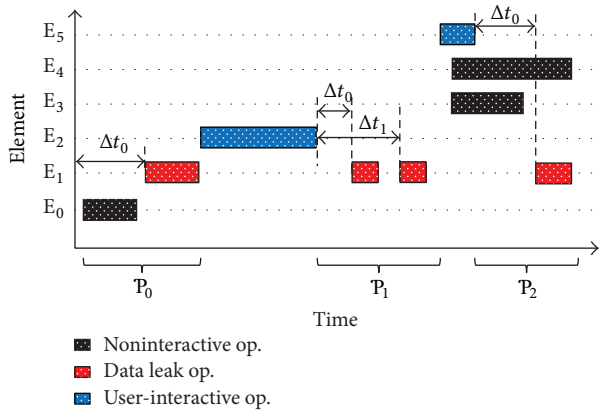


Figure 3: Measurement factors for data leak intention.

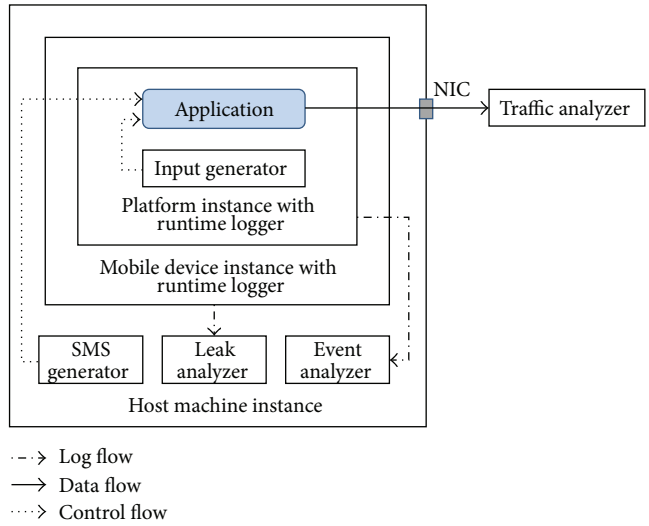


Figure 4: Core building blocks and operating platform.

rectangles, responding to the user inputs. The element E_1, running as a background service without user interaction, includes data leak operations at the four distinct time slices in the red rectangles. The distance variable (Δt) designates the elapsed time between the data leak operation and its preceding user-interactive operation. The period variable, noted as $\mathcal{P}$, specifies the time span during which all corresponding data leaks are assumed to be triggered by the identical user-interactive operation. The time distances Δt_0 and Δt_1 in period $\mathcal{P}_1$ are assumed to be triggered by the same user-interactive operation.

Generally, app can be launched by clicking the app's icon. So, we assume that there is an explicit user interaction involved. Then, by our definition, the first period $\mathcal{P}_0$ begins right after the app gets started. Among all execution elements in Figure 3, the elements E_0, E_3, and E_4 in the black rectangles are running without any explicit user interaction during the app's lifetime. They do not involve any data leak either, so we exclude them from the measurement procedure. Considering the malicious applications' behavior, malware manages to hide its internal activities from the user's consciousness. In other words, the device user hardly notices the activities and events which are happening behind the user-interactive operations. As a basic unit of measurement, a time distance (Δt) of an event represents a quantitative value of a user awareness on the event. By combining all timing distances, we aim to measure a degree of intention with which the application tries to hide its internal activities from the user consciousness. As the application hides more, the distance values would be higher.

3.2. Overall Architecture. To conduct our measurement on an application in question, the runtime information inside the application should be monitored while it is running. Hence, the crucial function in our approach is to track and record all activities that take place inside the application and let the analyzer examine the gathered information afterwards. With three major design principles, we propose a platform architecture that collects the runtime information at two different checkpoints. First, tracking the data flow inside the application is an initial step to trace the flow of sensitive information. When the data under monitoring leaves the application through network or interapplication communication (IPC), this potential leak should be reported with supplemental information. To accomplish this task, application's operational information needs to be provided from the base platform on which the application is running. Second, in addition to the local data flow, external data flow such as network traffic should be taken into account. Our preliminary analysis showed that a good amount of malware communicates with an external server acting as a C&C server. Moreover, analyzing network traffic of the application allows getting a better understanding of tactics used by malware. Finally, the measurement architecture can be deployed in cloud computing environment, so the entire processes should be done automatically without human intervention.

The proposed architecture shown in Figure 4 integrates core building blocks and surrounding environment. The required functions from the design principles above are incorporated into core building blocks. And the operating platform is projected on the basis of cloud computing's virtual environment. Overall, the architecture contains three instance layers of virtual execution environment: host machine, mobile device, and software execution platform. To obtain operational information inside and outside of the application, two runtime loggers are positioned in the device and software platform instance, respectively. If we take Android as a reference model, a mobile device along with the Android software platform can be emulated on a host machine with x86 CPU. Even further, the host machine can be virtualized into multiple instances in the cloud. In order to examine an Android app, a virtualized host machine creates an instance of a mobile device with the Android platform installed. The runtime loggers instrumented in the Android platform and mobile device allows the analyzers to retrieve app's operational information. The app feels the same environment as in the real Android device so that running the app would be straightforward.

3.2.1. Analyzers. The system shown in Figure 4 has three main building blocks for analyzing application's runtime behavior while the application is running. First, the leak analyzer collects all logs of internal data traces and the outgoing data flow. Once logs are generated, the data leak analyzer inspects the leakage and then the event analyzer looks into the related event causing the data leak. Based on the gathered logs, the leak analyzer measures the time distances described in the previous section to gauge the degree of user consciousness of the data leak. For every seemingly data leak, the leak analyzer needs to locate unconscious data leak using the information provided by the event analyzer. Aside from direct data flow of the application, supplementary information such as system resources and configuration is helpful to understand application's context. Therefore, the runtime loggers located in the base platforms and mobile device are responsible for providing information to the two analyzers. Two loggers located in the instance layers of platform and mobile device generate logs which include system events, emulated user events, and their time information. In this paper, we use the logcat as a ground logging system with which the Google's Android Software Development Kit (SDK) [3] comes. Along with these internal analyzers, there is an external analyzer, traffic analyzer. When an application is transmitting data outside, the outgoing packets would be generated. Then the traffic analyzer can examine the data flow passing through the network interface card (NIC).

As for the data leak monitoring, many outstanding approaches [2, 5, 18–20] have been suggested using dynamic analysis. Generally, the dynamic analysis examines the behavior of an application in execution within a controlled environment. Among those available in public domain, TaintDroid was adopted as a runtime logger to trace data flow and identify any data leak. The TaintDroid, a type of dynamic taint tracking technique, traces local data and variables through the instrumented Dalvik virtual machine. This Android-based approach aims to identify the sensitive data leakage in real phone.

3.2.2. Generators. Interactive mobile applications are usually triggered by user inputs or events from system services. To deploy our architecture in the cloud, we build core blocks on the virtual instances. For the same reason, one of most important issues in automating the analysis process is to emulate user interaction without actual human involvement. The input generator of Figure 4 controls the application under examination by emulating user input events to trigger all activities inside. In addition to the input generator, the SMS generator makes the application believe that it runs on a real phone by emulating SMS messages. When testing an application without human intervention, the performance of the input generator determines the accuracy of the entire dynamic analysis system.

Our preliminary static analysis in the previous section showed that most malware has excessive permissions on the SMS and IP networking access. Similarly, the previous studies [1, 12] suggest that a good number of malware samples utilize SMS messages to activate themselves. Therefore, emulating SMS is critical for activating malicious applications as well as normal applications. In practice, the SMS generator is tightly coupled with a mobile device emulator. The generator used in our experiment utilizes the Google's Android Debug Bridge (hereinafter ADB). The ADB provides a command to simulate an incoming message with a bogus phone number.

The proposed input generator scans all objects on the view layout [21] of the application and traverses each view object one by one using emulated valid inputs. This process can be achieved in a depth-first search data structure which keeps track of the activities in a stack and visits them thoroughly, thereby traversing through all the views of the application. Consider an application with three different visual activities shown in Figure 5. The proposed algorithm works by generating appropriate input to its text field and an event for clicking the button. And then clicking the button gives rise to new activity, Activity 2 for button 1. If the next activity gets started, then the same procedure applies recursively to the activity to explore further. The search flow path of our depth-first search described so far is represented as a tree in solid line.

Once the recursive search routine reaches the right-most leaf node, we can get the search flow back to the previous activity via an explicit or implicit back button. Before getting started at every activity's entry, a loop-check routine is in place to see if there is a cycle in the already traversed nodes. To avoid a loop situation, the traversed activity nodes are maintained in the list and we compare them with the current node before getting started.

With the help of these two generators, any application can be tested without any knowledge about the application's source code. However, recent mobile applications are implemented with more sophisticated view objects so that the input generator needs to understand the object's detailed information to generate valid input events. This often results in considerable overhead when writing robust, automatic black box test cases. Moreover, as privacy sensitive apps tend to require user authentication, generating valid text for one-time passcode remains as a challenging problem to emulate correct user inputs.

4. Experiments

In this section, experimental data sets and results are presented. With representative real-world Android apps, user's awareness of data leaks is measured based on our methodology introduced in Section 3. For the underlying virtual instance layers of operating platform, VirtualBox [22], QEMU [23], and Dalvik virtual machine are used. The QEMU-based device emulator and the TaintDroid-ported Dalvik virtual machine have equipped the runtime logger, providing application's operational information. The host machine for our experiments is based on 64-bit Linux with 3.5 GHz 8 cores and 32 GB memory. The mobile software platform is Android 4.3 Jelly Bean which is the latest applicable version for the TaintDroid integration. There are several ways to get Android apps for experiment but downloading from Google Play, known as a formal Android application market place, is selected. There are several criteria for selecting apps to analyze the experiment. First, apps need to be installed and

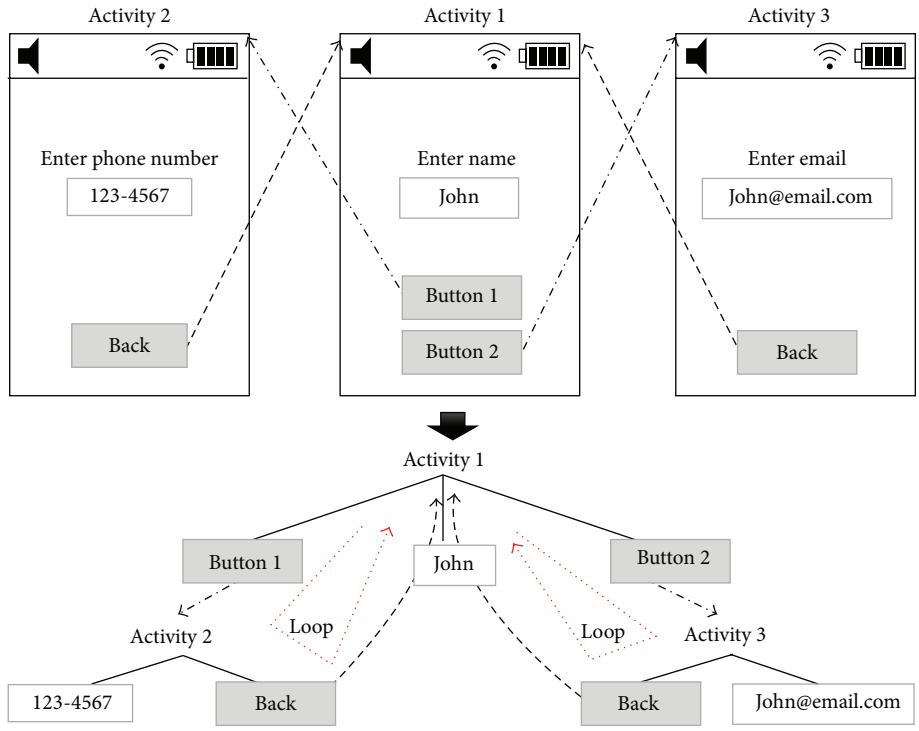


Figure 5: Depth-first search tree in user inputs emulation.

run successfully on our emulated platform. Some apps check whether a Subscriber Identification Module (SIM) is inserted and ready for use. The emulator-based software platform fails to install those apps due to the difficulty in emulating the physical SIM card. Second, to examine the relationship between data leak and user awareness, the app under test should transmit sensitive data to an external server or a group of servers. This requires the app to gain the permissions of full network access and location. Before downloading the candidate app, a full list of permissions should be checked to be granted and make sure that the app is likely causing data leaks. After the candidate meets this screening requirements, then a measurement sample is selected. Third, data set needs to represent the same type of services. We mainly focus on the personal service apps which utilize sensitive data. The coverage of the finally selected apps ranges from health, weather, and traffic to chatting messenger services.

Before starting our measurement experiment, an emulator is set up with bogus private data such as IMEI, GPS location coordinate, and contacts. This will let the app under test run as normally as in a real phone. The Google's ADB tool provides a set of commands for developers to configure the phone's system information. For example, the emulated device's geographic location can be set by the geolocation command. The power status of the phone can be manipulated to trigger the phone into a power save mode. Other than editable system information at the ADB tool, the IMEI number can be directly inserted into the platform's source code. And the contacts and photos are manually filled in and saved for every experiment. Other delicate issues related with the emulator testing environment will be addressed later in the discussion section.

Once an Android app is downloaded from the Google Play and installed, our analysis process is ready to start off. With the help of the TaintDroid notification function, a data leak can be easily spotted. The pictures of Figure 6 present some screen captures of the notifications informed by the TaintDroid. When the TaintDroid icon shows up on the status bar, the icon can be pulled down and show the detailed information about the corresponding data leak. The information includes the name of app causing the data leak, destination IP address, leaked data type, time of this event, and data itself. Among them, the data type, marked as tainted, and the timestamp are measured. Eight sample apps were tested, representing different service families. The preliminary analysis result shows that the 8 samples are sending sensitive data to remote server(s) more than one time. In practice, location-aware apps try to locate the phone in order to provide neighboring services or location-dependent information.

Given this fact, it is expected that the Glimpse and Weather Forecast Pro, location-related services, transmit the location data to the server. Also the personal workout assistant app, RunDouble, transmits the phone's IMEI, location, and GPS location multiple times during the test. The personalized apps usually transmit devices' IMEI information as well as location data to the remote server. Likewise, sending IMEI out is an essential part to personalized or device-oriented services. As with these 3 samples, other 5 apps provide their unique information such as device's IMEI in exchange for device-identifiable services.

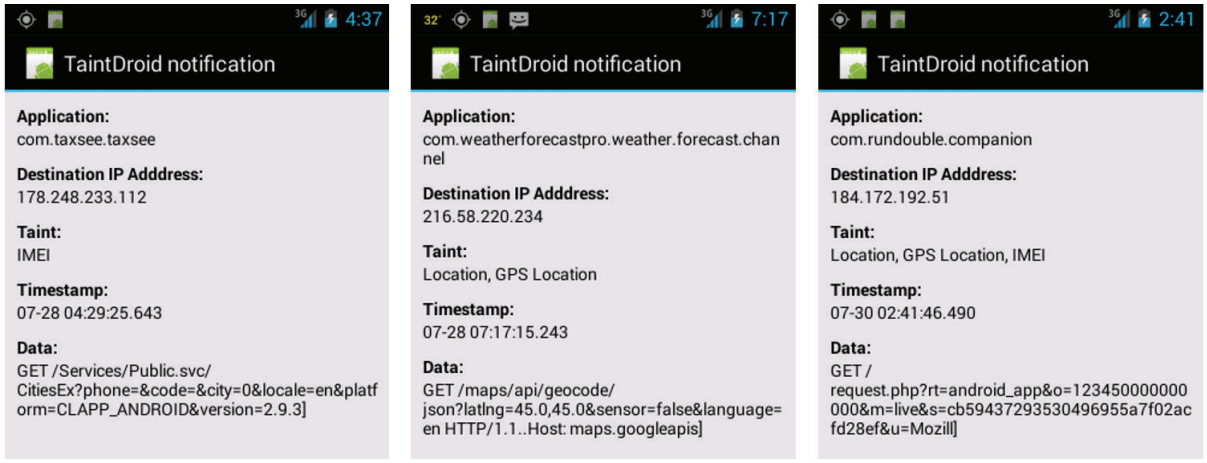


Figure 6: Data leak notification screenshots.

There are several findings to notice in our experiment result of Table 1. First, the case of IMEI information leaks accounts for about 50 percent of all the data leaks. It means that the counterpart service server requests the phone's ID for every service transaction. So, we expect more frequent IMEI transmissions as an app provides more various services through communications with its server. Interestingly, the traffic information app Bey2ollak, a cross platform mobile application, transmits device's IMEI 16 times to several servers as the app starts up. This can show an exemplary behavior of cross platform app, putting pieces together from crowdsource. Second, apps with periodic data updates show higher number of information transmissions. The RunDouble tracks the distance of the running course and allows the phone user to calculate the calories. This kind of app should read location data at a certain time interval while it is in active mode. This is also shown in the Glimpse app when the function of sharing location is activated. However, most privacy leaks occur at the early time of app's execution and discontinue without explicit inputs.

Among the leaked data types, the location and GPS location type seem similar but actually have a delicate difference in getting the location information. There could be several ways to find a location of the phone. Among them, the location means to get the location information through network instead of the GPS. In particular, when the user is inside the building or in downtown surrounded by high-rise buildings, known as GPS Dead Zone, the location should be resolved by network. In our experiment, the location and GPS location are both tainted whenever the location data leaks happen.

For the timing analysis of our experiment result, leaked information types which are IMEI and location were plotted. In the measurement methodology, each data leak has a corresponding input event triggering the leak. The x-axis represents a time distance from the event to the data leak, denoted as Δt in Section 3. Apps may have multiple triggering input events and corresponding data leaks and so there exists a set of periods $\{\mathcal{P}_0, \mathcal{P}_1, \ldots, \mathcal{P}_n\}$ in one app. But others like the Weather Forecast Pro and My Backup have only one input

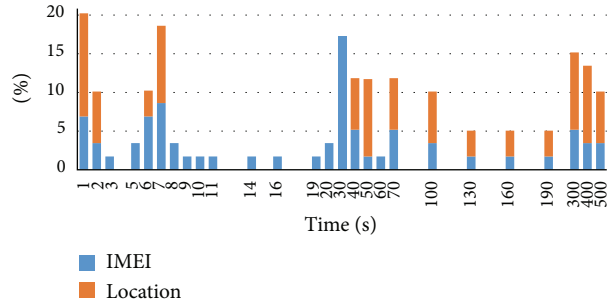


Figure 7: Time distance graph for all data leaks.

event, triggering actual data leaks. And then they have one period, $\mathcal{P}_0$, and a set of time distances, $\{\Delta t_0, \Delta t_1, \ldots, \Delta t_n\}$, in $\mathcal{P}_0$. The graph in Figure 7 depicts the aggregated time distances in the periods of all the apps tested.

As for the activation mechanism, clicking the app's icon is a common trigger for the data leak to happen. Therefore, many privacy leaks have been detected during the app's launch time. The next common triggering element is button object. In general, user interaction with mobile application works by clicking GUI view objects like button. Many Android apps utilize the button object to get a consent from its user. In this sense, monitoring the button object in the screen is crucial to identifying the triggering event correctly. The last device to activate the privacy leak is timeout events to get updated information from the server. As explained before, some mobile applications track location or get updated as time goes by. Such applications send out privacy data whenever they keep the local data synchronized with that of the server. And this update continues until the applications get deactivated. In all, our data set includes all privacy leaks activated by these three types of mechanism. According to the timing graph, we notice that a big group of data leaks comes within 11 seconds.

In particular, the data leak within 1 second stands at 21 percent. Considering these two facts, we assume that these apps respond to user events immediately by sending out

Table 1: Privacy leak and information type.

App name	App type	Transmitted information			API calls
		IMEI	Location	GPS location	
RunDouble	Health	●	●	●	57
Maxim	Taxi order	●			4
Trucaller	Call ID and block	●			8
Bey2ollak	Traffic	●			26
Glimpse	Location share		●	●	12
Weather Forcast Pro	Weather		●	●	6
My Bakcup	Backup	●			1
Hi	Communication	●			4
Total		**58**	**30**	**30**	**118**

necessary privacy data to the server. The next noticeable group comes around 30 seconds. Looking into this group further, most of the data leaks come mainly from the Glimpse, Bey2ollak, and RunDouble. Those apps have common characteristics, keeping the local information updated regularly. The remaining sporadic data leaks after the second group are also from those three apps. Therefore, two distinct data leak patterns, which are triggered by user's input events and recurrent timeout events, were discovered.

Getting a closer look at the first group, two separate subgroups were also discovered. The front subgroup within 3 seconds looks normal in that the data leaks in that group stem from the immediate responses from the user input events. But, the rear subgroup forming around 7 seconds needs more explanation. When an app get started or triggered into activation, it usually changes the screen for next activities. This screen change is a time-consuming job, compared to arithmetic operations or network communications, and leads to delays. Therefore, the seeming delayed date leak in the rear subgroup is attributed to the screen change and network connection latency. Overall, the time distance of data leak can depend on the application's programming design as well as activating event types.

Getting back to our design principal, we manage to correlate the privacy leak and user inputs events. That is considered to be a way to measure how much the user is aware of what the application is doing, especially leaking sensitive data. Our experiment shows that about 40% of the IMEI leaks and 33% of the location transmission happen within around 10 seconds right after user input events. All these data leaks respond to user willingness to get services by providing the personal data to external server. Meanwhile, the remaining 60% of the IMEI and 67% of the location leaks seem to take place far away from user's inputs. However, all these data leaks occur while the apps under test are running foreground with visual activities. In other words, the user assumes to be interacting with the apps without explicit user inputs. In this situation, the visual presentation shown on the screen is considered as a user's implicit intention. To be more quantitative argument, we need to devise a method to gauge the data leaks implicitly knowledgeable to the user. This will be a challenging topic and we leave it as future work.

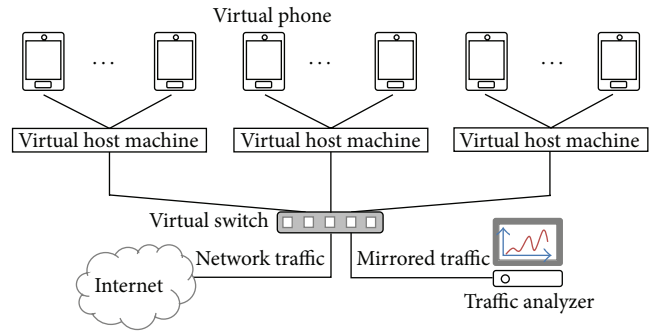


Figure 8: Cloud-based dynamic traffic analysis framework.

5. Case Study: Cloud-Based Analysis

To put the whole system into the cloud, the host machine as well as the mobile device under test should be virtualized. When an application is installed and activated by the input generator shown in Figure 4, the application will try to communicate with a remote server to send sensitive information. In that case, the traffic between the application running on the mobile device and the remote counterpart server can be analyzed with the traffic analyzer. For the network traffic analysis [24], the traffic analyzer can receive aggregated traffic from multiple instances of emulated mobile device. The overall traffic allows us to detect various traffic behavior which can only be shown in a group of mobile devices combined. For the traffic analyzer to exercise an extensive analysis over multiple devices, the traffic analyzer residing on a physical machine could be separated. With the isolated traffic analyzer, a scalable framework consisting of multiple virtual host machines is shown in Figure 8. A virtual host machine represents a host machine instance of Figure 4. The instance of host machine can connect to other virtual host machines, joining a huge virtual network which encompasses all virtual phones.

The overall framework of Figure 8 shows the traffic analyzer residing on the switch where all traffic from virtual phones merges. With this the analysis model, traffic coming from all virtual phones can be examined by a single traffic

analyzer. This architecture has the following advantages against the traffic analysis running on each host machine.

(i) *Analysis on the Logical Network*. The traffic analyzer inspects the aggregated traffic pattern as well as a single stream of traffic from a device. For example, virtual devices of the same affiliation can construct a logical network. This model allows us to apply the traffic analysis to a logical user group without regard to their actual physical location.

(ii) *Efficient Policy Management*. Single check point model at gateway is efficient to deploy consistent policy over all virtual devices. This obviates the need for replicating identical policy to each host machine whenever policy changes.

(iii) *Flexible Traffic Analyzer Implementation*. No design constraints on the traffic analyzer allow using a dedicated hardware with special accelerator(s) in analyzing packets. Or we can deploy software-based traffic analyzer on general-purpose host machine.

Through the network analysis in the experiment, personally sensitive data are transmitted through the network in a plain text form. Only 16 out of 182 data leaks have been secured by the SSL encryption. IMEI and location information might be less sensitive than personal messages and contacts. And sending data through mobile data communications is relatively secure due to the closed cellular network. However, they can have a harmful impact on the user when compromised with criminal purposes. Likewise, network-level analysis helps to understand the security vulnerabilities and come up with the measures mitigating them.

6. Discussion

In this paper, the privacy data leaks transmitted through IP network are traced. However, SMS and MMS message may also be used for a mobile app to send user's sensitive data to external server. The purpose of our analysis methodology is to measure user awareness of the privacy leak and discern useful apps against malicious apps leaking personal data by stealth. While the outgoing SMS and MMS without user's explicit consent are likely to be malicious and harmful, it is obvious to make a line between normal messages transmission and anomalous behavior with bad intention. Also, our methodology can facilitate detecting malware which targets premium-rate messages. Another applicable use case is lost phone services. These applications have built-in functions for the user to make a remote control over the phone. When the user lost the phone, the GPS location information is transmitted to the management server so that the user can locate the phone. In this case, the location-related privacy data leaks take place without user's input onto phone's view object. However, we make an assumption that the phone users are already conscious about the data transmission from the lost phone. For this reason, we exclude this case from the experimental data set.

When planning our experiment in the beginning, certain types of apps were expected to cause privacy leak. Particularly, map location services are expected to yield personal data leaks frequently. However, they have rarely produced privacy leak during the test. Downloading map information before starting the navigation services explains this situation. Map and the geographic information obtained by GPS module can provide the finding-location services without incurring any data leak. As a side effect, our experiment result helps to understand mobile application's internal operation.

To test real-world applications, the emulator should be set up with experimental configuration for an application under test to feel like a real phone. Even though we put some bogus personal data into the emulator, there are still more issues to be addressed. For example, whenever Android SDK makes an instance of mobile device emulator, the MAC and IP address of the instance are allocated all the same. Malicious apps use this information to circumvent the emulated system featuring our measurement module. Fortunately, normal apps do not check their emulator information to see if they are running on the emulator or real phone.

More apps tend to require robust user authentication or device authentication instead of simply requesting email address or IMEI number. As a common user authentication method, the sever checks if the email address given by the user is effective or not by confirming reply email from the user. For device authentication, pass code validation is used in several apps. This strict authentication process is the most challenging issue for our automatic analysis approach.

7. Related Work

With the increase in sales of smartphones, there also has been a steep rise in the number of malicious applications entering the online market. Given the enormous growth of the malware, security researchers and vendors must analyze more and more applications in a given period of time to understand the purpose of the software and to develop countermeasures. Until recently, analysis is done by using tools like decompilers and runtime debuggers. This process can be very time-consuming and error-prone depending on the skill set of the analyst. On the other hand, an automatic analysis [25] investigates the downloaded application without human intervention. Main technique that the automatic analysis applies is binary forensics including decompilation, decryption, pattern matching, static system call analysis, and control flow analysis. However, malware developers also put their efforts in finding new ways for the malware to circumvent the detection mechanism [1, 26].

As one of best well-known analysis approaches, TaintDroid [2] tracks runtime data flow through variables and intercomponent communication to detect privacy data leaks. However, this fine-grained tracking technique only covers the Dalvik's Java byte codes. So, it bears limitation in identifying the data leaks occurring in the native library written in C and C++. The Mobile-Sandbox [19] integrates the TaintDroid into a virtualized phone and extends the coverage of code into native libraries written in programming languages other than Java. However, the limitation of the taint tracking lies in the lack of understating of surrounding circumstances including the user's intent. To improve the runtime taint tracking,

the VetDroid [27] analyzes permission use behavior to detect malicious information leaks. As an integrated framework model, AppsPlayground [28] combines several analysis approaches such as dynamic taint tracking, API monitoring, and kernel level monitoring to supplement individual approach. Pegasus [29] uses formal method in specifying permissions and APIs property. By enforcing a permission event graph constructed from model checking, the formal method detects malicious behavior. However, the permission misuse and abnormal API call patterns only focus on the application's internal operations, not considering the user's interaction with the application. Even when a data leak is detected, determining whether the leak is maliciousness or not requires additional information.

To make all tests happen without human intervention, the input generator should control the application under examination by generating events which trigger all activities inside. Therefore, the accuracy of the input generator determines the overall performance of an automatic analysis system. The widely used Monkey tool [30] with which Android SDK comes simulates user inputs without the knowledge about the application's source code. However, this tool generates random events automatically regardless of the application's actual view layouts. Therefore, incorrect inputs sometimes lead to a crash due to input format mismatch or invalid values. Similarly, the MonkeyRunner [31] is an automatic input generator with more valid input creation. However, this tool is made to test a device at the functional or framework level and so it tends to be more application-specific in nature. SmartDroid [32] manages to find effective user inputs that trigger sensitive behavior. This approach uses both static analysis [33] constructed from function call graph and dynamic analysis exploring the UI elements to reach the sensitive APIs. Like our approach, AppIntent [20] manages to discern user intended data leak from unintended one by providing efficient sequence of GUI interactions to result in privacy data leaks. The event-space constraints model of AppIntent reduces the search space with the similar code coverage to other approaches. Unlike the above tools, Brahmastra [34] and the TriggerMetric of [35] use static analysis to construct execution paths to invoke sensitive APIs. Rather than focusing on GUI elements, Brahmastra rewrites the application to trigger the callback functions that reach privacy-sensitive APIs. Instead of instrumentation profiling, A^3E [36] improves coverage by using a static, taint-style, dataflow analysis on the bytecode.

In addition to the data leakage analysis, traffic analysis for mobile platform has been proposed in some literatures. A hacked cellular station provides a chance of identifying traffic with the malware signature [16]. That is a fundamental concept of wired network-based intrusion detection system (NIDS). Routing all the traffic from the mobile device to VPN server allows monitoring the packets the same way as the NIDS [37]. By emulating the virtualized network environment, we can apply a network-level detection to screen all traffic coming from or to the mobile device. In addition, the virtual network constructed from individual devices can provide an overall traffic behavior from the logical network perspective. This will let the traffic analyzer identify the potential threats from the network perspective as well as a single device perspective. Similar to the dynamic analysis framework proposed in the paper, Andlantis [5] provides a good scalability in a clustered environment, being capable of processing 3000 Android applications per hour. But Andlantis uses MonkeyRunner to generate user input events.

8. Conclusion

Given the pervasiveness of mobile device in modern life, proactive prevention measures against malicious applications should be put in place along with existing security solutions. In this paper, we have presented a methodology and an architecture for measuring user awareness of sensitive data leakage, which features runtime application analysis over timing distance between the user input event and actual privacy data leak. Mobile apps may request privacy data in exchange for useful services. And this seemingly voluntary data leak leads to difficulty making a clear line over whether the intention of the data leaks is malicious or not. From our experiment on real-world Android apps, we discover that the IMEI and location information are used for device identification and location-aware services. For normal apps, most data leaks stem from user's direct input events or implicit interaction with visual presentation on the screen. Moreover, the proposed methodology helps understand the mobile application's internal operations. Combined with existing malware detection systems, we expected the user awareness measurement can assist in reducing the false positive in a delicate situation by measuring the app's malicious intent.

To overcome the limited resource and computing power of mobile device, cloud computing is a great platform upon which we can build a solution free of resource constraints. Another main contribution of the paper is to employ the network-based monitoring in mobile traffic analysis. The virtual network constructed from individual phone emulators can provide a more complete network landscape the same as in physical network. From the network-perspective analysis in our experiment, we observed the vulnerable practices of transmitting the IMEI and location information in a plain text form.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This work was supported by Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (no. R-20150518-001267, Development of Operating System Security Core Technology for the Smart Lightweight IoT Devices).

References

[1] Y. Zhou and X. Jiang, "Dissecting Android malware: characterization and evolution," in *Proceedings of the 33rd IEEE Symposium on Security and Privacy*, pp. 95–109, San Francisco, Calif, USA, May 2012.

[2] W. Enck, P. Gilbert, S. Han et al., "TaintDroid: an information flow tracking system for real-time privacy monitoring on smartphones," *ACM Transactions on Computer Systems*, vol. 32, no. 2, article 5, 2014.

[3] Android SDK, http://developer.android.com/sdk/index.html.

[4] J. Oberheide, K. Veeraraghavan, E. Cooke, J. Flinn, and F. Jahanian, "Virtualized in-cloud security services for mobile devices," in *Proceedings of the 1st Workshop on Virtualization in Mobile Computing*, pp. 31–35, ACM, Breckenridge, Colo, USA, June 2008.

[5] M. Bierma, E. Gustafson, J. Erickson, D. Fritz, and Y. Choe, "Andlantis: large-scale Android dynamic analysis," in *Proceedings of the 3rd Workshop on Mobile Security Technologies (MoST '14)*, San Jose, Calif, USA, May 2014.

[6] C. Xiang, F. Binxing, Y. Lihua, L. Xiaoyi, and Z. Tianning, "Andbot: towards advanced mobile botnets," in *Proceedings of the 4th USENIX Conference on Large-scale Exploits and Emergent Threats*, p. 11, Boston, Mass, USA, March 2011.

[7] G. Portokalidis, P. Homburg, K. Anagnostakis, and H. Bos, "Paranoid Android: versatile protection for smartphones," in *Proceedings of the Annual Computer Security Applications Conference (ACSAC '10)*, pp. 347–356, Austin, Tex, USA, December 2010.

[8] C. Jarabek, D. Barrera, and J. Aycock, "ThinAV: truly lightweight mobile cloud-based anti-malware," in *Proceedings of the 28th Annual Computer Security Applications Conference (ACSAC '12)*, pp. 209–218, ACM, Los Angeles, Calif, USA, December 2012.

[9] P. Gilbert, B.-G. Chun, L. P. Cox, and J. Jung, "Vision: automated security validation of mobile apps at app markets," in *Proceedings of the 2nd International Workshop on Mobile Cloud Computing and Services (MCS '11)*, pp. 21–26, ACM, Bethesda, Md, USA, June-July 2011.

[10] A. Nazar, M. M. Seeger, and H. Baier, "Rooting Android—extending the ADB by an auto-connecting WiFi-accessible service," in *Information Security Technology for Applications*, P. Laud, Ed., vol. 7161 of *Lecture Notes in Computer Science*, pp. 189–204, Springer, Berlin, Germany, 2012.

[11] Androguard, http://code.google.com/p/androguard.

[12] W. Enck, P. Traynor, P. McDaniel, and T. La Porta, "Exploiting open functionality in SMS-capable cellular networks," in *Proceedings of the 12th ACM Conference on Computer and Communications Security (CCS '05)*, pp. 393–404, November 2005.

[13] T. Bläsing, L. Batyuk, A.-D. Schmidt, S. A. Camtepe, and S. Albayrak, "An Android Application Sandbox System for suspicious software detection," in *Proceedings of the 5th International Conference on Malicious and Unwanted Software (MALWARE '10)*, pp. 55–62, IEEE, Lorraine, France, October 2010.

[14] L. K. Yan and H. Yin, "DroidScope: seamlessly reconstructing the OS and Dalvik semantic views for dynamic Android malware analysis," in *Proceedings of the 21st USENIX Conference on Security Symposium*, p. 29, Bellevue, Wash, USA, August 2012.

[15] Foursquare, https://foursquare.com/.

[16] S. Davido, D. Harrison, R. Price, and S. Fretheim, "Do-it-yourself cellular intrusion detection system," LMG Security Whitepaper, 2013.

[17] IDC: Smartphone Market Share, http://www.idc.com/prodserv/smartphone-os-market-share.jsp.

[18] I. Burguera, U. Zurutuza, and S. Nadjm-Tehrani, "Crowdroid: behavior-based malware detection system for Android," in *Proceedings of the 1st ACM Workshop on Security and Privacy in Smartphones and Mobile Devices (SPSM '11)*, pp. 15–26, Chicago, Ill, USA, October 2011.

[19] M. Spreitzenbarth, T. Schreck, F. Echtler, D. Arp, and J. Hoffmann, "Mobile-Sandbox: combining static and dynamic analysis with machine-learning techniques," *International Journal of Information Security*, vol. 14, no. 2, pp. 141–153, 2015.

[20] Z. Yang, M. Yang, Y. Zhang, G. Gu, P. Ning, and X. S. Wang, "AppIntent: analyzing sensitive data transmission in Android for privacy leakage detection," in *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS '13)*, pp. 1043–1054, Berlin, Germany, November 2013.

[21] Robotium, https://code.google.com/p/robotium.

[22] Oracle VirtualBox, https://www.virtualbox.org.

[23] QEMU, http://www.qemu.org.

[24] B.-H. Chang and C. Y. Jeong, "An efficient network attack visualization using security quad and cube," *ETRI Journal*, vol. 33, no. 5, pp. 770–779, 2011.

[25] M. Egele, T. Scholte, E. Kirda, and C. Kruegel, "A survey on automated dynamic malware-analysis techniques and tools," *ACM Computing Surveys*, vol. 44, no. 2, article 6, 2012.

[26] R. Roemer, E. Buchanan, H. Shacham, and S. Savage, "Return-oriented programming: systems, languages, and applications," *ACM Transactions on Information and System Security*, vol. 15, no. 1, article 2, 2012.

[27] Y. Zhang, M. Yang, B. Xu et al., "Vetting undesirable behaviors in Android apps with permission use analysis," in *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS '13)*, pp. 611–622, ACM, Berlin, Germany, November 2013.

[28] V. Rastogi, Y. Chen, and W. Enck, "AppsPlayground: automatic security analysis of smartphone applications," in *Proceedings of the 3rd ACM Conference on Data and Application Security and Privacy (CODASPY '13)*, pp. 209–220, New Orleans, La, USA, February 2013.

[29] K. Chen, H. Johnson, V. D'Silva et al., "Contextual policy enforcement in android Applications with permission event graphs," in *Proceedings of the 20th Annual Network and Distributed System Security Symposium, (NDSS '13)*, San Diego, Calif, USA, February 2013.

[30] Monkey tool, http://developer.android.com/tools/help/monkey.html.

[31] Monkeyrunner, http://developer.android.com/tools/help/monkeyrunner_concepts.html.

[32] C. Zheng, S. Zhu, S. Dai et al., "Smartdroid: an automatic system for revealing UI-based trigger conditions in Android applications," in *Proceedings of the 2nd ACM Workshop on Security and Privacy in Smartphones and Mobile Devices (SPSM '12)*, pp. 93–104, ACM, October 2012.

[33] W. Enck, D. Octeau, P. McDaniel, and S. Chaudhuri, "A study of Android application security," in *Proceedings of the 20th USENIX Conference on Security (SEC '11)*, p. 21, San Francisco, Calif, USA, August 2011.

[34] R. Bhoraskar, S. Han, J. Jeon et al., "Brahmastra: driving apps to test the security of third-party components," in *Proceedings of the 23rd USENIX Conference on Security Symposium*, pp. 1021–1036, San Diego, Calif, USA, August 2014.

[35] K. O. Elish, X. Shu, D. Yao, B. G. Ryder, and X. Jiang, "Profiling user-trigger dependence for Android malware detection," *Computers & Security*, vol. 49, pp. 255–273, 2015.

[36] T. Azim and I. Neamtiu, "Targeted and depth-first exploration for systematic testing of Android apps," in *Proceedings of the ACM SIGPLAN Conference on Object Oriented Programming Systems Languages & Applications*, pp. 641–660, Indianapolis, Ind, USA, October 2013.

[37] A. Parrizas and D. Adrianto, *Monitoring Network Traffic for Android Devices*, SANS Institute InfoSec Reading Room, 2013.

Hindawi Publishing Corporation
Mobile Information Systems
Volume 2015, Article ID 971538, 11 pages
http://dx.doi.org/10.1155/2015/971538

Research Article

Power Saving Scheduling Scheme for Internet of Things over LTE/LTE-Advanced Networks

Yen-Wei Kuo and Li-Der Chou

Department of Computer Science and Information Engineering, National Central University, No. 300, Jhongda Road, Jhongli, Taoyuan County 32001, Taiwan

Correspondence should be addressed to Li-Der Chou; cld@csie.ncu.edu.tw

Received 1 October 2015; Accepted 1 December 2015

Academic Editor: Jong-Hyouk Lee

The devices of Internet of Things (IoT) will grow rapidly in the near future, and the power consumption and radio spectrum management will become the most critical issues in the IoT networks. Long Term Evolution (LTE) technology will become a promising technology used in IoT networks due to its flat architecture, all-IP network, and greater spectrum efficiency. The 3rd Generation Partnership Project (3GPP) specified the Discontinuous Reception (DRX) to reduce device's power consumption. However, the DRX may pose unexpected communication delay due to missing Physical Downlink Control Channel (PDCCH) information in sleep mode. Recent studies mainly focus on optimizing DRX parameters to manage the tradeoff between the energy consumption and communication latency. In this paper, we proposed a fuzzy-based power saving scheduling scheme for IoT over the LTE/LTE-Advanced networks to deal with the issues of the radio resource management and power consumption from the scheduling and resource allocation perspective. The proposed scheme considers not only individual IoT device's real-time requirement but also the overall network performance. The simulation results show that our proposed scheme can meet the requirements of the DRX cycle and scheduling latency and can save about half of energy consumption for IoT devices compared to conventional approaches.

1. Introduction

The concept of Internet of Things (IoT) enables devices to connect to the Internet to sense data and interact with each other. Different type of IoT devices, such as fitness, entertainment, location and tracking, and surveillance, may have different real-time requirement for different purpose. It can be predicted that IoT devices also referred to as IoT User Equipment (UE) pieces will grow rapidly and the power consumption and radio spectrum management will become the critical issues in the IoT networks. Moreover, the wireless connectivity is a promising way to the Internet for IoT devices. The Long Term Evolution (LTE) that, nowadays, is the dominant technology in wireless communication makes it ideal for IoT applications due to its flat architecture, all-IP network, and greater spectrum efficiency than 2G or 3G [1].

In the LTE network, the 3rd Generation Partnership Project (3GPP) specified the Discontinuous Reception/Transmission (DRX/DTX) mechanism to alleviate the power consumption issue which can severely affect the battery lifetime of the IoT UEs. The DRX mechanism allows UEs to stop monitoring the Physical Downlink Control Channel (PDCCH) and to enter low-power consumption mode to extend its battery lifetime. Without DRX mechanism, the LTE UEs have to continuously monitor the PDCCH in every subframe in order to check UE-specific scheduling assignments such as Downlink (DL) resource allocation, Uplink (UL) grants, and PRACH (Physical Random Access Channel) responses. Not only the UEs but also the Evolved Node B (eNodeB) can also benefit from the DRX mechanism to reduce its power and signal consumptions, such as the Channel State Information (CSI) or Sounding Reference Signal (SRS), and to improve its resource utilization. However, the main issue of the DRX mechanism is that it may pose unexpected delay if the UEs are in the sleep mode while the scheduling information arrived. Particularly, it should be noted that the PDCCH carries the scheduling information rather than service data and the scheduling information is

interleaved in the PDCCH in order to reduce the collision of blind decoding of the UEs. This implies that extending the on duration time of the DRX mechanism may not help to overcome this issue. Unfortunately, the specifications say nothing about the means of the DRX parameter configuration. Furthermore, different from the single DRX cycle in the Universal Mobile Telecommunications System (UMTS), the LTE network introduces two types of DRX cycles, the short and long DRX cycles, which complicates the DRX parameters configuration.

The DRX parameters must be configured to maximize power saving while the constraints of the communication latency and system throughput are satisfied. The DRX cycles allow UE to temporarily turn off its radio interface and enter a sleep mode for several Transmission Time Intervals (TTIs). On the one hand, too short of DRX cycle will lead to inefficient power saving. Too long DRX cycle, on the other hand, will result in long communication latency decreasing system throughput. Thus, the scheduler of the IoT networks must manage the tradeoff among the power consumption, communication latency, and the system throughput.

In this paper, we designed a fuzzy-based power saving scheduling scheme for IoT over LTE/LTE-Advanced networks to deal with the issues of the radio spectrum management and the power consumption as well as the unexpected delay from UE's scheduling and resource allocation perspective instead of complicated DRX parameters adjustment. Not only individual IoT UE's real-time requirement but also the overall network performance is taken into account. The main issue of the DRX parameters adjustment is that it will cause the inter-parameter interaction. Although the power saving can be achieved by adjusting the DRX parameters dynamically, higher signaling overhead is also introduced [2]. Since the UE's DRX parameters are semistatically configured by the eNodeB, the parameters can only be adjusted in a signaling procedure [3]. In addition, it is more difficult to adjust the DRX parameters dynamically because the UEs automatically enter the sleep mode, and the eNodeB does not know about this change. Moreover, because of the uncertainty of the wireless network and service traffic, optimizing the DRX parameters is challenging [2]. In contrast, dealing with these issues from the perspective of the UE's scheduling and resource allocation is more suitable to be realized in practice. The main idea behind the proposed scheme is to schedule the UEs which in the period of the on duration to prevent the unexpected delay. To successfully receive the scheduling information, reduce the power consumption, and meet the real-time requirement of IoT applications, the proposed scheme considers the UE's DRX cycle and scheduling latency when allocating radio resources. Furthermore, in addition to satisfying individual UE's requirements, maintaining overall system throughput is necessary.

The major contributions of this paper are that this paper is the first paper taking these two important issues, the power consumption and radio spectrum management, into account for IoT over LTE/LTE-Advanced networks, and our proposed scheme that benefited from the characteristics of low complexity of fuzzy theory can be realized in practice. In particular, different from previous works, our design considers the key aspects of the LTE/LTE-Advanced networks. In addition, we proposed a systematic simulation model to investigate the DRX performance based on the concept of the LTE networks. The simulation results show that our proposed scheme can provide approximately decision to meet the requirements of the DRX cycle and scheduling latency and to save about half of energy consumption for IoT UEs compared to conventional approaches.

The rest of this paper is organized as follows. Section 2 provides an overview on related work. Section 3 introduces the discontinuous reception mechanism used in LTE/LTE-Advanced networks. Section 4 formulates the problem of maximum radio resource utilization. Section 5 describes the proposed fuzzy-based power saving scheduling scheme. Section 6 presents the simulation model. The simulation results and the discussions are presented in Section 7. Finally, Section 8 concludes this paper.

2. Related Work

The recent researches [2–9] are concentrated on modeling, investigating, or analyzing the DRX mechanism for 3GPP LTE/LTE-Advanced networks, and these studies show that the DRX mechanism has the ability to save UE's power consumption. However, the research [3] points out that many studies neglect the key aspects of LTE. In [4, 5, 7], the semi-Markov chain is used to model and analyze the wake-up latency and power saving factor of the DRX mechanism. A light sleeping mode is proposed in [10]. From hardware perspective, the light sleep mode allows UE to just turn off its power amplifier in order to reduce energy consumption while satisfying the delay requirement of Quality of Service (QoS).

In [2], the authors proposed a concept of burst-based scheduling scheme for the LTE networks to increase power saving efficiency while the desired QoS of UEs are satisfied. The main idea of this paper is to prevent UEs from entering the period of the opportunity for DRX with no data reception by adjusting scheduling priority using forward and backward strategies. However, the paper only considers fixed inactive timers, and the multi-DRX cycle scenario and the multiple access scheme of the LTE networks have not been taken into account.

The guaranteeing quality of service is used to optimize the DRX parameters in [1, 11]. In [1], the authors proposed two schemes, the three-stage scheme and the packet scheduling scheme, for the UEs and the eNodeB to cooperate with each other, respectively. The UE's power is saved by reducing its wake-up period. To accomplish this goal, these algorithms must determine the UE's DRX parameters, such as the short and long DRX cycles, on duration time, and the in active timer, in advance. However, there may be an issue of starvation in the packet scheduling scheme since this paper only considers higher channel rate and stringent data that are going to be swapped out from the buffer.

In [11], the authors mainly proposed two decision algorithms, the DRX parameters decision algorithm and scheduling-start offset decision algorithm. The former is used to determine the UE's on duration time, and the latter is

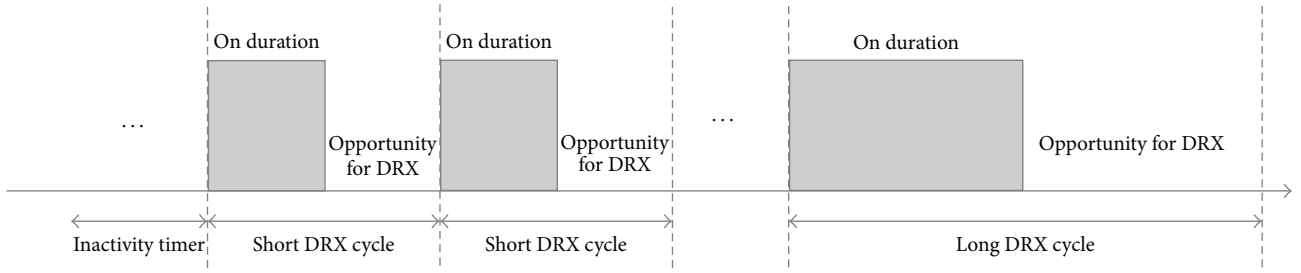


Figure 1: Operation of the DRX cycle in RRC_CONNECTED state.

used to disperse the UE's on duration time. The main idea of this paper is to prevent overlapping of UE's on duration time in order to fully utilize the network resource. However, the algorithms might be hard to be deployed in practice since the UE's DRX parameters are semistatically configured by the eNodeB rather than the UEs and the parameters can only be adjusted in a signaling procedure [3]. In addition, the authors also neglect the key aspects of the multiple access scheme of the LTE networks.

However, from our understanding, the following two key aspects of the LTE networks are usually neglected by recent research. (1) Extending the on duration time may not help to improve QoS of service traffic since the UEs check the PDCCH scheduling information rather than service data in the on duration period and the scheduling information is interleaved in the PDCCH in order to reduce the collision of blind decoding. (2) The UE's DRX parameters are semistatically configured by the eNodeB rather than the UEs; the parameters can only be adjusted in a signaling procedure [3]. In addition, most of them do not consider multicellular networks in their simulation works.

3. Discontinuous Reception in LTE

The DRX mechanism is known as a sleep mode in which UEs can stop monitoring the PDCCH and enter low-power consumption states to extend its battery lifetime. As mentioned in Section 1, the PDCCH carries the scheduling information, such as the UE-specific scheduling assignments for Downlink (DL) resource allocation, Uplink (UL) grants, and PRACH (Physical Random Access Channel) responses, rather than service data for the UEs, and the scheduling information is interleaved in the PDCCH in order to reduce the collision of the UE's blind decoding. Depending on UEs in active or idle mode, two mobile's Radio Resource Control (RRC) states, RRC_IDLE and RRC_CONNECTED, are used in the DRX mechanism. When a UE in RRC_IDLE state, the eNodeB will free up its resources for other users to maximize the UE's battery life and will know nothing about this UE. After that, the procedure of paging request will be typically used if the eNodeB wishes to contact this UE, and procedure will take more time to establish the connection between them. The eNodeB can negotiate the DRX parameters with UE through the RRC configuration procedure. In this paper, we concentrated on dealing with the DRX issues in the RRC_CONNECTED state since the state enabling data transmission and reception is important.

In the LTE networks, different from the single DRX cycle used in the UMTS, two types of DRX cycles, the short and long DRX cycles, are introduced. The long DRX cycle offers greater opportunity to achieve better power saving than the short DRX cycle. The short DRX cycle is optional. If both are configured by the eNobeB, the UE starts with 16 short DRX cycles, 2 to 640 subframes, and then enters a long DRX cycle, 10 to 2560 subframes, without receiving any scheduling information on the PDCCH.

The DRX parameters, typically, consist of three timers, the inactivity timer, on duration timer, and opportunity for DRX timer as shown in Figure 1. The UE should start the on duration timer in the subframe which satisfied (1) and (2) for the long DRX cycle and short DRX cycle, respectively [12–14]:

$$[(\text{SFN} \times 10) + \text{subframe number}] \cdot \bmod (\text{Short DRX cycle}) = (\text{DRX start offset}) \cdot \bmod (\text{Short DRX cycle}), \quad (1)$$

$$[(\text{SFN} \times 10) + \text{subframe number}] \cdot \bmod (\text{Long DRX cycle}) = (\text{DRX start offset}), \quad (2)$$

where the System Frame Number (SFN) which consists of 10 subframes will be reset to zero if it is greater than 1023 and the DRX start offset indicates the start of the on duration timer.

The three timers, the inactivity timer, on duration timer, and the opportunity for DRX timer, are contributed to the DRX mechanism. First, the UE stays awake for an inactivity time, 1 to 2560 subframes, when it received every PDCCH information. If the inactivity timer expires, the UE monitors the PDCCH for an on duration time, 1 to 200 subframes, and then it may go to sleep for an opportunity for DRX time, periodically. However, it should be noted that the DRX-to-connected latency is specified as less than 50 ms for the LTE networks and less than tightly 10 ms for the LTE-Advanced networks in the 3GPP specification [15]. This implies that the maximum DRX cycle must be 32 Transmission Time Interval (TTI) or 32 one-millisecond subframes for the LTE networks. In addition, the DRX cycles are expressed as power of 2 and all DRX cycles are always a multiple of the maximum DRX cycle in order to maintain the periodicity of DRX cycle. Therefore, there are six DRX cycle combinations and six DRX cycle

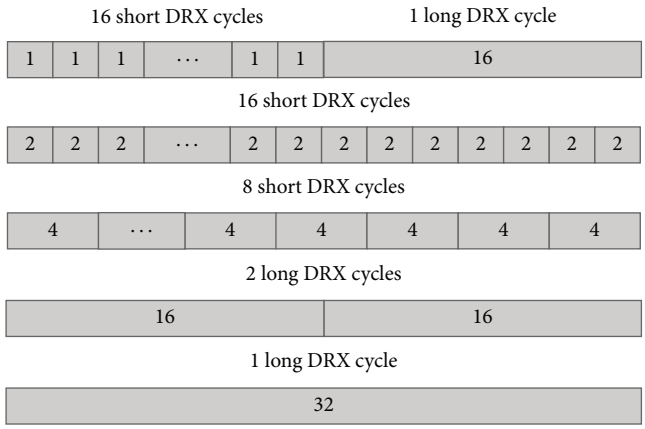


Figure 2: Six DRX cycle combinations.

granularities, 32, 16, 8, 4, and 2 subframes, and 1 subframe, can be used in LTE DRX mechanism as shown in Figure 2.

The selection of DRX combination is based on the Control Channel Elements (CCEs) aggregation level of PDCCH, which indicates the numbers of consecutive CCEs interleaved in the control region. The CCE aggregation level is set based on the UE's channel condition. The CCE carries the Downlink Control Information (DCI), which contains resource assignments and control information for each UE. For example, in the case of the UE with good channel condition, one CCE might be sufficient, but eight CCEs might be required in the case of the UE with poor channel condition. Therefore, the UE's reported Channel Quality Indicator (CQI) should be taken into account in the scheduling decision to ensure the overall system performance.

4. Problem Formulation

We consider the downlink transmission of the LTE networks in the proposed scheme. In this section, the problem of UE scheduling and radio resource allocation of the LTE networks is formulated as maximum radio resource utilization based on the fuzzy theory. We aim to find the UEs with higher scheduling priority and then to allocate radio resources for them. Given an eNodeB and a set of UEs UE $= \{\mathrm{UE}_k, 1 \le k \le N_{\mathrm{UE}}\}$ with its corresponding Channel Quality Indicator (CQI) C_k, the scheduling efficiency which involves the exact DRX time, CQI, and scheduling latency is given as

$$\mathrm{SE}_k = \frac{C_k}{E_k L_k}, \quad \forall k \in \mathrm{UE},\ 0 \le \forall C_k \le 15, \tag{3}$$

where E_k is the exact DRX timer of kth UE and C_k and L_k denote the CQI index and scheduling latency of UE_k, respectively. The CQI index 0 represents the out of range due to bad signal quality or DTX. Thus, the maximum radio resource utilization can be defined as

$$J = \max_{\alpha} \sum_{n}^{N_{\mathrm{RB}}} \sum_{k}^{N_{\mathrm{UE}}} \alpha_{k,n} \mathrm{SE}_k \tag{4}$$

$$\text{subject to} \quad \sum_{k}^{N_{\mathrm{UE}}} \alpha_{k,n} = 1, \quad \forall \alpha_{k,n} \in [0,1], \tag{5}$$

where N_{RB} denotes the number of Resource Blocks (RBs) in downlink transmission and N_{UE} represents the total number of UEs served by the eNodeB. The number of RBs depends on the system bandwidth of the eNodeB. The restriction of $\sum_{k}^{N_{\mathrm{UE}}} \alpha_{k,n} = 1$ entails that each resource block can only be allocated to a UE. Particularly, the $\alpha \in [0,1]$ is fuzzy set not crisp set in the proposed scheme in order to make appropriate decision. The resource block is the smallest entity that can be scheduled for the UEs in the frequency domain. The primary goal of (4) is to maximize the scheduling efficiency of the system bandwidth by adjusting the fuzzy set α. The appropriate fuzzy set α can be figured out by using three fuzzy membership functions and a low-complexity intersection function described in the next section.

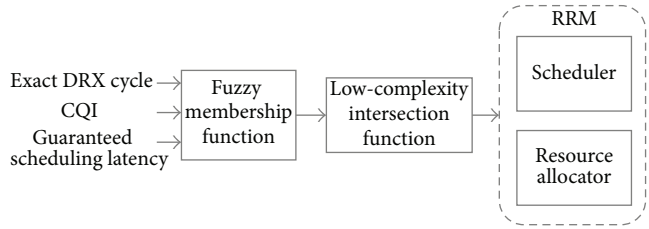


Figure 3: The structure of the proposed scheme.

5. Proposed Fuzzy-Based Power Saving Scheduling Scheme

In this section, we introduce a fuzzy-based power saving scheduling scheme to deal with the issues of the radio spectrum management and the power consumption as well as the unexpected delay for IoT over LTE networks. The proposed scheme considers not only individual IoT UE's real-time requirement but also overall system performance. We intend to allocate RBs to those UEs who have stringent real-time requirement, exact DRX cycle, and higher CQI. The tradeoff among them should be determined to achieve this goal. Fortunately, fuzzy theory provides means to make approximate decisions especially in multiobjective problems. The data even in different result spaces can be combined with each other through fuzzy set operations, such as union, intersection, and complement operation. An approximate decision is typically decided according to the highest membership values intersected by fuzzy operation.

A block diagram of designed fuzzy-based approach is shown in Figure 3. The proposed scheme mainly consists of two fuzzy functions, the fuzzy membership function and low-complexity fuzzy intersection function. The IoT UE's

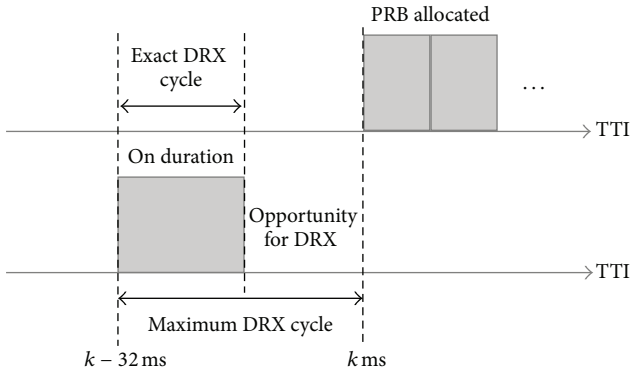


Figure 4: The idea of the exact DRX cycle.

CQI, exact DRX cycle, and guaranteed scheduling latency are considered as the scheduling metrics. The approximate scheduling decision is made according to the result of the low-complexity intersection function and performed by the scheduler and resource allocator of the Radio Resource Management (RRM) in the LTE/LTE-Advanced networks. In the LTE networks, the RRM comprises scheduler and resource allocator that are responsible for the UE selection and radio resource assignment, respectively [16]. The guaranteed scheduling latency and exact DRX cycle is used to ensure individual IoT UE's requirements such as the power saving and real-time demand. In addition, we maintain overall system performance according to the CQI. The details of three metrics as well as the fuzzy membership function, low-complexity fuzzy intersection function, and an adaptive power control function are described as follows.

5.1. Exact DRX Cycle. In the proposed scheme, the exact DRX cycle timer is used to prevent the unexpected delay caused by the sleep mode. To achieve this goal, the eNodeB must schedule the UEs within the on duration period in order to successfully receive the scheduling information. Because the DRX parameters are configured by the eNodeB, the eNodeB has the ability to handle the UE's exact DRX cycle. The maximum DRX cycle that is considered as the timer of the short and long DRX cycle combination is shown in Figure 4. Because of the constraint of 50 subframes DRX-to-connected latency, the maximum DRX cycle must be within the $k-32$ ms and k ms for the LTE networks. The exact DRX cycle must be the period of $k - 32$ ms to k-Opportunity for DRX time. In the proposed scheme, if the exact DRX cycle expires, it will be restarted when the IoT UEs are scheduled and allocated RBs again by the eNodeB. The RB, which consists of 12 subcarriers of 15 kHz, and the 1 ms subframe, which consists of two RBs, are the smallest entity that can be scheduled for the UEs in the frequency domain and time domain, respectively. Furthermore, network operators can adjust the length of the maximum DRX cycle to apply to the LTE-Advanced networks.

5.2. Guaranteed Scheduling Latency. In the proposed scheme, the guaranteed scheduling latency configured by IoT manufacture is used to specify and to ensure the real-time requirement of IoT UEs. The scheduling latency, a real-time indicator, is the nonproductive time between the end of last RB and the start of next RB as shown in Figure 5. The scheduling latency usually consists of the context switching time and scheduling decision time. In the proposed scheme, each IoT UE has a timer of the guaranteed scheduling latency to ensure the scheduling latency. The timer will be restarted when the UEs are scheduled and allocated RBs again by the eNodeB if the guaranteed scheduling latency expires.

Table 1: CQI table.

CQI index	Modulation scheme	Coding rate ∗ 1024	Efficiency
0		Out of range	
1	QPSK	78	0.15
2	QPSK	120	0.23
3	QPSK	193	0.38
4	QPSK	308	0.60
5	QPSK	449	0.88
6	QPSK	602	1.18
7	16-QAM	378	1.48
8	16-QAM	490	1.91
9	16-QAM	616	2.41
10	64-QAM	466	2.73
11	64-QAM	567	3.32
12	64-QAM	666	3.90
13	64-QAM	772	4.52
14	64-QAM	873	5.12
15	64-QAM	948	5.55

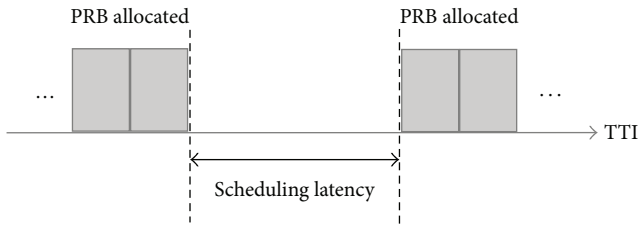


Figure 5: An example of the scheduling latency.

5.3. Channel Quality Indicator. The four-bit channel quality indicator ranged from 0 to 15, as shown in Table 1, which indicates the highest Modulation and Coding Scheme (MCS) and the maximum data rate with less than 10% block error ratio that UE can support. The CQI is calculated at the UEs according to its Signal to Interference plus Noise Ratio (SINR) and the CQI is regularly reported from the UE to the eNodeB. The reporting intervals of the CQI lie between 2 and 160 ms [13]. A higher CQI value will result in higher data rate and higher MCS profile used. In the proposed scheme, we maintain the overall system performance based on the CQI as high as possible. Since relatively sufficient buffer is used in the eNodeB, the channel quality rather than service traffic dominates the QoS in the cellular networks.

5.4. Fuzzy Membership Function. In the proposed scheme, the goal of the membership functions is to convert three metrics, the UE's exact DRX cycle, scheduling latency, and CQI, into degree of memberships. Recall that, in Problem Formulation, we attempt to maximize the scheduling efficiency by adjusting the fuzzy set α. Since this fuzzy set is related to these three metrics, we fuzzified them into memberships to figure out the intersection of the three metrics by the intersection function. The three fuzzy membership functions used in the scheme are given as

$$\mu_{e,k} = \left(\frac{{E_k}^2}{\sum_j {E_j}^2} \right)^{-2},$$

$$\mu_{l,k} = \left(\frac{{L_k}^2}{\sum_j {L_j}^2} \right)^{-2}, \tag{6}$$

$$\mu_{c,k} = \left(\frac{\left(\mathrm{CQI}_{\max} - C_k\right)^2}{\sum_j \left(\mathrm{CQI}_{\max} - C_j\right)^2} \right)^{-2},$$

where the notation μ denotes the fuzzy membership without loss of generality. The E_k, L_k, and C_k are the exact DRX cycle, guaranteed scheduling latency, and the CQI of kth IoT UE, respectively. The $\mathrm{CQI}_{\max}$ which equals 15 is the maximum CQI index.

5.5. Low-Complexity Fuzzy Intersection Function. In this paper, to reduce the computational complexity of the scheduling decision, we propose a low-complexity and simple intersection approach. The scheduling decision is made according to the three metrics, the UE's exact DRX cycle, scheduling latency, and CQI. The intersection function is formulated as

$$\alpha_k \left(E_k, L_k, C_k\right) = \mu_{e,k} \cdot \mu_{l,k} \cdot \mu_{c,k}, \tag{7}$$

where each IoT UE's membership intersection can be regarded as the volume or capacity in the exact DRX cycle, guaranteed scheduling latency, and the CQI planes from the capacity point of view. To present our ideas clearly, we illustrate an example of the IoT UEs intersections as shown in Figure 6, in which each node presents the product of the UE's three metric memberships.

The IoT UEs which are closer to the upper right side have higher priority to be considered for scheduling. The number of IoT UEs can be scheduled and allocated RBs at a time instant depending on the available system bandwidth of the eNodeB.

5.6. Adaptive Power Control Algorithm. In the future, a lot of IOT UEs are expected to be deployed indoors where the small cells, such as femtocells, are necessary to be deployed to achieve a high spectral efficiency and to provide better QoS for IoT UEs. In our previous study [17], we proposed an effective Adaptive Smart Power Control Algorithm (ASPCA), which can be applied to cluster IoT UEs in the coverage hole and to deal with the cross-tier interference issues by determining an appropriate serving range of home eNodeB without requiring complicated negotiation among them. The proposed ASPCA not only improves the overall system performance but also takes QoS of UEs into account.

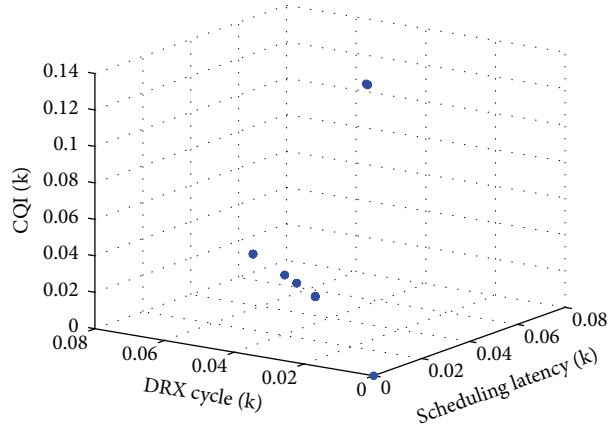


Figure 6: The example of three metrics intersection, the CQI, exact DRX cycle, and the guaranteed scheduling latency.

6. Simulation Model

In this section, we introduce a heterogeneous simulation model, in which an outdoor macro eNodeB is located in center of the map, and six surrounding femto eNodeBs are deployed around the macro eNodeB. The macro eNodeB's coverage is overlaid with the six femto eNodeBs. In addition, about 70% of IoT UEs are randomly deployed in a $50 \times 50\ \mathrm{m}^2$ indoor area, and others are randomly deployed in the range of 100 meters of each femto eNodeB. In addition, all IoT UEs of macro eNodeB are randomly deployed in the Range of Interest (ROI). The main simulation parameters are summarized in Table 2. The transmission power of the femto eNodeB is 20 dBm, and the transmission power of the macro eNodeB is 46 dBm. The details of channel fading model, propagation model, data rate calculation, scheduling latency parameter, and the power consumption model are described as follows.

6.1. Channel Fading Model. To provide realistic results, we use the Rayleigh fading channel model to reflect the effect of urban environments. Because of multipath propagation, the channel fading will lead to Bit Error Rate (BER) increasing. The BER is defined as the number of received bit errors divided by the total number of transmitted bits over a communication channel. The analytical equation in the Rayleigh channel [18] is given as

$$\mathrm{BER} = \frac{M-1}{M \log_2 (M)} \left(1 - \sqrt{\frac{\delta}{\delta + 1}}\right),$$

$$\delta = \frac{3 \log_2 (M) \cdot E_{\mathrm{rb}}}{\left(I_{\mathrm{rb}}/\Delta f_{\mathrm{rb}} + N_o\right) \Delta f_{\mathrm{rb}}} \gamma \cdot \frac{1}{M^2 - 1}, \tag{8}$$

Table 2: Simulation configuration.

Network layout	Hexagonal/2-tier
Transmission scheme	Downlink
Intercell distance	100 m
Carrier frequency	Band 3, 1800 MHz
System bandwidth	20 MHz for macrocell, 1.4 MHz for femtocells
Antenna type	Omnidirectional with 0 dB gain
Noise density	−174 dBm/Hz
Propagation model	TS 36.942 for urban environment
Penetration loss for femtocells	12 dBm
Channel fading model	Rayleigh fading
Number of macrocells	1
Number of femtocells	6
Transmit power of macrocell	46 dBm
Maximum transmission power of femtocells	20 dBm
Minimum transmission power of femtocells	−20 dBm
Number of macrocell users	200
Number of users per femtocell	20
Proportion of cell-edge users per femtocell	0.3
Simulation duration	512 ms

where M denotes the constellation size of adopted M-ary QAM signal which is related to the adopted MCS. The adopted MCS is configured according to the CQI reported by each IoT UE.

6.2. Propagation Model and Data Rate Calculation. The propagation model specified by 3GPP [19] has been considered and implemented for the urban environment as

$$\begin{aligned} \mathrm{PL}_{\mathrm{outdoor}} &= 40\left(1 - 4e^{-3}\,\mathrm{Dhb}\right)\log_{10}(R) \\ &\quad - 10\log_{10}(\mathrm{Dhb}) + 21\log_{10}(f) + 80\,\mathrm{dB}, \\ \mathrm{PL}_{\mathrm{indoor}} &= 40\left(1 - 4e^{-3}\,\mathrm{Dhb}\right)\log_{10}(R) \\ &\quad - 10\log_{10}(\mathrm{Dhb}) + 21\log_{10}(f) + 80\,\mathrm{dB} \\ &\quad + \text{pentration loss}, \end{aligned} \tag{9}$$

where R is the separation from the cell to the user in kilometers, f is the carrier frequency in MHz, and Dhb is the antenna height of cell in meters. In our simulation, a lookup table, as given in Table 3, is used to map SINR estimate to spectral efficiency [20] in order to calculate the data rate.

6.3. Scheduling Latency Parameters. The QoS Class Identifier (QCI) has been standardized [1] as shown in Table 4, in which the packet delay budget is the latency that a packet receives between the IoT UE and the eNodeB. In the simulation, fixed 50 ms delay budgets are used as the scheduling latency of the IoT UEs for performance evaluation.

Table 3: Lookup table for mapping SINR estimate to spectral efficiency.

CQI index	SINR estimate (dB)	Data rate (bps/Hz)
1	−6.7	0.1523
2	−4.7	0.2344
3	−2.3	0.3770
4	0.2	0.6016
5	2.4	0.8770
6	4.3	1.1758
7	5.9	1.4766
8	8.1	1.9141
9	10.3	2.4063
10	11.7	2.7305
11	14.1	3.3323
12	16.3	3.9023
13	18.7	4.5234
14	21.0	5.1152
15	22.7	5.5547

Table 4: QCI characteristics.

QCI index	Packet error rate	Packet delay budget (ms)
1	10^{-2}	100
2	10^{-3}	150
3	10^{-3}	50
4	10^{-6}	300
5	10^{-6}	100
6	10^{-6}	300
7	10^{-3}	100
8	10^{-6}	300
9	10^{-6}	300

6.4. Power Consumption Model. In this section, we introduce a power efficiency indicator to emulate the power consumption of DRX mechanism. In the simulation, unlike other studies, we focus on investigating the power consumption of the scheduling information reception on the PDCCH. A lookup table to map CQI index to power consumption is shown in Table 5. The power efficiency indicator is given as

$$\mathrm{PE} = \frac{P_{\mathrm{subframe}}}{N_{\mathrm{bit}}^{\mathrm{subframe}}}, \tag{10}$$

where the $P_{\mathrm{subframe}} = 500\,\mathrm{mw/ms}$ denotes the UE's power consumption of reception and the $N_{\mathrm{bit}}^{\mathrm{subframe}}$ denotes the number of bits per subframe. By defining the power efficiency, we can calculate the IoT UE's power consumption according to its signal quality. In the case of DRX adapted, the IoT UE's power consumption is defined as the PE multiplied by the number of PDCCH bits. On the other hand, the P_{subframe}

Table 5: Lookup table for mapping CQI to power consumption.

CQI index	CCE aggregation level	Number of PDCCH bits	Power consumption (mw)
1	8	576	11520.00
2	4	288	3789.47
3	4	288	2285.71
4	4	288	1440.00
5	2	144	489.79
6	2	144	363.63
7	2	144	290.32
8	2	144	225.00
9	2	144	178.21
10	2	144	157.20
11	2	144	129.26
12	2	144	109.92
13	1	72	47.43
14	1	72	41.86
15	1	72	38.62

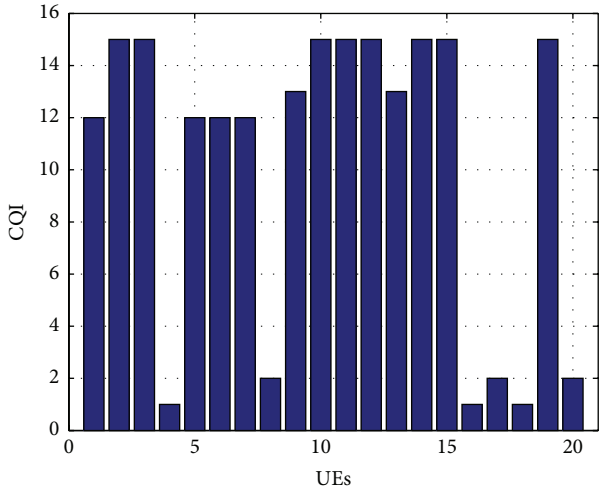


Figure 7: The CQI of IoT UEs of a femto eNodeB.

multiplied by the UE's scheduling latency is defined as the IoT UE's power consumption in the case of being without DRX.

7. Results and Discussion

In this section, we present simulation results to verify that our proposed scheme has the ability to ensure the DRX constraint and real-time requirement and to maintain overall system performance in IoT over the LTE/LTE-Advanced networks. First, we investigate the scheduling performance in terms of the exact DRX cycle and guaranteed scheduling latency as shown in Figure 7 to Figure 10. After that, we examine the power consumption efficiency of our proposed scheme as shown in Figure 11. Then we investigate the performance of the DRX mechanism in terms of average data rate as

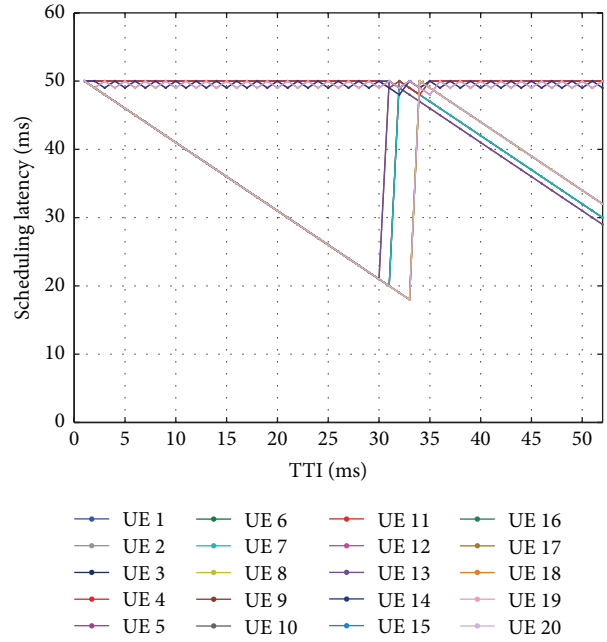


Figure 8: The performance of guaranteed scheduling latency in proposed scheme.

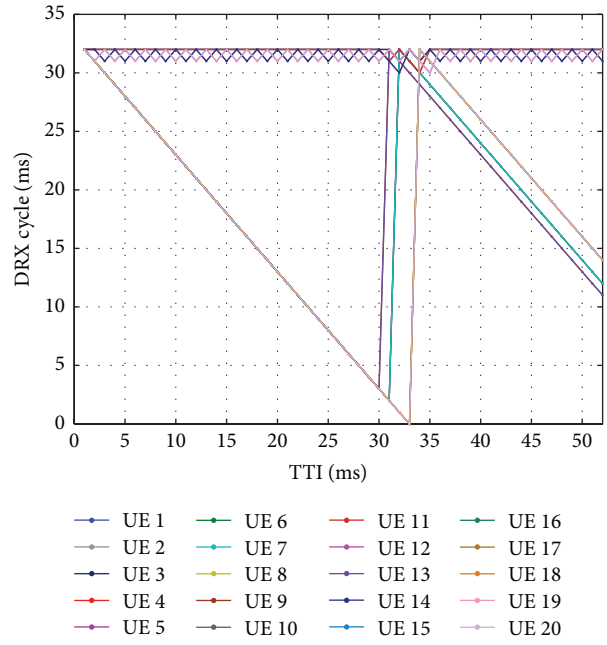


Figure 9: The performance of exact DRX cycle in proposed scheme.

shown in Figure 12. Next, we compare the proposed scheme with conventional schemes, Round-Robin (RR), and Max-C/I scheme as shown in Figure 13. Finally, we show that the requirements of exact DRX cycle and guaranteed scheduling latency can be satisfied as shown in Figure 14.

In our simulation, the CQI does not vary since the low-mobility of IoT UEs is assumed and deployed in the indoor environment, thence no fast-fading effects. In addition, 32 ms exact DRX cycle, 50 ms guaranteed scheduling latency, and

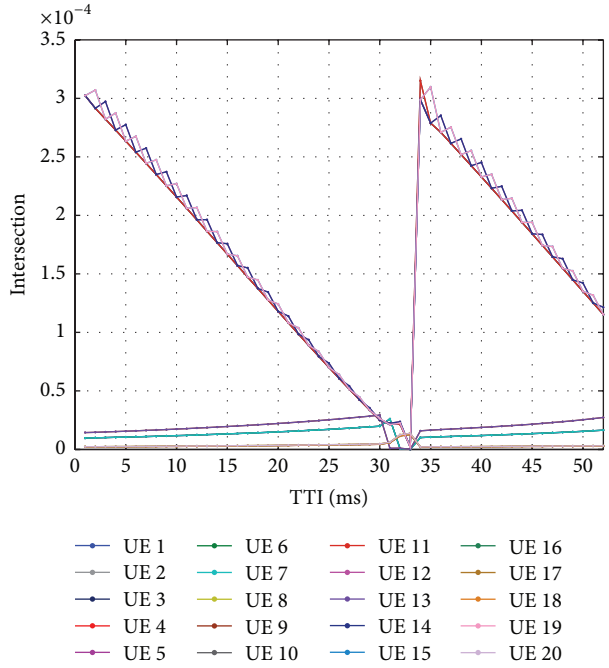


Figure 10: The fuzzy membership intersections of three metrics.

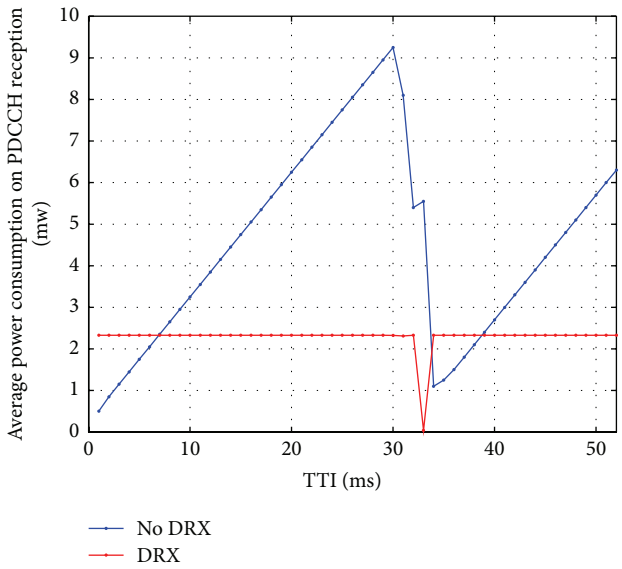


Figure 11: The average power consumption comparison of the proposed scheme with DRX and the proposed scheme without DRX.

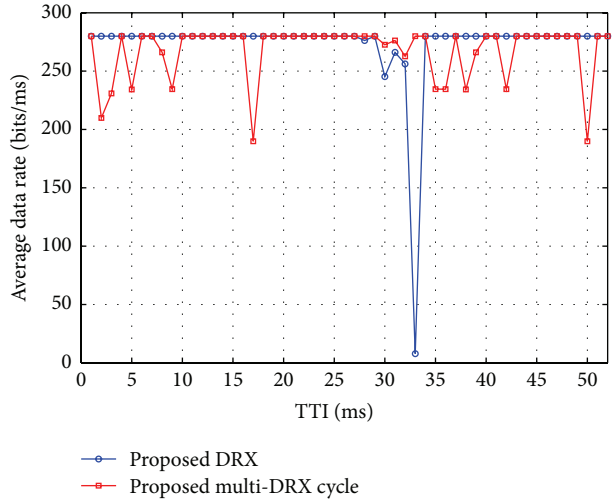


Figure 12: The average data rate comparison of single DRX cycle and multi-DRX cycle.

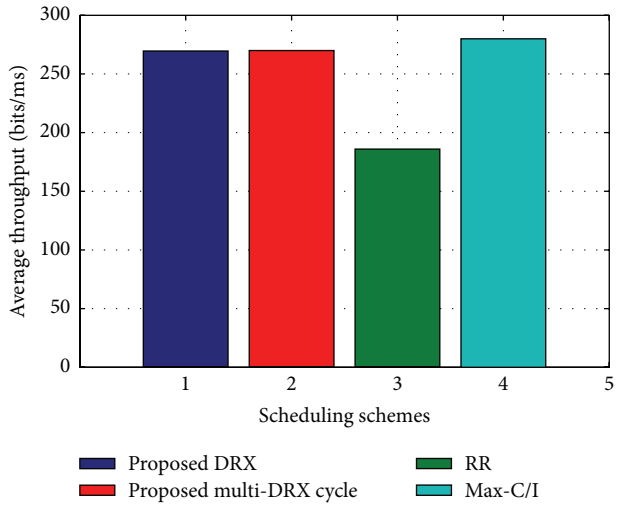


Figure 13: The throughput comparison of the proposed scheme, Round-Robin (RR) scheme, and the Max-C/I scheme.

1 ms inactivity timer are used in the normal case. In the multi-DRX cycle case, we examine the impact of six different DRX cycles, 32 ms, 16 ms, 8 ms, 2 ms, and 1 ms, which are randomly assigned to the IoT UEs.

Our simulator generates three metrics, the CQI, guaranteed scheduling latency, and the exact RDX cycle as shown in Figures 7, 8, and 9. The results of these three metrics intersections are shown in Figure 10, in which we can clearly see that there are few intersections at about 32 ms since fixed 32 ms exact DRX cycle is used. Even though fixed 50 ms of guaranteed scheduling latency is adopted, the scheduling decision is still dominated by short exact DRX cycle due to more stringent time restriction. Comparing Figure 6 with Figures 7 and 8, we can observe that even though the outdoor UEs of femtocell 0, UE 4, UE 8, UE 16, UE 17, UE 18, and UE 19, have bad CQI, they can still reach their constraints of the guaranteed scheduling latency and exact DRX cycle. In addition, it should be noted that there are only 6 RBs that can be allocated to IoT UEs for a TTI due to intended narrowing of 1.4 MHz system bandwidth.

A comparison of average power consumption between the case of being with DRX and the case of being without DRX mechanism is shown in Figure 11. The figure shows that the proposed scheme can conserve about half of energy consumption compared to the case of being without DRX mechanism.

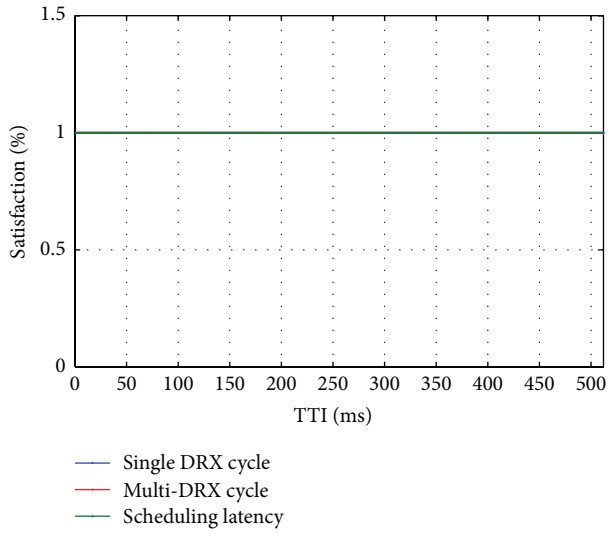


Figure 14: The satisfaction of the exact DRX cycle and guaranteed scheduling latency.

A comparison of single DRX cycle and multi-DRX cycle is shown in Figure 12. We can observe that the multi-DRX cycle has more stable data rate than the single DRX cycle. Evidently, the diversity of metrics, such as multi-DRX cycle and multischeduling latency, can benefit our scheme to make better decision.

The throughput comparison of the proposed scheme, RR, and the Max-C/I scheme is shown in Figure 13. Since Round-Robin scheme serves IoT UEs in turn without taking the instantaneous channel quality into account, it gets the worst throughput. Because the Max-C/I scheme prefers to serve UEs which have better channel condition, it gets the best throughput but loses fairness. The proposed scheme can provide a better throughput as well as fairness for the IoT UEs.

Finally, the satisfaction of the proposed scheme is shown in Figure 14, which shows that the requirement of the exact DRX cycle and the guaranteed scheduling latency can be accomplished by our fuzzy-based power saving scheduling scheme.

8. Conclusions

In this paper, we proposed a fuzzy-based power saving scheduling scheme for IoT over the LTE/LTE-Advanced networks to deal with the issues of the radio resource management and power consumption as well as the unexpected delay caused by the DRX mechanism from the scheduling and resource allocation perspective. The simulation results show that our scheme has the ability to leverage the tradeoff among three individual metrics and overall system performance. Hence, the IoT UE's requirements can be guaranteed. In addition, we exam the performance of multi-DRX cycle, and we find that the diversity of the system metrics can help fuzzy-based approach make better decision. The proposed scheme is useful to be applied to the IoT over the LTE/LTE-Advance networks in practice.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This work was partially supported by Ministry of Science and Technology, Taiwan, under Grants nos. NSC 102-2221-E-008-039-MY3 and 104-2221-E-008-039-MY3.

References

[1] J.-M. Liang, J.-J. Chen, H.-H. Cheng, and Y.-C. Tseng, "An energy-efficient sleep scheduling with QoS consideration in 3GPP LTE-advanced networks for internet of things," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 3, no. 1, pp. 13–22, 2013.

[2] Y.-W. Chen, M.-H. Lin, and Y.-T. Su, "Power saving efficiency analysis of QoS scheduling in the LTE network featuring discontinuous reception operation," *IEICE Transactions on Communications*, vol. 97, no. 10, pp. 2212–2221, 2014.

[3] G. Stea and A. Virdis, "A comprehensive simulation analysis of LTE discontinuous reception (DRX)," *Computer Networks*, vol. 73, pp. 22–40, 2014.

[4] K. Wang, X. Li, and H. Ji, "Modeling 3GPP LTE advanced DRX mechanism under multimedia traffic," *IEEE Communications Letters*, vol. 18, no. 7, pp. 1238–1241, 2014.

[5] K. Zhou, N. Nikaein, and T. Spyropoulos, "LTE/LTE-a discontinuous reception modeling for machine type communications," *IEEE Wireless Communications Letters*, vol. 2, no. 1, pp. 102–105, 2013.

[6] S. Jin and D. Qiao, "Numerical analysis of the power saving with a bursty traffic model in LTE-Advanced networks," *Computer Networks*, vol. 73, pp. 72–83, 2014.

[7] S. Fowler, A. O. Shahidullah, M. Osman, J. M. Karlsson, and D. Yuan, "Analytical evaluation of extended DRX with additional active cycles for light traffic," *Computer Networks*, vol. 77, pp. 90–102, 2015.

[8] E. Liu and W. Ren, "Performance analysis of a generalized and autonomous DRX scheme," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 5, pp. 2148–2153, 2014.

[9] A. T. Koc, S. C. Jha, R. Vannithamby, and M. Torlak, "Device power saving and latency optimization in LTE-a networks through DRX configuration," *IEEE Transactions on Wireless Communications*, vol. 13, no. 5, pp. 2614–2625, 2014.

[10] K.-C. Ting, H.-C. Wang, C.-C. Tseng, and F.-C. Kuo, "Energy-efficient DRX scheduling for QoS traffic in LTE networks," in *Proceedings of the 9th IEEE International Symposium on Parallel and Distributed Processing with Applications (ISPA '11)*, pp. 213–218, IEEE, Busan, South Korea, May 2011.

[11] L.-P. Tung, Y.-D. Lin, Y.-H. Kuo, Y.-C. Lai, and K. M. Sivalingam, "Reducing power consumption in LTE data scheduling with the constraints of channel condition and QoS," *Computer Networks*, vol. 75, pp. 149–159, 2014.

[12] 3GPP TS 36.321, *Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) Protocol Specification*, v12.7.0, 2015.

[13] 3GPP TS 36.321, *Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Layer Procedures*, v12.7.0, 2015.

[14] 3GPP TS 36.331, "Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification," v 12.7.0, 2015.

[15] I. Grigorik, *High Performance Browser Networking: What Every Web Developer Should Know about Networking and Web performance*, O'Reilly Media, 2013.

[16] M. I. Salman, M. Q. Abdulhasan, C. K. Ng, N. K. Noordin, A. Sali, and B. Mohd Ali, "Radio resource management for green 3gpp long term evolution cellular networks: review and tradeoffs," *IETE Technical Review*, vol. 30, no. 3, pp. 257–269, 2013.

[17] Y.-W. Kuo, L.-D. Chou, and Y.-M. Chen, "Adaptive smart power control algorithm for LTE downlink cross-tier interference avoidance," in *Proceedings of the 11th EAI International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness (QSHINE '15)*, Taipei, Taiwan, August 2015.

[18] Y. S. Cho, J. Kim, W. Y. Yang, and C. G. Kang, *MIMO-OFDM Wireless Communications with MATLAB*, Wiley, 2010.

[19] 3GPP TS 36.942, "Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Frequency (RF) system scenarios," Tech. Rep. v12.0.0, 2014.

[20] H. Zarrinkoub, *Understanding LTE with MATLAB: From Mathematical Modeling to Simulation and Prototyping*, John Wiley & Sons, 2014.

Hindawi Publishing Corporation
Mobile Information Systems
Volume 2015, Article ID 931765, 11 pages
http://dx.doi.org/10.1155/2015/931765

Research Article

Data-Sharing Method for Multi-Smart Devices at Close Range

Myoungbeom Chung[1] and Ilju Ko[2]

[1]*Division of Computer Engineering, Sungkyul University, Gyeonggi-do 430-742, Republic of Korea*
[2]*Department of Global Media, Soongsil University, Seoul 156-743, Republic of Korea*

Correspondence should be addressed to Myoungbeom Chung; nzin@sungkyul.ac.kr and Ilju Ko; andy@ssu.ac.kr

Received 8 July 2015; Revised 22 October 2015; Accepted 9 November 2015

Academic Editor: Kamal Deep Singh

We proposed a useful data-sharing method among multi-smart devices at close range using inaudible frequencies and Wi-Fi. The existing near data-sharing methods mostly use Bluetooth technology, but these methods have the problem of being unable to be operated using different operating systems. To correct this flaw, the proposed method that uses inaudible frequencies through the inner speaker and microphone of smart device can solve the problems of the existing methods. Using the proposed method, the sending device generates trigger signals composed of inaudible sound. Moreover, smart devices that receive the signals obtain the shared data from the sending device through Wi-Fi. To evaluate the efficacy of the proposed method, we developed a near data-sharing application based on the trigger signals and conducted a performance evaluation experiment. The success rate of the proposed method was 98.8%. Furthermore, we tested the user usability of the Bump application and the proposed method and found that the proposed method is more useful than Bump. Therefore, the proposed method is an effective approach for sharing data practically among multi-smart devices at close range.

1. Introduction

With the development of the mobile phone and communication technology, the existing mobile phone has changed into a smart device that has various functions, such as camera, mp3 player, and on/offline mobile games. Smart device users can share their daily lives and thoughts online through social network services (SNS) such as Facebook and Twitter. Users can share not only photos, such as those of their daily lives or vacation, but also data such as name card and work materials, among others, during business meetings. The data-sharing methods among smart devices are divided into those that can be used at close range using Bluetooth or Wi-Fi Direct and at a long distance through Wi-Fi and a server computer.

The data-sharing methods for close range use Bluetooth, Wi-Fi Direct, Airdrop, or NFC. Data can be sent directly without the use of extra service applications. The method that uses Bluetooth requires the same application to be installed on each smart device, and data sharing can occur through the pairing of smart devices [1]. The maximum speed of data transmission using Bluetooth is 24 Mbps, and it requires only the inner Bluetooth module of a smart device. However, as this method can only be used with the same operating system (OS), it cannot transmit data between a smart device that uses an Android OS and a smart device that uses iOS. Wi-Fi Direct is the upgraded model of the Bluetooth feature for near data transmission and has a data transmission speed of 300 Mbps [2]. Therefore, it is more suitable for sending big data to multimedia devices such as cameras or printers than using data transmission technology among smart devices. Similarly, Apple Inc. created the Airdrop technology that uses Bluetooth instead of Wi-Fi Direct [3]. Airdrop can transmit data among smart devices at close range. However, similar to Bluetooth, this technology can only be used with the same OS. In addition, Android Beam and S Beam are newly released data sharing methods. They can work well with smart devices with built-in NFC. Android Beam can share much data, such as address lists, URLs of web browsers, GPS data of Google Maps, market addresses of the Play store, and applications downloaded from the Android market [4, 5]. S Beam can share media data, such as mp3 music files, photo images, and movie files [6, 7]. However, Android Beam and S Beam cannot work with smart devices without NFC, such as iPhones and iPads.

Data-sharing methods for long distance use Wi-Fi and a server computer. Users are required to use extra services

or SNS and become friends with other users to be able to share data. The typical applications for long-distance data sharing are Kakao Talk, Whatsapp, and Viber, among others [8–10]. These applications can be used for sharing data and exchanging photo files between two friends. The advantage of this method is that smart device users can share data at a long distance, such as when users are in different countries. However, these data-sharing methods for long distance have the disadvantage of being unable to send data to any smart device if the user does not join the application. Users must be friends with each other to share data as a pretask. To overcome these disadvantages, Bump Technologies created the Bump application in 2010 [11]. The Bump application uses Wi-Fi, global positioning system (GPS) information, and acceleration sensor to share data at close range. This technology can share data among smart devices without joining any application or becoming friends with other users. Moreover, it can be used with different OS. As Bump Technologies use Open API and the Bump server, the Bump application can be used for the various applications of smart phones. However, Bump can only conduct 1 : 1 data sharing. If three smart devices are used to share data in the same place and at the same time, the Bump application will not work as the smart device that sends the data cannot know which of the other smart devices will receive the data.

Therefore, we propose a useful data-sharing method among multi-smart devices at close range that can solve the problems of the existing methods. The proposed method uses inaudible frequencies as a trigger signal, instead of the shaking motion of the Bump application, and Wi-Fi and GPS information for sharing data among smart devices at close range. The audible range known scientifically is 20 Hz to 22,000 Hz. However, the audible range for most people is below 20,000 Hz [12, 13], and the older the people are, the lower their audible range is. People in the 40s and 50s can hear frequencies below 13,000 Hz [14]. Therefore, the proposed method uses a frequency of over 18,000 Hz. The smart device (sending device) that shares data generates a trigger signal from an inner speaker using a combination of inaudible frequencies. At the same time, the sending device sends the shared data and its current GPS information to the sharing server. Smart devices (receiving devices) are located around the sending device and receive the data-sharing process surrounded by sound through an inner microphone. If the receiving device perceives the trigger signal assigned as the data-sharing signal, it sends its current GPS information to the server and downloads the shared data from the server. We upgrade the high-frequency signal as the trigger signal rather than the existing high-frequency signal in the literature. As the trigger signal of the proposed method uses two or more high frequencies, we can upgrade the reaching distance of the trigger and the accuracy of data transmission.

We developed a data-sharing application based on a smart device and conducted a data-sharing experiment according to distance to evaluate whether the proposed method is suitable for data sharing among smart devices at close range. The result of the experiment showed 98% accuracy in the data-sharing test within a 5 m distance irrespective of the OS of the smart devices. We then conducted a 1 : 1 and 1 : N sharing accuracy test as a comparative experiment using the Bump application. We surveyed the experiment participants to evaluate the usability of the proposed method. As the Bump application has not supported the Bump API and Bump server since January 31, 2014, we developed an application that works like the Bump application based on the Bump algorithm, and we set up a server like the Bump server to share data. The results of the sharing accuracy experiment and of the survey on usability showed that the proposed method was better than the Bump application. Therefore, the proposed method could be a useful method for sharing data effectively among multi-smart devices at close range.

The present paper is organized as follows. In Section 2, we explain the Bump application that is useful in sharing data at close range among smart devices and the existing research that uses high frequency, which we use as the trigger signal in the proposed method. In Section 3, we present the general architecture of the proposed method for sharing data among multi-smart devices, the process of the trigger signal using high frequency, and the sharing server. In Section 4, we show the data-sharing application that applies the proposed method and report the experimental results of the data transmission accuracy and usability of the application. We conclude the study in Section 5.

2. Previous Work

This section briefly explains the existing methods for near data-sharing technology and presents the general architecture of the Bump application. We explain the existing research that uses high frequency to send information. Although near data-sharing technologies among smart devices use Bluetooth, Wi-Fi Direct, Airdrop, and NFC, among others, these technologies can only use the same OS [15–17]. To solve this problem, Bump Technologies created the Bump application and opened the Bump API. This technology has been adopted in various smart device applications of iOS and Android since 2012 [18–20].

The Bump application uses three inner sensors of the smart device, unlike Bluetooth, Wi-Fi Direct, and Airdrop. Three inner sensors are the accelerometer, GPS, and Wi-Fi. The accelerometer detects the user's shaking action to begin data sharing, and the GPS obtains the location information of each device [21, 22]. The sending device sends the location information and shared data to the Bump server through Wi-Fi, while the receiving device downloads the shared data from the Bump server also through Wi-Fi. Therefore, the Bump application involves two smart devices and one Bump server. The general architecture of the Bump application is shown in Figure 1.

In Figure 1, we redraw the workflow of the Bump API, which Ahren suggested [23]. If the sending device and the receiving device do the shaking action to share data, each smart device sends its GPS information through 3G, long-term evolution (LTE), or Wi-Fi to the Bump server (①). The Bump server assesses whether the sending device and the

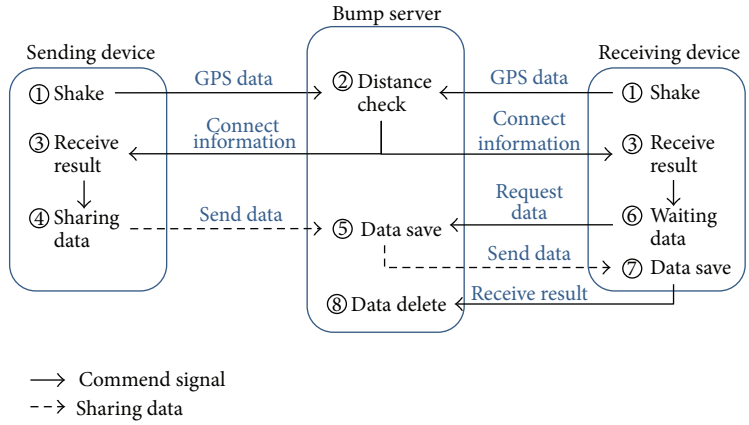


Figure 1: General architecture of the Bump application.

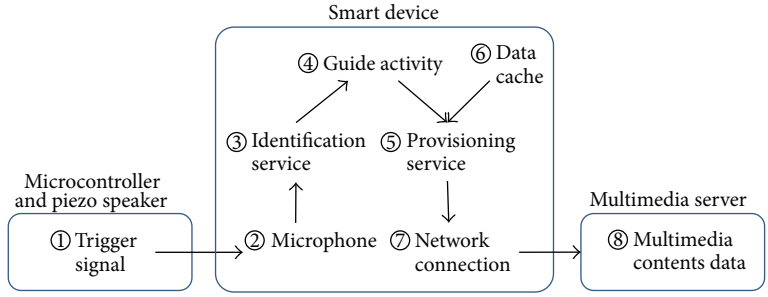


Figure 2: General architecture of the SmartGuide.

receiving device are close using the GPS information of each smart device and connecting time to the Bump server (②). The Bump server then sends the connect information to each smart device on whether the shaking action is successful (③). If the connect information is successful, the sending device sends the shared data to the Bump server (④), which saves the data (⑤). At this time, the receiving device requests the shared data and waits to save it (⑥). When the Bump server sends the data, the receiving device begins to save the data in its memory space through 3G, LTE, or Wi-Fi (⑦). Finally, when the transmission of the shared data ends, the receiving device sends the transmission result of the shared data to the Bump server, which deletes the shared data from the server (⑧).

The Bump application can share data easily among smart devices at close range regardless of the OS of the smart devices. It began to support the data-sharing service between a computer and a smart device in 2013 [24]. The sharable data formats of the Bump application are picture, video, and contact information, among others. By 2013, it had been downloaded 125 million times [25]. However, the Bump application requires the coinciding shaking action of the sending device and the receiving device, and it can only be used for 1 : 1 and not for 1 : N devices.

The existing methods that use high frequency are indoor positioning technology and indoor information transmission technology. Indoor positioning technology is a user position trace technology that uses a smart device inside a room or a building where the GPS of a smart device does not work well. Viacheslav proposed an indoor position trace algorithm based on high frequency using one mobile phone and four microphones [12]. He tested the output performance of the speaker of each mobile phone with various smart devices to determine the inaudible high frequency, which is suitable for a user's position trace. Bihler et al. suggested a trigger signal using the high frequency and developed a SmartGuide that could support user-related information according to a user's indoor position. Figure 2 shows that we redrew the general architecture of Bihler et al.'s SmartGuide [26].

In Figure 2, the trigger signal is generated by an 8-bit 3.2 MHz Freescale microcontroller and a simple piezo speaker (①). Bihler et al. used 20 kHz and 22 kHz as the high frequencies for the trigger signal, which could send 8-bit data at 208 ms. The smart device confirms the identification service through a microphone (②, ③) and runs the guide activity (④). If the smart device already has the information in the data cache (⑥), it sets the provisioning service (⑤). However, if the smart device does not have the information, the smart device connects to the multimedia server through 3G, LTE, or Wi-Fi to download the multimedia content (⑦, ⑧) and sets the provisioning service. Bihler's method is suitable for quiet indoor spaces or museum where GPS does not work well. However, as the trigger signal of this method requires many changes in high frequency for a short time, the trigger signal generates some

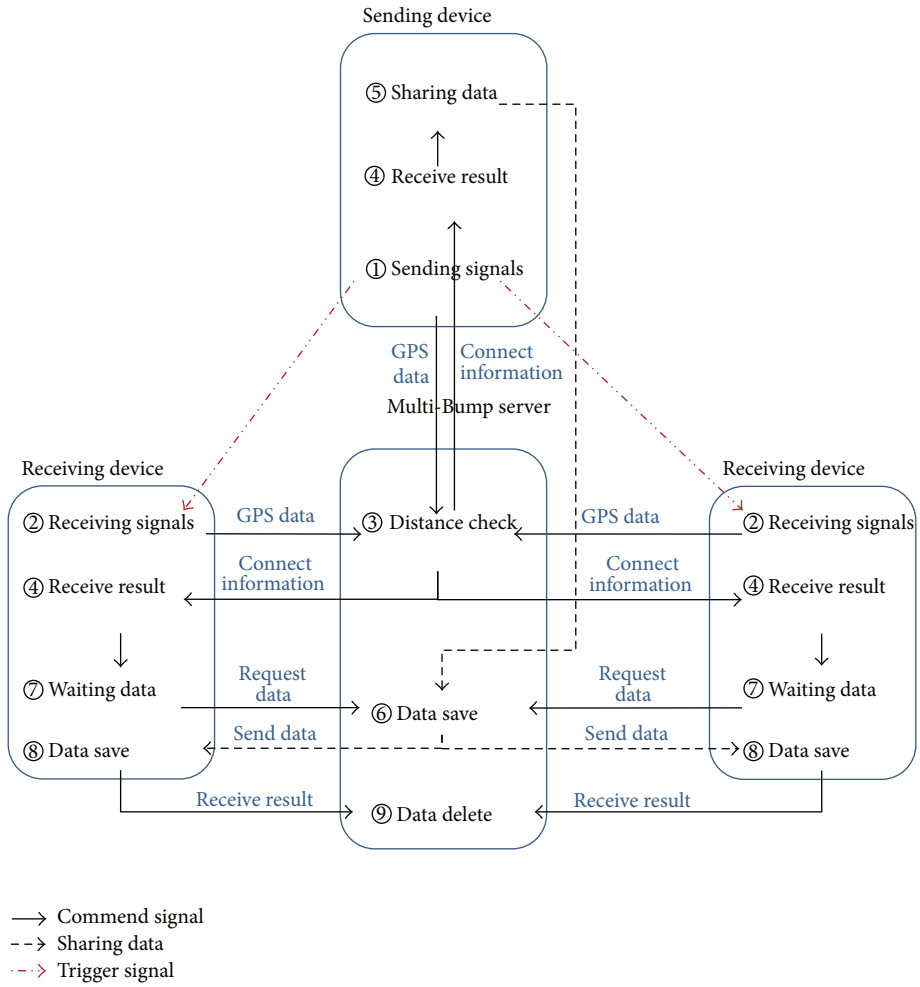


FIGURE 3: General architecture of the data-sharing method using high frequency as the trigger signal.

noise and the transmission accuracy of the trigger signal is low.

Next, the existing methods that use high frequency as little information transmission technology indoors were SonicData proposed by Nittala and Authentication method proposed by Kim [27, 28]. However, those methods have a problem that they need so many times to transmit a small quantity of data.

3. Data-Sharing Method for Multi-Smart Devices

This section explains the general architecture of the proposed data-sharing method among multi-smart devices using inaudible high frequency. The processing method of the trigger signals using high frequency and the sharing server is also presented. Figure 3 shows the general architecture of the proposed method. As shown in Figure 3, the sending device generates a trigger signal to share data and sends its own GPS information to the Multi-Bump server (①) at the same time. The nearby receiving devices analyze the surrounding sounds using the inner microphone in the smart device. When the receiving devices detect the trigger signal, the devices send their own SPG information to the Multi-Bump server (②). Then, the Multi-Bump server analyzes the received GPS data from each device and calculates the distance among the smart devices (③).

The Multi-Bump server checks the distance interval and sends the connect information according to the distance check to the sending device and the receiving devices (④). When the sending device receives the result that the data have been shared to all available devices, the sending device uploads the shared data to the Multi-Bump server (⑤). The Multi-Bump server then saves the uploaded shared data temporarily (⑥). When the waiting receiving devices request the shared data (⑦), the Multi-Bump server sends the shared data to each receiving device. The receiving devices begin to save the shared data in their respective memories (⑧), and each device sends the result of the received data to the Multi-Bump server when saving of the shared data ends. Finally, when the Multi-Bump server checks all the received results from the receiving devices, it deletes the shared data (⑨).

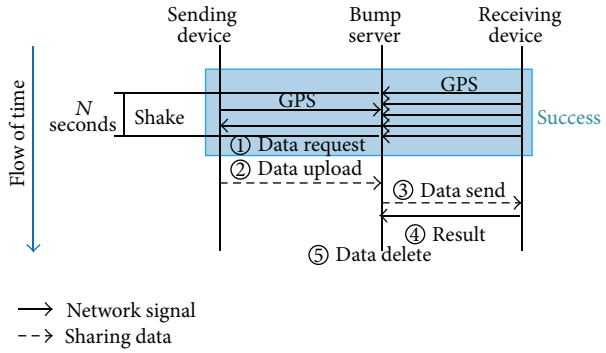


Figure 4: General data-sharing flow of the Bump application.

The difference between the proposed method and the existing Bump method is denoted by ① and ② movements in Figure 3. The Bump method uses a shaking action as the data-sharing signal. Moreover, it can share data only when each smart device shakes at the same time, as shown in Figure 4. Conversely, as the proposed method uses inaudible frequency as the data-sharing signal instead of a shaking action, it can share data with several smart devices at the same time. As shown in Figure 4, the Bump method can share data when the sending device and the receiving device shake within N seconds. In other words, if shake permission time of the Bump API is N seconds, the receiving device must shake within $\pm N/2$ seconds from the shaking time T of the sending device to share data.

Then, the sending device and the receiving device progress from ① to ⑤ in order and they share the data. However, if the shaking time is over, as shown in Figure 5, the Bump method for sharing data fails. In Figure 5, the reason for the first failure is that the receiving device performs the shaking action too early.

As the receiving device does the shaking action before $T - N/2$ seconds, the Bump method fails (①). The reason for the second fail is that the receiving device does the shaking action too late. As the receiving device does the shaking action after $T + N/2$ seconds, the Bump method fails (②). Conversely, the proposed method can solve this problem because the sending device generates the trigger signal using inaudible frequency for sharing data consistently until the receiving device detects the signal. Figure 6 shows the flow of the proposed method using inaudible frequency for data sharing.

As shown in Figure 6, we do not use N seconds, which is used in the Bump method. The sending device generates the trigger signal consistently instead of N seconds until the receiving device responds to detecting the trigger signal and sends its GPS information to the Multi-Bump server. When the receiving device detects the trigger signal, it sends its own GPS information to the Multi-Bump server. At this time, the server calculates the distance from the GPS information of each device and requests the shared data for the sending device according to the connection result (①). The sending device that requests the shared data stops to generate the trigger signal and continues the data sharing from step ③ to step ⑤. Furthermore, the proposed method can share data from one sending device to many receiving devices at the same time. Figure 7 shows the data-sharing flow of one sending device and two receiving devices.

In Figure 7, the sending device generates the trigger signal (①). At the same time, it sends its own GPS information to the Multi-Bump server and is waiting for the request for sharing data from the server (②). The receiving devices that detect the trigger signal send the GPS information of each device to the Multi-Bump server (③), which calculates the distance from the GPS information of all devices. If the connection result is successful, the server requests the sending of shared data from the sending device (④). Then, the server saves the number of receiving devices temporarily and sends the shared data, which are received by each receiving device from the sending device (⑤, ⑥). When the receiving devices obtain the shared data, the receiving devices send a message about the transmission result of the shared data to the server (⑦). The server counts the number of messages about the transmission result. If the number of the receiving devices is equal to the number of messages, the server deletes the shared data uploaded from the sending device (⑧). Therefore, the proposed method can share data easily in $1:N$ because it does not need to consider the time of the shaking action.

Next, the trigger signal used in the proposed method is applied to the control signal proposed by Chung and Choo [29]. Chung and Choo used two base signals and one low-latency key as the trigger signal. Two base signals are fixed high frequencies and the low-latency key is a high frequency that can be changed by a mobile central processing unit (CPU). Chung's method is used to determine why the arrival distance of the trigger signal is so far and why its accuracy is high. Therefore, we use two base signals and one low-latency key and the frequency as the trigger signal ranges from 18 kHz to 22 kHz. However, as the length of the trigger signal used in Chung and Choo's study is short, we make the trigger signal longer and generate it many times to send the trigger signal to various smart devices. To avoid distinguishing the trigger signal from the surrounding noise, we use the changeable low latency as the sharing key value in near data sharing. The pseudocode in Pseudocode 1 is the detecting code of the trigger signal applied to the sharing key using the low-latency. It also works in the receiving smart devices. In Pseudocode 1, $\mathbf{k}$ is the parameter for counting the trigger signal and $\mathbf{A_t}$ is the audio data obtained by the inner microphone of the smart device during $\mathbf{t}$ time. $\mathbf{A_t}$ is separated from each frequency $\mathbf{F_t}$ as fast Fourier transform (FFT). If the base signals and the low-latency key of the $\mathbf{F_t}$ values exist, $\mathbf{k}$ increases. If the base signals and the low-latency key of the $\mathbf{F_t}$ values do not exist, $\mathbf{k}$ sets the initialization as zero.

When the $\mathbf{k}$ parameter is over the threshold value $\boldsymbol{\alpha}$, the receiving device makes an assessment whether to receive the trigger signal and sends its own GPS information to the Multi-Bump server. Then, the receiving device requests the shared data and ends the process to detect the trigger signal.

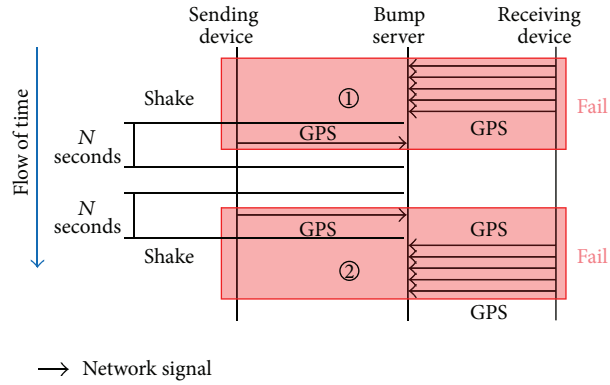


Figure 5: Example of a data sharing failure using the Bump method.

```
k ← 0;
While (receive A_t) Do
    F_t ← FFT(A_t);
    If (F_t has the base signals) Then
        If (F_t has key low-latency) Then
            k ← k + 1;
            If (k is more than threshold α) Then
                GPS data send to server;
                Request sharing data;
                Break;
            END If
        ELSE
        k ← 0;
        END If
    ELSE
        k ← 0;
    END IF
END WHILE
```

Pseudocode 1: Pseudocode for the trigger signal detection of the receiving devices.

4. Experiments on and Analysis of the Data-Sharing Method among Multi-Smart Devices

This section introduces the development of the application for sharing data among multi-smart devices using the proposed method. To evaluate the performance of the proposed method, we conducted a comparative experiment between the existing Bump method and the proposed method in near data sharing. Moreover, we discuss the survey on and the result of the usability of the proposed method. Figure 8 shows the screen of the Multi-Bump application using the proposed method. The initial screen of the Multi-Bump application is shown in Figure 8(a), and the screen of the Multi-Bump application when the application receives the shared data is presented in Figure 8(b). Figure 8(a) shows the application's image preview of the shared data, the "Select File" button, the "Change Key" button for selecting the trigger signal and now the trigger signal value, and the "Start Data Share" toggle button to start and end data sharing in order from the top. When the application initially runs, the preview shows "No Image" because it does not select any shared data. When the user selects a shared file, the preview shows a thumbnail image of the shared file. The base signals of the trigger signal use 19.0 kHz and 22.0 kHz within the 18 kHz to 22 kHz range. The range of the low-latency key value that the user can select is from 19.6 kHz to 21.4 kHz. The start frequency of the low-latency key value is 19.6, and this frequency is 600 Hz greater than 19.0 kHz as the base signal. Furthermore, the end frequency of the low-latency key value is 21.4 kHz, and this frequency is 600 Hz lower than 22.0 kHz as the base signal [28]. These frequencies are used to avoid interference between the base signals and the low-latency key signal. As shown in Figure 8(a), the Multi-Bump application can have 19 types of trigger signals. The low-latency key value shown in Figure 8(a) is 20.3 kHz (eight-channel).

If the receiving device wants to receive the shared data, it must set the same low-latency key value of the sending device. In Figure 8(b), the three position (①–③) states of the receiving device screen change when the shared data are received. Position ① as the preview for the shared data shows an activity indicator during the receiving of the shared data. Two buttons of position ② are changed from a usable state to a nonusable state to avoid the button control error during the receiving of the shared data. The progress bar of position ③ appears only when the receiving device is receiving the shared data. It disappears when the receiving device finishes receiving the shared data.

In the experiment on the performance evaluation of the Multi-Bump application, we conducted a test to determine the suitable length of a trigger signal and threshold α. Chung and Choo used 104 ms as the length of the trigger signal, and threshold α was 52 ms. Thus, we tested changing the length of the trigger signal to 52, 78, 104, 130, 156, 182, 208, and 234 ms and set threshold α to 50%. The experimental environment was an office laboratory where the noise level was maintained at less than 40 dB. The distance of the sending device from the receiving device was from 1 m to 5 m. We sent the trigger signal 100 times according to the length of the trigger signal and the distance of the devices. We used 19.6 kHz as the low-latency key value. iPhone 6 was used as the sending

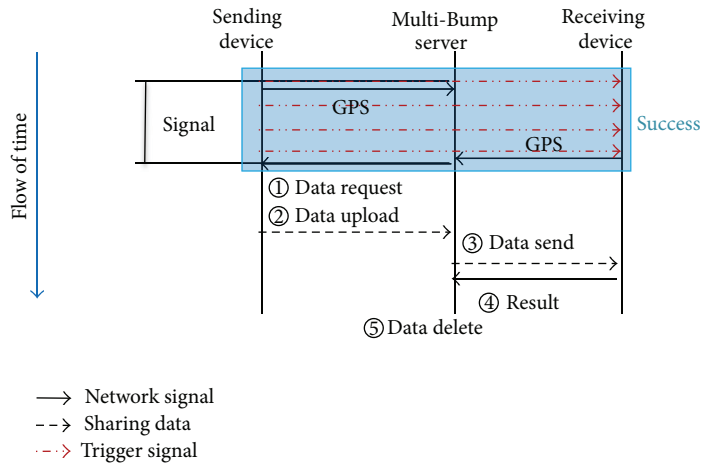


FIGURE 6: Flow of the proposed method using inaudible frequency for data sharing.

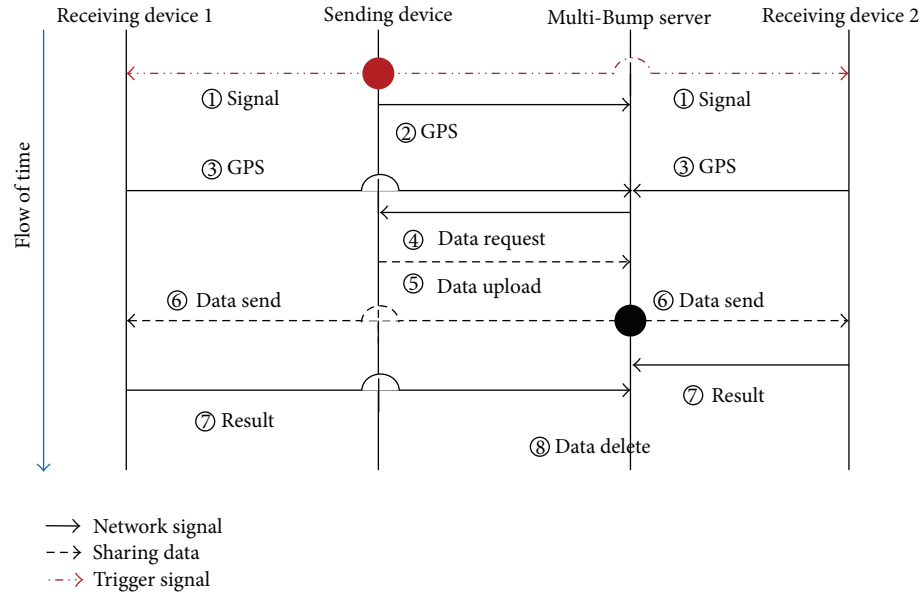


FIGURE 7: Data-sharing flow of one sending device and two receiving devices.

smart device and Galaxy S6 as the receiving smart device. Figure 9 shows the result according to the length of the trigger signal.

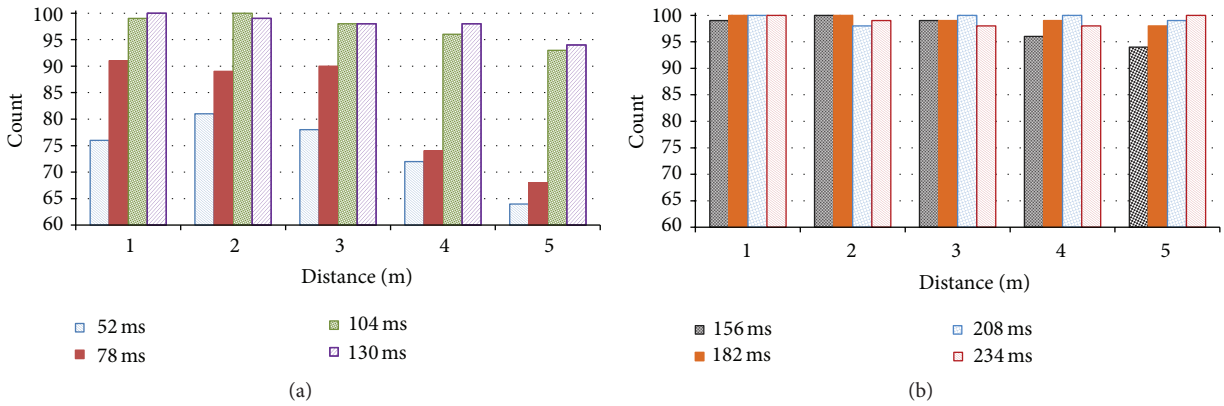
As shown in Figure 9(a), when the length of the trigger signal is short, such as 52 ms and 78 ms, the transmission accuracy is under 83%. As shown in Figures 9(a) and 9(b), when the length of the trigger signal is in the middle, such as 104, 130, and 156 ms, the transmission accuracy is over 97%. When the length of the trigger signal is long, such as 182, 208, and 234 ms, the transmission accuracy is over 99%. Thus, as the trigger signal length of 208 ms obtained the best accuracy in this test, we set the length of the trigger signal to 208 ms.

Next, to find the value of threshold α, which causes the high accuracy of the proposed method, we conducted a transmission test according to the distance between the smart devices with a changing threshold α. The experimental environment was the same as that in the previous experiment, and the length of the trigger signal was 208 ms. The changes in threshold α are 104 ms (50% of 208 ms), 125 ms (60%), 146 ms (70%), 166 ms (80%), and 187 ms (90%). Figure 10 shows the result of the transmission accuracy according to the distance between the sending smart device and the receiving smart device with a changing threshold α. As shown in Figure 10, the transmission accuracy values according to the changing threshold α are 99.2% (α: 104 ms), 99.2% (α: 125 ms), 99.8% (α: 146 ms), 97.2% (α: 166 ms), and 97.2% (α: 187 ms). When threshold α is 104 ms, 125 ms, and 146 ms, the transmission accuracy is over 99%. However, when threshold α is 166 ms and 187 ms, the transmission accuracy is under 98%. Therefore, in our experiment, we set threshold α to 146 ms.

The Multi-Bump server was made up of an Intel Core i5-4690 CPU and 8 G RAM. Linux was used as the server OS.

(a) (b)

Figure 8: Application example applied to the proposed method. (a) Screen used to launch application; (b) screen used to receive the shared data.



(a) (b)

Figure 9: Transmission accuracy result according to the length of the trigger signal. (a) Result of using time duration from 52 ms to 130 ms; (b) result of using time duration from 156 ms to 234 ms.

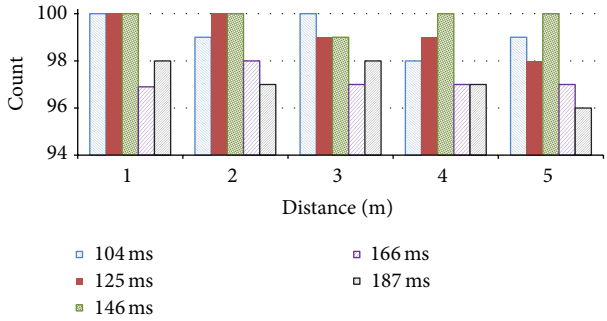


Figure 10: Result of the transmission accuracy according to the distance among smart devices with a changing threshold α.

Apache 1.3.41, PHP 5.2.6, and MySQL 5.0.51 were installed on the server. We used the POST function of PHP as the transmission method between the server and the smart devices. The table schema of the database is presented in Table 1.

As shown in Table 1, no is the autoincrement index number, and regDate is the start time of the sending device for data sharing. signalCh is the low-latency key value when the sending device and the receiving device transmit the shared data. sGPSlo and sGPSla are the GPS information of the sending device, and receiveNo is the number of receiving smart devices joining the Multi-Bump data sharing. Count is the number of receiving smart devices that start to receive the shared data, and filename is the real filename of the shared data. To evaluate the performance of the proposed method, we conducted Chung's experiment. Chung and Choo did conduct an experiment where one smart device could control another smart device using a high frequency as the trigger signal [29]. The experiment environment had four conditions:

Table 1: Table schema of the database for data transmission between smart devices.

Schema	Data value	Description
no	int(11) unsigned auto_increment	Index number
regDate	int(11) unsigned	Data transmission time
signalCh	tinyint(3)	Using the channel number of low-latency key
sGPSlo	int(10) unsigned	GPS location of the data-sending device (longitude)
sGPSla	int(10) unsigned	GPS location of the data-sending device (latitude)
receiveNo	tinyint(2)	Number of devices that will receive the data
count	tinyint(2)	Number of devices that start the receiving of data
filename	varchar(40)	Real file name to be transmitted

Table 2: Result of the performance evaluation of the proposed method according to the surrounding noise.

Environment	Distance							
	1 m	2 m	3 m	4 m	5 m	6 m	7 m	Total
Quiet indoors	100	100	100	99	98	92	91	97.1%
Noisy indoors	100	99	99	98	97	92	89	96.3%
Quiet outdoors	100	100	99	99	95	94	90	96.7%
Noisy outdoors	100	100	99	98	96	91	89	96.1%

"quiet indoors," "noisy indoors," "quiet outdoors," and "noisy outdoors." In addition, the measurement distance for the near wireless-control was from 1 to 7 m, and each distance was tested 100 times. "Quiet" is 40 dB, such as typical household noise without conversation, and "noisy" is about 60 dB, the noise level of people talking in a typical conversational tone. "Quiet indoors" was in a laboratory that remained at about 40 dB, and "noisy indoors" was based on the same location, but with the addition of a water flow sound that was played from a PC speaker and with three subjects speaking freely. "Quiet outdoors" was a peaceful garden, which was maintained at about 40 dB, and "noisy outdoors" was a bus station where many vehicles passed; it was maintained at about 60 dB. Therefore, we conducted our experiment in a space similar to Chung's experimental environment. The distance of the test was from 1 m to 7 m, and 100 attempts were made in each distance. A 2 MB photo image was used as the shared data, and 20.0 kHz (five-channel) was used as the low-latency key. Table 2 shows the experimental result of the data-sharing transmission. In Table 2, the transmission accuracy of the proposed method is similar to that of Chung's experimental result. Chung's experiment showed that the accuracy of the total average was 98.36% within 5 m and 96.09% within 7 m. The accuracy of the total average is 98.80% within 5 m and 96.57% within 7 m. Therefore, the proposed method is robust to the interference by unexpected sound signals.

Next, in Section 1, we wrote that the Bump application has not supported the Bump API and Bump server since January 31, 2014. Therefore, we developed an application and sharing server that work similarly to Bump API, and we conducted a comparative experiment on 1 : 1 and 1 : N data sharing using the proposed method and the developed application,

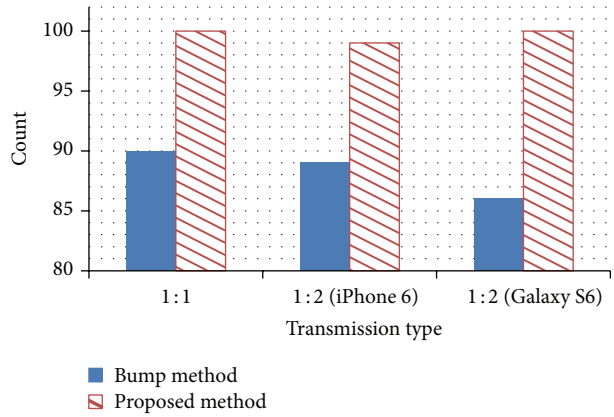


Figure 11: Experimental results of the 1 : 1 and 1 : N data sharing using the Bump application and the proposed method.

which works like Bump. The test for the 1 : 1 data sharing was performed in the quiet indoors. The distance between the sending device and the receiving device was 2 m, and 100 attempts were made using each method. We used one sending device and two receiving devices for the 1 : N experiment similar to the previous experiment. The sending device was an iPhone 6, and the receiving devices were an iPhone 6 and a Galaxy S6. The distance between the sending device and the receiving devices was 2 m. To determine the usability of the proposed method, 20 college students were asked to join the experiment as participants. The students were divided into two groups. In the 1 : 1 data-sharing test, each group made 10 attempts using each method. In the 1 : N data-sharing test, each group and the researcher made 10 attempts using each method. Figure 11 shows the result of the 1 : 1 and 1 : N data-sharing experiments.

As shown in Figure 11, the Bump application (i.e., the existing method) succeeded 90 times, and the proposed method succeeded 100 times in the 1 : 1 data-sharing experiment. In the 1 : N data-sharing experiment, the Bump method succeeded 89 times using the iPhone 6 and 86 times using the Galaxy S6 as the receiving device. The proposed method succeeded 99 times using the iPhone 6 and 100 times using the Galaxy S6. We considered that the reason why the accuracy of the Bump method was lower than that of the proposed method was that the time of the shaking action

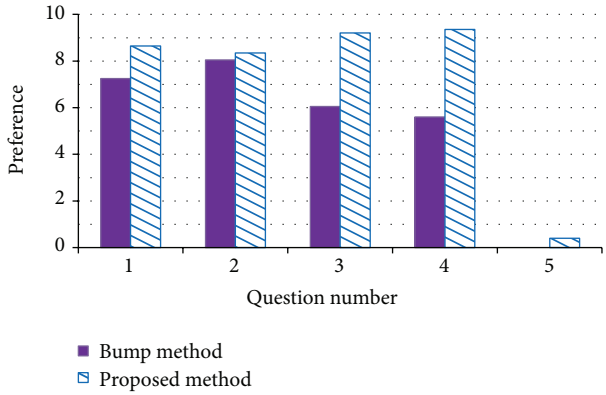


Figure 12: Survey results of the comparison between the Bump method and the proposed method.

for the data sharing sometimes did not exactly correspond. Moreover, in the 1 : N data-sharing experiment, the sending device must perform the shaking action to each receiving device separately, as the Bump method can only perform 1 : 1 dating sharing. Conversely, the proposed method has 99% high accuracy in both 1 : 1 and 1 : N data sharing. This accuracy is similar to that of the previous experiment (four environments). Therefore, the proposed method is more accurate than the existing method in the 1 : 1 and 1 : N data sharing.

Finally, we surveyed the 20 college students from the previous experiment using the following list to determine the usability of the Bump method and the proposed method. Each item is marked from one to ten points:

(1) Convenience of the 1 : 1 data sharing.

(2) Preference in the transmission time for the 1 : 1 data sharing.

(3) Convenience of the 1 : N data sharing.

(4) Preference in the transmission time for the 1 : N data sharing.

(5) Strident disadvantage of the user in the high frequency as the trigger signal.

Figure 12 shows the survey results.

As shown in Figure 12, convenience of the 1 : 1 data sharing and preference in transmission time for the data sharing both gained 7 points. However, as the Bump method obtained 6 points in convenience of the 1 : N data sharing and 5.5 points in preference in the transmission time for 1 : N data sharing, we consider the Bump method to be unsuitable for 1 : N data sharing. Conversely, the proposed method gained 9.2 points in convenience of the 1 : N data sharing and 9.35 points in preference in transmission time for the 1 : N data sharing. Therefore, we confirm that the proposed method is more suitable than the Bump method. As the points of the strident disadvantage of the user in high frequency as the trigger signal is under 0.5 points, the proposed method can be used in real life without any difficulty.

5. Conclusion

In this study, a method that uses inaudible high frequency for effectively sharing data among multi-smart devices at close range is proposed. The proposed method can solve the problems (e.g., Bluetooth pairing, different OS, and pretasks such as adding a friend) of the existing methods, and it is more accurate than the Bump method in 1 : 1 data sharing. Indeed, the proposed method can share data among multi-smart devices unlike the Bump method. Therefore, as the Multi-Bump method that uses inaudible high frequency does not need extra transmission modules, the proposed method can be applied to data sharing between a personal computer and smart devices as well as data sharing among multi-smart devices.

In our future research, we will examine direct data sharing among multi-smart devices without the Multi-Bump server by upgrading the trigger signal using inaudible high frequency. We will also study system development that can confirm people's focus on a specific space using a trigger signal and smart devices. Finally, in this paper, the sharing server did not consider security and privacy concerns related to the data stored on the Bump server. Thus, we had to handle the security and privacy problem. The developed application in this paper has a "Change Key" button that can change the high frequency of the trigger signal. If we apply the value of this button and the additional value to the application, we can use an encryption algorithm, such as the RSA method, and it should solve the security problem. Then, in Figure 3 of Section 3, when the data sharing between smart devices is concluded, the sharing server deletes the sharing data immediately. We think this function of the proposed method should solve the privacy concerns related to the data stored on the sharing server. Therefore, in our future research, we will study security and privacy concerns related to the data stored on the sharing server.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This research was supported in part by Ministry of Education, under Basic Science Research Program (NRF-2013R1A1A2061478).

References

[1] A. Iera, L. Militano, L. P. Romeo, and F. Scarcello, "Fair cost allocation in cellular-bluetooth cooperation scenarios," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2566–2576, 2011.

[2] H. Yoon and J. W. Kim, "Collaborative streaming-based media content sharing in WiFi-enabled home networks," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 4, pp. 2193–2200, 2010.

[3] Apple Inc., https://support.apple.com/en-us/HT204144.

[4] Z. Wang, Z. Xu, W. Xin, and Z. Chen, "Implementation and analysis of a practical NFC relay attack example," in *Proceedings of the 2nd International Conference on Instrumentation, Measurement, Computer, Communication and Control (IMCCC '12)*, pp. 143–146, IEEE, Harbin, China, December 2012.

[5] M. Roland, "Software card emulation in NFC-enabled mobile phones: great advantage or security nightmare," in *Proceedings of the 4th International Workshop on Security and Privacy in Spontaneous Interaction and Mobile Phone Use*, June 2012.

[6] C.-H. Huang and S.-L. Chang, "Study on the feasibility of NFC P2P communication for nursing care daily work," *Journal of Computers*, vol. 24, no. 2, pp. 33–45, 2013.

[7] S. Bouzefrane, A. F. B. Mostefa, F. Houacine, and H. Cagnon, "Cloudlets authentication in NFC-based mobile computing," in *Proceedings of the 2nd IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud '14)*, pp. 267–272, April 2014.

[8] C. Eunjeong, "Kakaotalk, a mobile social platform pioneer," *SERI Quarterly*, vol. 6, no. 1, pp. 63–69, 2013.

[9] K. Church and R. De Oliveira, "What's up with whatsapp?: comparing mobile instant messaging behaviors with traditional SMS," in *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '13)*, pp. 352–361, Munich, Germany, August 2013.

[10] J. D. Rachid, "Global communication apps," *Network Journal*, vol. 21, no. 1, p. 61, 2014.

[11] S. L. Kane-Gill, J. P. Post, P. L. Smithburger, and A. L. Seybert, "'Bump': using a mobile app to enhance learning in simulation scenarios," *Simulation in Healthcare*, vol. 7, no. 5, pp. 326–327, 2012.

[12] V. Filonenko, C. Cullen, and J. Carswell, "Indoor positioning for smartphones using asynchronous ultrasound trilateration," *ISPRS International Journal of Geo-Information*, vol. 2, no. 3, pp. 598–620, 2013.

[13] B. Thiel, K. Kloch, and P. Lukowicz, "Sound-based proximity detection with mobile phones," in *Proceedings of the 3rd International Workshop on Sensing Applications on Mobile Phones (PhoneSense '12)*, p. 4, ACM, Toronto, Canada, November 2012.

[14] G. Leventhall, "What is infrasound?" *Progress in Biophysics and Molecular Biology*, vol. 93, no. 1–3, pp. 130–137, 2007.

[15] V. Coskun, B. Ozdenizci, and K. Ok, "A survey on near field communication (NFC) technology," *Wireless Personal Communications*, vol. 71, no. 3, pp. 2259–2294, 2013.

[16] A. Kumar, A. Arora, and C. J. Islam, "Near field communication (NFC): an expertise primer," *Discovery*, vol. 2, no. 4, pp. 20–25, 2012.

[17] M. H. Eldefrawy and M. K. Khan, "Banknote validation through an embedded RFID chip and an NFC-enabled smartphon," *Mathematical Problems in Engineering*, vol. 2015, Article ID 264514, 8 pages, 2015.

[18] J. Constine, "Bump pay lets you PayPal someone with a tap, but only in-person," *TechCrunch Hot Topics*, March 2012.

[19] R. Lai, *ING Direct's iOS App Adds 'Bump' Money Transfer Feature for Well-Heeled Posses*, Engadget, May 2011.

[20] G. Tan, M. Lu, F. Jiang, K. Chen, X. Huang, and J. Wu, "Bumping: A bump-aided inertial navigation method for indoor vehicles using smartphones," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 7, pp. 1670–1680, 2014.

[21] Bump-api-ios, https://github.com/bumptech/bump-api-ios.

[22] T. Halevi, D. Ma, N. Saxena, and T. Xiang, "Secure proximity detection for NFC devices based on ambient sensor data," in *Computer Security—ESORICS 2012: 17th European Symposium on Research in Computer Security, Pisa, Italy, September 10–12, 2012. Proceedings*, vol. 7459 of *Lecture Notes in Computer Science*, pp. 379–396, Springer, Berlin, Germany, 2012.

[23] A. Studer, T. Passaro, and L. Bauer, "Don't Bump, Shake on It: the exploitation of a popular accelerometer-based smart phone exchange and its secure replacement," in *Proceedings of the 27th Annual Computer Security Applications Conference (ACSAC '11)*, pp. 333–342, Austin, Tex, USA, December 2011.

[24] J. Edwards, "Signal processing: the driving force behind smarter, safer, and more connected vehicles," *IEEE Signal Processing Magazine*, vol. 28, no. 5, pp. 8–13, 2011.

[25] P. Sarah, "With its latest update, Bump's mobile app replaces USB flash drives," Techcrunch, http://techcrunch.com/2013/02/14/with-its-latest-update-bumps-mobile-app-replaces-usb-flash-drives/.

[26] P. Bihler, P. Imhoff, and A. B. Cremers, "SmartGuide—a smartphone museum guide with ultrasound control," *Procedia Computer Science*, vol. 5, pp. 586–592, 2011.

[27] A. S. Nittala, X. D. Yang, E. Sharlin, S. Bateman, and S. Greenberg, "SonicData: broadcasting data via sound for smartphones," Research Report 2014-1064-15, PRISM: University of Calgary Digital Repository, Calgary, Canada, 2014.

[28] J. B. Kim, J. E. Song, and M. K. Lee, "Authentication of a smart phone user using audio frequency analysis," *Journal of the Korea Institute of Information Security and Cryptology*, vol. 22, no. 2, pp. 327–336, 2012.

[29] M. B. Chung and H. S. Choo, "Near wireless-control technology between smart devices using inaudible high-frequencies," *Multimedia Tools and Applications*, vol. 74, no. 15, pp. 5955–5971, 2015.

Hindawi Publishing Corporation
Mobile Information Systems
Volume 2015, Article ID 605357, 8 pages
http://dx.doi.org/10.1155/2015/605357

Research Article

Learning-Based QoS Control Algorithms for Next Generation Internet of Things

Sungwook Kim

Department of Computer Science, Sogang University, 35 Baekbeom-ro (Sinsu-dong), Mapo-gu, Seoul 121-742, Republic of Korea

Correspondence should be addressed to Sungwook Kim; swkim01@sogang.ac.kr

Received 24 July 2015; Revised 2 October 2015; Accepted 20 October 2015

Academic Editor: Yassine Hadjadj-Aoul

The Internet has become an evolving entity, growing in importance and creating new value through its expansion and added utilization. The Internet of Things (IoT) is a new concept associated with the future Internet and has recently become popular in a dynamic and global network infrastructure. However, in an IoT implementation, it is difficult to satisfy different Quality of Service (QoS) requirements and achieve rapid service composition and deployment. In this paper, we propose a new QoS control scheme for IoT systems. Based on the Markov game model, the proposed scheme can effectively allocate IoT resources while maximizing system performance. In multiagent environments, a game theory approach can provide an effective decision-making framework for resource allocation problems. To verify the results of our study, we perform a simulation and confirm that the proposed scheme can achieve considerably improved system performance compared to existing schemes.

1. Introduction

In the past forty years, the Internet has grown into a network that connects an estimated 1.8 billion users and has attained a global penetration rate of almost 25%. Telecommunications and the Internet are forming an increasingly integrated system for processing, storing, accessing, and distributing information and managing content. This convergence is based on the rapid evolution of digital technology and the diffusion of the concept of the Internet. In recent years, the steps of penetration of digital technologies, the evolution towards integrated telecommunications, information technology, and the electronic media sector have been actively presented. Developments in many different technologies are creating a significant, innovative, technical potential for the production, distribution, and consumption of information services [1, 2].

In 1999, Ashton first presented the concept of the Internet of Things (IoT) [2], a technological revolution that promotes a new ubiquitous connectivity, computing, and communication era. The IoT is a vision wherein the Internet extends into our everyday lives through a wireless network of uniquely identifiable objects. Therefore, the development of the IoT depends on dynamic technical innovations in a number of fields, including wireless sensors and nanotechnology [3]. Furthermore, the IoT service infrastructure is expected to promptly evaluate the Quality of Services (QoS) and provide satisfying services by considering things such as the preferences of users' device capability and current network status. However, the definition of QoS in the IoT is not clear because it has been poorly studied. To adaptively manage an IoT system, a new QoS control model is necessary. This model must be able to balance network availability with information accuracy in delivering data [4–8].

A fundamental challenge of QoS management is that relatively scarce network resources must be selected and allocated in a prudent manner to maximize system performance [7, 8]. To adaptively allocate network resources, game theory has been widely applied in mission-critical network management problems. Typically, game theory is used to study strategic situations where players choose different actions in an attempt to maximize their payoffs, depending upon the choices of other individuals. Therefore, game theory provides a framework for modeling and analyzing various interactions between intelligent and rational game players in conflict situations [6].

In traditional game models, it is important to define equilibrium strategies as game solutions. Equilibrium strategies are assumed to be the optimal reaction to others, given full knowledge and observability of the payoffs and actions of the other players. Therefore, most equilibrium concepts require that the payoffs and strategies of the other players be known in advance and observed by all players. However, this is a strong assumption that is not the case in the majority of real-life problems. Players in actual situations have only partial knowledge, or no knowledge at all, regarding their environments and the other players evolving around them [6]. To alleviate this difficulty, van der Wal developed the Markov game model [9]. This approach relaxes the strict game model assumptions by implementing learning algorithms. Through repeated plays, Markov game players effectively consider their current payoffs and a history of observations regarding the strategies of the other players [9, 10].

The main purpose of this paper is to develop an effective QoS control scheme for IoT systems. Based on the Markov game model, we build an intelligent decision-making process that addresses the critical QoS problem of an IoT system. With a real-time learning feedback mechanism, the proposed scheme adapts well to the dynamic requirements of IoT applications. Through online-oriented strategic decisions, the proposed scheme attempts to attain a self-confirming equilibrium, the new solution concept for real-time network systems.

1.1. Related Work. To improve IoT system performance, several QoS control schemes have been proposed to efficiently and integrally allocate IoT resources. The Time-Controlled Resource Sharing (TCRS) scheme [11] is a scheduling scheme that shares resources between Machine-to-Machine (M2M) and Human-to-Human (H2H) communication traffic services. This scheme analytically focuses solely on resource utilization and the QoS of the M2M and H2H traffic and derives expressions for blocking probabilities of the M2M and H2H traffic and percentage resource utilization [11].

The *IoT Service Selection* (*IoTSS*) scheme [12] is a model to select the appropriate service, from many services, that satisfies a user's requirements. This scheme considers three core concepts, device, resource, and service, while specifying their relationships. To dynamically aggregate individual QoS ratings and select physical services, the *IoTSS* scheme designs a Physical Service Selection (PSS) method that considers a user preference and an absolute dominance relationship among the physical services.

The *Approximate Dynamic Programming based Prediction* (*ADPP*) scheme [13] is a novel evaluation approach employing prediction strategies to obtain accurate QoS values. Unlike the traditional QoS prediction approaches, the *ADPP* scheme is realized by incorporating an approximate dynamic programming based online parameter tuning strategy into the QoS prediction approach. The *Services-oriented QoS-aware Scheduling* (*SQoSS*) scheme [5] is a layered QoS scheduling scheme for service-oriented IoT. The *SQoSS* scheme explores optimal QoS-aware service composition using the knowledge of each component service. This scheme can effectively operate the scheduling problem in heterogeneous network environments. The main goal of the *SQoSS* scheme is to optimize the scheduling performance of the IoT network while minimizing the resource costs [5].

The *Intelligent Decision-Making Service* (*IDMS*) scheme [4] constructs a context-oriented QoS model according to the Analytical Hierarchy Process (AHP). Using this hierarchical clustering algorithm, the *IDMS* scheme can effect intelligent decisions while fully considering the users' feedback. The earlier study has attracted significant attention and introduced unique challenges to efficiently solve the QoS control problem. Compared to these schemes [4, 5, 13], the proposed scheme attains improved performance during the IoT system operations.

The remainder of this paper is organized as follows. The proposed game model is formulated in Section 2, where we introduce a Markov decision process to solve the QoS problem and explain the proposed IoT resource allocation algorithm in detail. In Section 3, we verify the effectiveness and efficiency of the proposed scheme from simulation results. We draw conclusions in Section 4.

2. Proposed QoS Control Algorithms for IoT Systems

In this section, we describe the proposed algorithm in detail. The algorithm implements a game theory technique and appears to be a natural approach to the QoS control problem. Employing a Markov game process, we can effectively model the uncertainties in the current system environment. The proposed algorithm significantly improves the success rate of the IoT services.

2.1. Markov Game Model for IoT Systems. Network services are operated based on the Open Systems Interconnection model (OSI Model). In this study, we design the proposed scheme using a three-layered (i.e., application, network, and sensing layers) QoS architecture. At the application layer, an application is selected to establish a connection and decisions are made by the user and the QoS scheduling engine. In general, the QoS module must allocate network resources to the services that are selected in the application layer [5]. At the network layer, the QoS module must allocate network resources to the selected services. The decision-making process at this layer may involve QoS attributes that are used in traditional QoS mechanisms over networks [5]. At the sensing layer, the decision-making process involves the selection of a basic sensing infrastructure based on sensing ability and the required QoS for applications. The QoS module at the sensing layer is responsible for the selection of the basic sensing devices [5].

In this study, we investigate learning algorithms using uncertain, dynamic, and incomplete information and develop a new adaptive QoS scheduling algorithm that has an intelligent decision-making process useful in IoT systems. For the interactive decisions of the IoT system agents, we formulate a multiple decision-making process using a game model while studying a multiagent learning approach. Using this

technique, the proposed scheme can effectively improve the QoS in IoT systems.

Learning is defined as the capability of making intelligent decisions by self-adapting to the dynamics of the environment, considering experience gained in the past and present system states, and using long-term benefit estimations. This approach can be viewed as self-play, where either a single player or a population of players evolves during competitions on a repeated game. During the operation of an IoT system, learning is driven by the amount of information available from every QoS scheduler [14]. As indicated in the traditional methods, complete information significantly improves performance with respect to partial observability; however, the control overhead results in a lack of practical implementations. Consequently, a tradeoff must be made considering that the capability to make autonomous decisions is a desirable property of self-organized IoT systems [5, 14].

The Markov decision-making process is a well-established mathematical framework for solving sequential decision problems using probabilities. It models a decision-making system where an action must be taken in each state. Each action may have different probabilistic outcomes that change the system's state. The goal of the Markov decision process is to determine a policy that dictates the best action to take in each state. By adopting the learning Markov game approach, the proposed model allows distributed QoS schedulers to learn the optimal strategy, one step at a time. Within each step, the repeated game strategy is applied to ensure cooperation among the QoS schedulers. The well-known Markov decision process can be extended in a straightforward manner to create multiplayer Markov games. In a Markov game, actions are the result of the joint action selection of all players and payoffs, and state transitions depend on these joint actions. Therefore, payoffs are sensed for combinations of actions taken by different players and players learn in a product or joint action space. From the obtained data, players can adapt to changing environments, improve performance based on their experience, and make progress in understanding fundamental issues [5, 9, 10].

In the proposed QoS control algorithm, the game model is defined as a tuple $\langle \mathbf{S}, N, \mathbb{A}_{k,1\le k\le N}, U_{k,1\le k\le N}, \mathcal{T}\rangle$, where $\mathbf{S}$ is the set of all possible states and N is the number of players. In the proposed model, each state is the resource allocation status in the IoT system. $\mathbb{A}_{k,1\le k\le N} = \{a_1, a_2, \ldots, a_m\}$ is the collection of strategies for player k, where m is the number of possible strategies. Actions are the joint result of multiple players choosing a strategy individually. In the proposed Markov game, QoS schedulers are assumed as game players and the collection of strategies for each player is the set of availabilities of system resources. $U_{k,1\le k\le N} : \mathbf{S} \times \mathbb{A}_1 \times \mathbb{A}_2 \times \cdots \times \mathbb{A}_N \to \mathfrak{N}$ is the utility function, where $\mathfrak{N}$ represents the set of real numbers. $\mathcal{T} : \mathbf{S} \times \mathbb{A}_1 \times \mathbb{A}_2 \times \cdots \times \mathbb{A}_N \to \Delta(\mathbf{S})$ is the state transition function, where $\Delta(\mathbf{S})$ is the set of discrete probability distributions over the set $\mathbf{S}$. Therefore, $\mathcal{T}(s^t, a_1, a_2, \ldots, a_N, s^{t+1})$ is the probability of arriving in state s^{t+1} when each agent takes an action a_i at state s^t, where $s^t, s^{t+1} \in \mathbf{S}$ [5, 9, 10].

In the developed game model, players seek to choose their strategy independently and self-interestedly to maximize their payoffs. Each strategy represents an amount of system resource and the utility function measures the outcome of this decision. Therefore, different players can receive different payoffs for the same state transition. By considering the allocated resource amount, delay, and price, the utility function (U) of each player is defined as follows:

$$U(x) = \omega \exp\left(\left(\frac{\mathrm{T}(x)}{\overline{\mathbb{T}}}\right)^{0.5-\epsilon} (\tau_M - \tau)^{0.5+\epsilon}\right) - c(x, \xi), \quad \text{s.t., } \epsilon \in \{-0.5; 0.5\}, \tag{1}$$

where ω represents the player's willingness to pay for his perceived service worth. $\overline{\mathbb{T}}$ is the system's average throughput and $\mathrm{T}(x)$ is the player's current throughput with the allocated resource x; this is the rate of successful data delivery over a communication channel. τ_M and τ are the maximum delay and the observed delay of the application services, respectively. τ is measured from real network operations. In a real-time online manner, each QoS scheduler actually measures $\overline{\mathbb{T}}$, $\mathrm{T}(x)$, and τ. $c(x, \xi)$ is the cost function and ξ is the price for a resource unit. τ is obtained according to the processing and arrival service rates. In a distributed self-regarding fashion, each player (i.e., QoS scheduler) is independently interested in the sole goal of maximizing his utility function as follows:

$$\max_x U(x), \quad \text{where } c(x, \xi) = \left(\frac{x\xi}{\overline{x}_A \xi}\right)^q, \tag{2}$$

where x is the allocated resource in its own QoS scheduler, $\overline{x}_A$ is the average resource amount of all QoS schedulers, and q is a cost parameter for the cost function $c(x, \xi)$. The cost function is defined as the ratio of its own obtained resource to the average resource amount of all the QoS schedulers. Therefore, other players' decisions are returned to each player. This iterative feedback procedure continues under IoT system dynamics. In this study, QoS schedulers can modify their actions in an effort to maximize their $U(x)$ in a distributed manner. This approach can significantly reduce the computational complexity and control overheads. Therefore, it is practical and suitable for real world system implementation.

2.2. Markov Decision Process for QoS Control Problems. In this work, we study the method that a player (i.e., QoS scheduler) in a dynamic IoT system uses to learn an uncertain network situation and arrives at a control decision by considering the online feedback mechanism. With an iterative learning process, the players' decision-making mechanism is developed as a Markov game model, which is an effective method for the players' decision mechanism. If players change their strategies, the system state may change. Based on the immediate payoff ($U(S_0, a_i(0))$) of the current state S_0 and action $a_i(0)$, players must consider the future payoffs. With the current payoff, player i's long-term expected payoff ($V_i(S_0, a_i(0))$) is given by [5]

$$V_i\left(S_0, a_i\left(0\right)\right) = \max_{a_i(t), 0 \le t \le \infty} \left[U_i\left(S_0, a_i\left(0\right)\right) + \sum_{t=1}^{\infty}\left(\beta^t U_i\left(S_t, a_i\left(t\right)\right)\right)\right], \quad \text{s.t.,} \quad a_i\left(t\right) \in \mathbb{A}_i, \tag{3}$$

where $a_i(t)$ and $U_i(S_t, a_i(t))$ are player i's action and expected payoff at time t, respectively. β is a discount factor for the future state. During game operations, each combination of starting state, action choice, and next state has an associated transition probability. Based on the transition probability, (3) can be rewritten by the recursive Bellman equation form given in [5]

$$V_i\left(S\right) = \max_{a_S} \left[U_i\left(Sa_S\right) + \gamma \sum_{S' \in \mathbf{S}} \left\{P^i\left(S' \mid S, a_S\right) V_i\left(S'\right)\right\}\right], \quad \text{s.t.,} \quad a_S \in \mathbb{A}_i, \tag{4}$$

where S' represents all possible next states of S and γ can be regarded as the probability that the player remains at the selected strategy. $P^i(S'|S, a_S)$ is the state transition probability from state S to the state S'; S and S' are elements of system state set $\mathbf{S}$. In this study, N is the number of QoS schedulers, and m is the number of possible strategies for each scheduler. Therefore, there are total m^N system states.

$P(S' \mid S, a_S)$ is a distributed multiplayer probability decision problem. Using the multiplayer-learning algorithm, each player independently learns the current IoT system situation to dynamically determine $P(S' \mid S, a_S)$. This approach can effectively control a Markov game process with unknown transition probabilities and payoffs. In the proposed algorithm, each player is assumed to be interconnected by allowing them to play in a repeated game with the same environment. Assume there is a finite set of strategies $\mathbb{A}_{1 \le k \le N}(t) = \{a_1^k(t), \ldots, a_m^k(t)\}$ chosen by player k at game iteration t; m is the number of possible strategies. Correspondingly, $\mathbf{U}^k(t) = (u_1^k(t), \ldots, u_m^k(t))$ is a vector of specified payoffs for player k. If player k plays action $a_{l,1 \le l \le m}^k$, he earns a payoff $u_{l,1 \le l \le m}^k$ with probability p_l^k. $\mathbf{P}^k(t) = \{p_1^k(t), \ldots, p_m^k(t)\}$ is defined as player k's probability distribution.

Actions chosen by the players are input to the environment and the environmental response to these actions serves as input to each player. Therefore, multiple players are connected in a feedback loop with the environment. When a player selects an action with his respective probability distribution $\mathbf{P}(\cdot)$, the environment produces a payoff $U(\cdot)$ according to (1). Therefore, $\mathbf{P}(\cdot)$ must be adjusted adaptively to contend with the payoff fluctuation. At every game round, all players update their probability distributions based on the online responses of the environment. If player k chooses a_l^k at time t, this player updates $\mathbf{P}^k(t+1)$ as follows:

$$p_j^k\left(t+1\right) = \begin{cases} f\left(p_j^k\left(t\right) + \psi\left[\dfrac{u_l^k\left(t\right) - u_l^k\left(t-1\right)}{u_l^k\left(t-1\right)}\right]\right), & \text{if } j = l \\ \varphi p_j^k\left(t\right), & \text{if } j \ne l, \end{cases} \tag{5}$$

$$\text{s.t.,} \begin{cases} f\left(\chi\right) = 0, & \text{if } \chi < 0 \\ f\left(\chi\right) = \chi, & \text{if } 0 < \chi < 1 \\ f\left(\chi\right) = 1, & \text{if } \chi > 1, \end{cases}$$

where φ is a discount factor and ψ is a parameter to control the learning size from $p(t)$ to $p(t+1)$. In general, small values of ψ correspond to slower rates of convergence, and vice versa. According to (5), $P^k(S' \mid S, a_S)$ is defined based on the Boltzmann distribution:

$$P^k\left(S' \mid S, a_S\right) = \frac{\exp\left(\left(1/\lambda\right) p_{a_S}^k\left(t\right)\right)}{\sum_{j \in \mathbb{A}_k} \exp\left(\left(1/\lambda\right) p_j^k\left(t\right)\right)}, \quad \text{s.t.,} \; a_S \in \mathbb{A}_k\left(t\right) = \left\{a_1^k\left(t\right), \ldots, a_m^k\left(t\right)\right\}, \tag{6}$$

where λ is a control parameter. Strategies are chosen in proportion to their payoffs; however, their relative probability is adjusted by λ. A value of λ close to zero allows minimal randomization and a large value of λ results in complete randomization.

2.3. The Main Steps of Proposed Scheme. To allow optimal movement in multischeduler systems, we consider the consequences of using the Markov game model by implementing the adaptive learning algorithm that attempts to learn an optimal action based on past actions and environmental feedback. Although there are learning algorithms to construct a game model, minimal research has been conducted on integrating learning algorithms with the decision-making process where players are uncertain regarding the real world and the influence of their decisions on each other.

In the proposed learning-based Markov decision process, a single QoS scheduler interacts with an environment defined by a probabilistic transition function. From the result of the individual learning experiences, each scheduler can learn how to effectively play under the dynamic network situations. As the proposed learning algorithm proceeds and the various

Table 1: System parameters used in the simulation experiments.

Traffic class	Message application	Bandwidth requirement	Connection duration average/sec
I	Delay-critical emergency applications	32 Kbps	30 sec (0.5 min)
II	Event-related applications	32 Kbps	120 sec (2 min)
		64 Kbps	180 sec (3 min)
III	General applications	128 Kbps	120 sec (2 min)
		256 Kbps	180 sec (3 min)
IV	Multimedia applications	384 Kbps	300 sec (5 min)
		512 Kbps	120 sec (2 min)

Parameter	Value	Description
ω	1.2	The player's willingness to pay for his perceived service worth
ϵ	−0.2	The control parameter between throughput and delay
q	1.1	The estimation parameters of the cost function
γ	0.3	A probability that the user keeps staying at the selected strategy
Δ	1	Predefined minimum bound for stable status
ξ	1	The price for resource unit in the cost function
m	3	The number of strategies for QoS schedulers
ψ	1	A parameter to control the learning size
φ	0.8	A discount factor for the respective probability distribution
λ	1	A control parameter on the Boltzmann distribution

```
Init ( )
{   1: p(·) = 1/m
    2: Control parameter values (ω, ϵ, q, γ, p, Δ, ψ, φ and λ) are given from Table 1.
}
Main_QoS control ( )
{   Start: Init ( );
    For (;; ) {
       3: U(x) is obtained from (1) and (2);
       4: p(·) is adjusted by using (5);
       5: P(S′ | S, a_S) is defined by using (6);
       6: a(t) is selected to maximize V(·) based on (4).
       7: IF (‖V^(t+1)(·) − V^(t)(·)‖ < Δ), Temp ( );
         ELSE continue;
       }
}
Temp ( )
{ 8: For (;; ) { IF (‖V^(t+1)(·) − V^(t)(·)‖ < Δ), continue; ELSE break; }
}
```

Pseudocode 1: IoT system QoS control procedure.

actions are tested, the QoS scheduler acquires increasingly more information. That is, the payoff estimation at each game iteration can be used to update $P(S' \mid S, a_S)$ in such a manner that those actions with a large payoff are more likely to be chosen again in the next iteration. To maximize their expected payoffs, QoS schedulers adaptively modify their current strategies. This adjustment process is sequentially repeated until the change of expected payoff ($V(\cdot)$) is within a predefined minimum bound (Δ). When no further strategy modifications are made by all the QoS schedulers, the IoT system has attained a stable status. The proposed algorithm for this approach is described by Pseudocode 1 and the following steps.

Step 1. To begin, $p(\cdot)$ is set to be equally distributed ($p(\cdot) = 1/m$, where m is the number of strategies). This starting guess guarantees that each strategy enjoys the same selection probability at the start of the game.

Step 2. Control parameters ω, ϵ, q, γ, p, Δ, ψ, φ, and λ are provided to each QoS scheduler from the simulation scenario (refer to Table 1).

Step 3. Based on the current IoT situation, each QoS scheduler estimates his utility function ($U(x)$) according to (1) and (2).

Step 4. Using (5), each QoS scheduler periodically adjusts the $p(\cdot)$ values.

Step 5. Based on the probability distribution $\mathbf{P}(\cdot)$, each $P(S' \mid S, a_S)$ is defined using the Boltzmann distribution.

Step 6. Iteratively, each QoS scheduler selects a strategy ($a(t)$) to maximize his long-term expected payoff ($V(\cdot)$). This sequential learning process is repeatedly executed in a distributed manner.

Step 7. If a QoS scheduler attains a stable status (i.e., $\|V^{(t+1)}(\cdot) - V^{(t)}(\cdot)\| < \Delta$), this scheduler is assumed to have obtained an equilibrium strategy. When all QoS schedulers achieve a stable status, the game process is temporarily stopped.

Step 8. Each QoS scheduler continuously self-monitors the current IoT situation and proceeds to Step 3 for the next iteration.

3. Performance Evaluation

In this section, we compare the performance of the proposed scheme with other existing schemes [4, 5, 13] and confirm the performance superiority of the proposed approach using a simulation model. Our simulation model is a representation of an IoT system that includes system entities and the behavior and interactions of these entities. To facilitate the development and implementation of our simulator, Table 1 lists the system parameters.

Our simulation results were achieved using MATLAB, which is widely used in academic and research institutions in addition to industrial enterprises. To emulate a real world scenario, the assumptions of our simulation environment were as follows.

(i) The simulated system consisted of four QoS schedulers for the IoT system.

(ii) In each scheduler coverage area, a new service request was Poisson with rate ρ (services/s) and the range of the offered service load was varied from 0 to 3.0.

(iii) There were three strategies (m) for the QoS schedulers and each strategy ($a_{i,1\le i\le m}$) was $a_i \in \{25\text{ Mbps}, 30\text{ Mbps}, 35\text{ Mbps}\}$. Therefore, there were total m^N, that is, 3^4, system states like $\mathbf{S} = \{(25\text{ Mbps}, 25\text{ Mbps}, 25\text{ Mbps}, 25\text{ Mbps}), \ldots, (35\text{ Mbps}, 35\text{ Mbps}, 35\text{ Mbps}, 35\text{ Mbps})\}$.

(iv) The resources of the IoT system, bandwidth (bps), and total resource amount were 140 Mbps.

(v) Network performance measures obtained based on 50 simulation runs were plotted as a function of the offered traffic load.

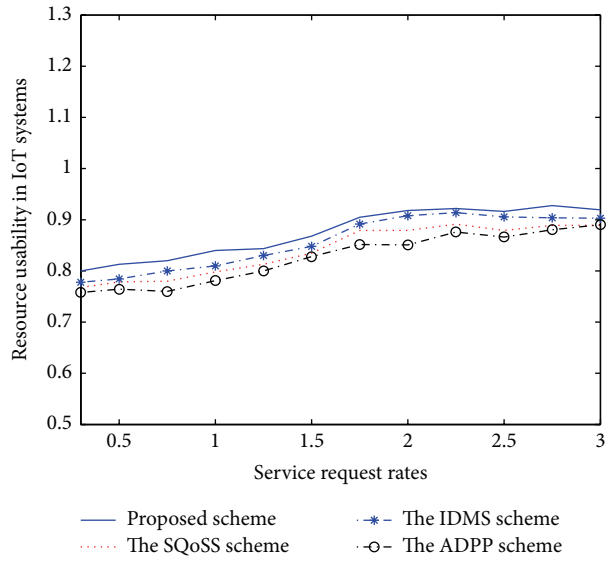


Figure 1: Resource usability in IoT systems.

(vi) The message size of each application was exponentially distributed with different means for different message applications.

(vii) For simplicity, we assumed the absence of physical obstacles in the experiments.

(viii) The performance criteria obtained through simulation were resource usability, service availability, and normalized service delay.

(ix) Resource usability was defined as the percentage of the actually used resource.

(x) Service availability was the success ratio of the service requests.

(xi) The normalized service delay was a normalized service delay measured from real network operations.

In this paper, we compared the performance of the proposed scheme with existing schemes: *SQoSS* [5], *IDMS* [4], and *ADPP* [13]. These existing schemes were recently developed as effective IoT management algorithms.

Figure 1 presents the performance comparison of each scheme in terms of resource usability in the IoT systems. In this study, resource usability is a measure of how system resources are used. Traditionally, monitoring how resources are used is one of the most critical aspects of IoT management. During the system operations, all schemes produced similar resource usability. However, the proposed scheme adaptively allocates resources to the IoT system in an incremental manner while ensuring different requirements. Therefore, the resource usability produced by the proposed scheme was higher than the other schemes from low to heavy service load intensities.

Figure 2 represents the service availability of each IoT control scheme. In this study, service availability is defined as the success ratio of the service requests. In general, excellent service availability is a highly desirable property for real

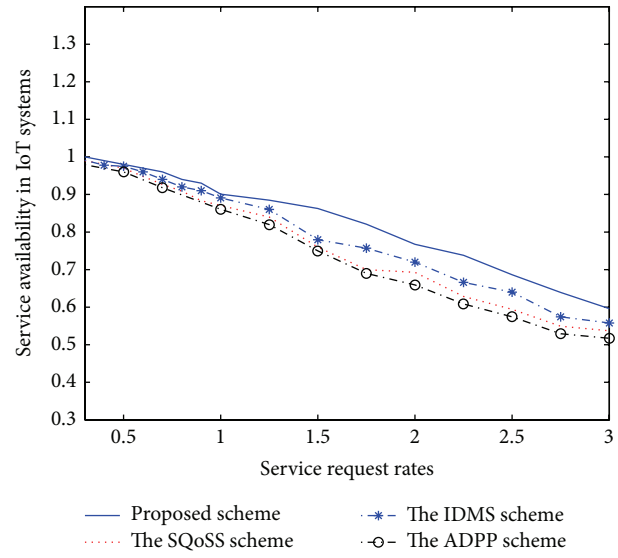


Figure 2: Service availability in IoT systems.

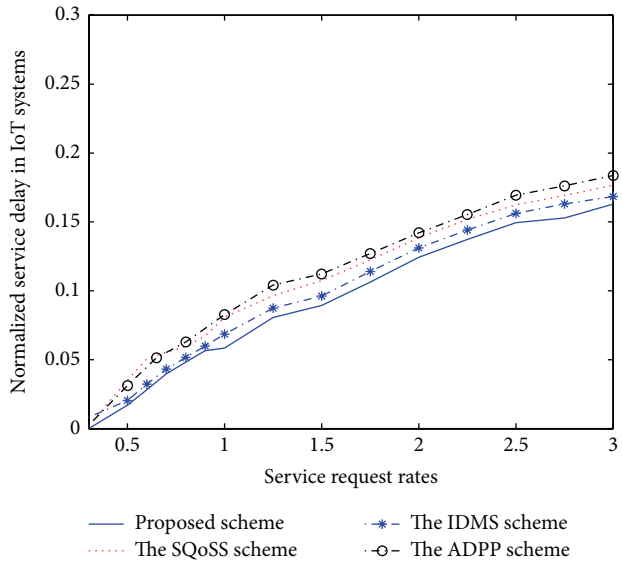


Figure 3: Normalized service delay in IoT systems.

world IoT operations. As indicated in the results, it is clear that performance trends are similar. As the service request rate increases, it can saturate or exceed the system capacity. Therefore, excessive service requests may lead to system congestion, decreasing the service availability. This is intuitively correct. Under various application service requests, the proposed game-based approach can provide a higher traffic service than the other schemes. From the above results, we conclude that the proposed scheme can provide a higher service availability in IoT systems.

The curves in Figure 3 illustrate the normalized service delay for IoT services under different service loads. Typically, service delay is an important QoS metric and can reveal the fitness or unfitness of system protocols for different delay-sensitive applications. Owing to the feedback-based Markov game approach, the proposed scheme can dynamically adapt the current situation and has significantly lower service delay than the other schemes. From the results, we can observe that the proposed approach can support delay-sensitive applications and ensure a latency reduction in IoT services.

The simulation results presented in Figures 1–3 demonstrate the performance of the proposed and other existing schemes and verify that the proposed Markov game-based scheme can provide attractive network performance. The main features of the proposed scheme are as follows: (i) a new Markov game model based on a distributed learning approach is established, (ii) each QoS scheduler learns the uncertain system state according to local information, (iii) schedulers make decisions to maximize their own expected payoff by considering network dynamics, and (iv) when selecting a strategy, schedulers consider not only the immediate payoff but also the subsequent decisions. The proposed scheme constantly monitors the current network conditions for an adaptive IoT system management and successfully exhibits excellent performance to approximate the optimized performance. As expected, the performance enhancements provided by the proposed scheme outperformed the existing schemes [4, 5, 13].

4. Summary and Conclusions

Today, IoT-based services and applications are becoming an integral part of our everyday life. It is foreseeable that the IoT will be a part of the future Internet where "things" can be wirelessly organized as a global network that can provide dynamic services for applications and users. Therefore, IoT technology can bridge the gap between the virtual network and the "real things" world. Innovative uses of IoT techniques on the Internet will not only provide benefits to users to access wide ranges of data sources but also generate challenges in accessing heterogeneous application data, especially in the dynamic environment of real-time IoT systems.

This paper addressed a QoS control algorithm for IoT systems. Using the learning-based Markov game model, QoS schedulers iteratively observed the current situation and repeatedly modified their strategies to effectively manage system resources. Using a step-by-step feedback process, the proposed scheme effectively approximated the optimized system performance in an entirely distributed manner. The most important novelties of the proposed scheme are its adaptability and responsiveness to current system conditions. Compared with the existing schemes, the simulation results confirmed that the proposed game-based approach could improve the performance under dynamically changing IoT system environments whereas other existing schemes could not offer such an attractive performance. Resource usability, service availability in IoT systems, normalized service delay, and accuracy were improved by approximately 5%, 10%, 10%, and 5%, respectively, compared to the existing schemes.

Furthermore, our study opens the door to several interesting extensions. In the future, we plan to design new reinforcement-learning models and develop adaptive online feedback algorithms. This is a potential direction and possible

extension to this study and can further improve the performance of IoT systems. Moreover, it would be interesting to extend the Markov game model to various decision-theoretic frameworks. Under uncertain system environments, this would be an interesting topic for future research.

Conflict of Interests

The author, Sungwook Kim, declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) Support Program (IITP-2015-H8501-15-1018) supervised by the IITP (Institute for Information & Communications Technology Promotion) and by the Sogang University Research Grant of 2014 (201410020.01).

References

[1] G. Sallai, "Chapters of future internet research," in *Proceedings of the 4th IEEE International Conference on Cognitive Infocommunications (CogInfoCom '13)*, pp. 161–166, IEEE, Budapest, Hungary, December 2013.

[2] K. Ashton, "That 'Internet of Things' thing in the real world, things matter more than ideas," *RFID Journal*, 2009, http://www.rfidjournal.com/article/print/4986.

[3] Q. Wu, G. Ding, Y. Xu et al., "Cognitive internet of things: a new paradigm beyond connection," *IEEE Internet of Things Journal*, vol. 1, no. 2, pp. 129–143, 2014.

[4] Q. Zhang and D. Peng, "Intelligent decision-making service framework based on QoS model in the internet of things," in *Proceedings of the 11th International Symposium on Distributed Computing and Applications to Business, Engineering and Science (DCABES '12)*, pp. 103–107, Guilin, China, October 2012.

[5] L. Li, S. Li, and S. Zhao, "QoS-Aware scheduling of services-oriented internet of things," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 2, pp. 1497–1507, 2014.

[6] S. Kim, "Adaptive ad-hoc network routing scheme by using incentive-based model," *Ad Hoc & Sensor Wireless Networks*, vol. 15, no. 2, pp. 107–125, 2012.

[7] G. Pujolle, "Metamorphic networks," *Journal of Computing Science and Engineering*, vol. 7, no. 3, pp. 198–203, 2013.

[8] I. Jang, D. Pyeon, S. Kim, and H. Yoon, "A survey on communication protocols for wireless sensor networks," *Journal of Computing Science and Engineering*, vol. 7, no. 4, pp. 231–241, 2013.

[9] J. van der Wal, "Discounted Markov games: successive approximation and stopping times," *International Journal of Game Theory*, vol. 6, no. 1, pp. 11–22, 1977.

[10] P. Vrancx, K. Verbeeck, and A. Nowe, "Decentralized learning in Markov games," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 38, no. 4, pp. 976–981, 2008.

[11] K. Edemacu and T. Bulega, "Resource sharing between M2M and H2H traffic under time-controlled scheduling scheme in LTE networks," in *Proceedings of the 8th International Conference on Telecommunication Systems Services and Applications (TSSA '14)*, pp. 1–6, Kuta, Indonesia, October 2014.

[12] X. Jin, S. Chun, J. Jung, and K.-H. Lee, "IoT service selection based on physical service model and absolute dominance relationship," in *Proceedings of the 7th IEEE International Conference on Service-Oriented Computing and Applications (SOCA '14)*, pp. 65–72, Matsue, Japan, November 2014.

[13] X. Luo, H. Luo, and X. Chang, "Online optimization of collaborative web service QoS prediction based on approximate dynamic programming," in *Proceedings of the International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI '14)*, pp. 80–83, IEEE, Beijing, China, October 2014.

[14] A. Imran, M. Bennis, and L. Giupponi, "Use of learning, game theory and optimization as biomimetic approaches for Self-Organization in macro-femtocell coexistence," in *Proceedings of the IEEE Wireless Communications and Networking Conference Workshops (WCNCW '12)*, pp. 103–108, IEEE, Paris, France, April 2012.

Hindawi Publishing Corporation
Mobile Information Systems
Volume 2015, Article ID 942858, 10 pages
http://dx.doi.org/10.1155/2015/942858

Research Article

Efficient DFSA Algorithm in RFID Systems for the Internet of Things

Hsing-Wen Wang

Department of Business Administration, National Changhua University of Education, Changhua 500, Taiwan

Correspondence should be addressed to Hsing-Wen Wang; wenwang329@gmail.com

Received 5 August 2015; Accepted 5 October 2015

Academic Editor: Jong-Hyouk Lee

Radio Frequency IDentification (RFID) used in business applications and international business management fields can create and sustain the competitive advantage, which is also one of the wireless telecommunication techniques for recognizing objects to realize Internet of Things (IoT) technologies. In construction of IoT network, the RFID technologies play the role of the front-end data collection via tag identification, as the basis of IoT. Hence, the adoption of RFID technologies is spurring innovation and the development of the IoT. However, in RFID system, one of the most important challenges is the collision resolution between the tags when these tags transmit their data to the reader simultaneously. Hence, in this paper I develop an efficient scheme to estimate the number of unidentified tags for Dynamic Framed Slotted Aloha (DFSA) based RFID system, with the view of increasing system performance. In addition to theoretical analysis, simulations are conducted to evaluate the performance of proposed scheme. The simulation results reveal the proposed scheme works very well in providing a substantial performance improvement in RFID system. The proposed algorithm promotes business effectiveness and efficiency while applying the RFID technologies to IoT.

1. Introduction

The Internet of Things (IoT) enables innovative business opportunities by connecting new era technology and objects to create new applications in business management and enterprise workflows [1–3]. IoT allows "things (objects)" to be sensed and controlled remotely across the Internet, creating opportunities for more direct integration between the real and cyber world and resulting in increased efficiency, accuracy, and economic benefit. For IoT, each object is uniquely identifiable through its embedded computing system but is able to interoperate within the network infrastructure. In general, the IoT requires a few necessary components to enable communication between devices and objects, and objects need to be augmented with an autoidentified technology, usually a RFID tag, so that the object can be uniquely identified. In other words, the IoT refer to embedding the devices such as RFID tag on the objects, and then through wireless communication technology the objects can automatically communicate the information to each other to achieve intelligent identification and management of the objects. Hence, RFID [4–6] plays a very important role in IoT industries and also makes IoT come to people's life or lives.

Because of the success of IoT, nowadays, RFID is one of the most popular wireless communication technologies applied in short range wireless transmission, and the RFID techniques have been successfully applied in a variety of areas, including the logistics and supply chains management, assets management, and inspection and safety compliance management. The advantages of RFID technology can make it have extensively more and more applications in various business areas [7], such as the well-known brand enterprises in the world: Zara, Prada, and Amazon [8, 9]. A RFID system, as illustrated in Figure 1, consists of a reader and a number of tags. At first the reader sends a request to ask the tags to send their data to the reader. After receiving the request from the reader, the tags will send their data to the reader. Each tag will randomly select a slot in the contending frame to send out its data. If there is only one tag sending its data in an empty slot, then data will be successfully received by the reader. However, sometimes collisions occur when there are more than one tag transmitting its data to the reader, and this

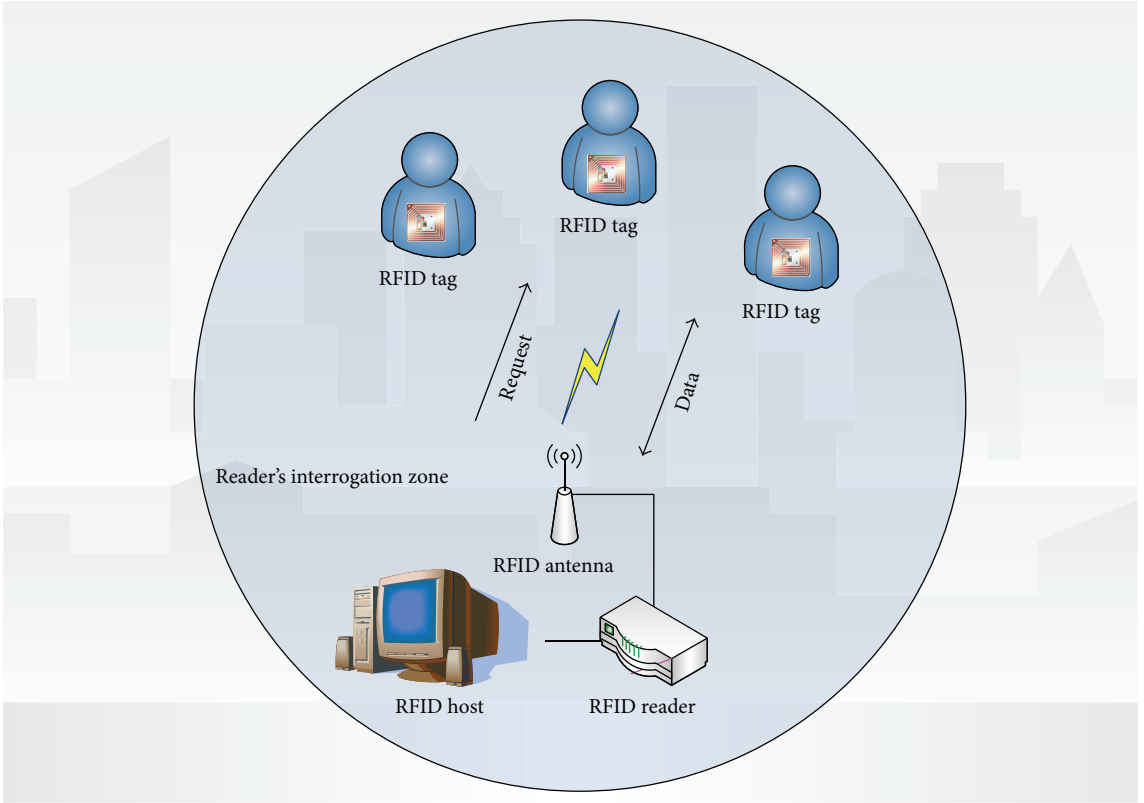


Figure 1: The conceptual framework of RFID system.

will lead to failed transmission, which means these collided tags have to transmit their data to the reader once again.

In RFID systems, each tag has its Unique ID (UID), and it will transmit its data, including its UID, to the reader when the tag receives the request from the reader. If there are more than one tag transmitting its data to the reader simultaneously in the same slot, the reader will not be able to decode the signal sent from the tags, and this means those tags will have to transmit their data once again, which causes degraded system performance and power consumption. According to the latest version of the RFID standard, EPC global UHF Gen2, Dynamic Framed Slotted Aloha (DFSA) has been accepted and employed as the de facto collision resolution algorithm to share the channel usage when multiple tags respond to the reader's signal command simultaneously. For DFSA, the size of contending frame will significantly affect the system performance because severe collisions occurred if the contending frame size is too small and bandwidth wastage is caused if the contending frame size is too large. However, unfortunately, according to current DFSA algorithm, the contending frame size is usually far from optimal since the reader does not have the information of how many unidentified tags are in its interrogation zone. In other words, if the reader is not able to get the information of how many tags exist, it is difficult to make the correct decision [10].

In the literatures, there have been adequate excellent discussions on the issues of collision resolution algorithm and its performance evaluation [11–13]. In general, collision resolution algorithms in RFID system can be divided into two categories, namely, Aloha-based approach [14, 15] and tree-based approach [16]. Tree-based algorithms repeatedly split the tags into two subsets until each subset has only one tag inside. Tree-based anticollision algorithm is efficient when the number of tags is small, but this approach is less efficient when the number of tags is large. Another kind of collision resolution algorithm, the Aloha-based algorithm, is a probabilistic approach. It assigns an amount of slots, called contending frame, for tags to transmit their data to the reader. Among all the Aloha-based algorithms, framed slotted Aloha (FSA) is preferred because of its simplicity and efficiency. The FSA algorithm is very similar to slotted Aloha protocol. The only difference is that, in FSA, it groups multiple time slots into a contending frame, and usually the contending frame length is decided based on the current number of unidentified tags. Particularly, according to previous works, the performance of FSA algorithm is optimal when the contending frame size equals the number of unidentified tags.

Since the performance of FSA algorithm depends on a suitable choosing contending frame size, a variant of FSA mechanism, Dynamic Framed Slotted Aloha (DFSA) collision resolution algorithm, is proposed [17] to improve the performance of FSA algorithm. In the DFSA algorithm, the reader can dynamically change the contending frame size according to the number of existing tags to increase the successful rate and decrease the collision rate with the view of

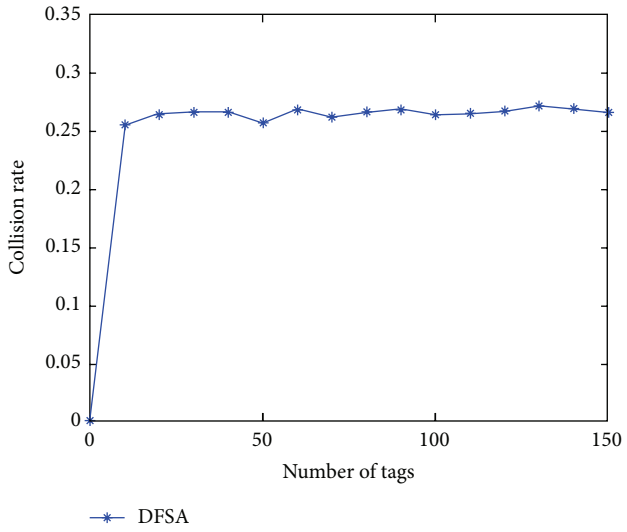


Figure 2: Collision rate for DFSA algorithm when the contending frame size is equal to the number of existing tags.

increasing the system efficiency. One may think that it is easy to get the conclusion of the maximum system throughput which could be obtained by setting the contending frame size equal to the number of existing tags [18]. As a matter of fact, according to my simulation results, the system throughput was not close to its theoretical limitation because the collisions still occurred when the tags transmit their data to the reader (see Figure 2). Besides, the information of the number of existing tags is difficult to get for reader in the current RFID system.

Although now we know that we should tune the frame size based on the current number of unidentified tags, however, the information of precise number of unidentified tags is difficult to get. In general, if the frame length is smaller than the optimal one, collisions occur frequently. On the contrary, idle time slots produced and system performance degraded. Therefore, correct estimation of the current number of unidentified tags is the key factor to successfully get an appropriate frame length for high system performance.

Based on the above observations, we can find that a small contending frame size could cause large numbers of collisions and bring bandwidth wastage in dense environments. However, a large contending frame size could cause large numbers of idle slots and this also brings bandwidth wastage in sparse environments. Hence, the reader should dynamically tune the contending frame size based on the information of number of existing unidentified tags. As a matter of fact, choosing an appropriate contending frame size is an effective technique to improve the overall RFID system performance. Now the question is as follows: how does the reader get the information of how many unidentified tags are in its surrounding area?

In this paper, I propose a run-time estimation scheme to effectively estimate the number of existing unidentified tags in reader's surrounding area in the RFID system, and this information can be used to adjust an appropriate contending frame size for DFSA collision resolution algorithm to improve the overall system performance. Instead of analytical analysis, I have also carried out a comprehensive simulation, developed by C language, to evaluate the performance of the proposed scheme. The simulation results confirmed that the proposed scheme can reduce the collision rate and increase the channel utilization comparing with the traditional DFSA collision resolution algorithms.

The remainder of this paper is organized as follows. Section 2 surveys the well-known collision resolution algorithms, FSA and DFSA, and some previous works on tags number estimation. Section 3 introduces the proposed algorithm. Section 4 shows the simulation results of the proposed scheme, and I compared it with some well-known algorithms. The conclusions are presented in Section 5.

2. Preliminary

According to the latest RFID standard, EPC global UHF Gen2, the Aloha-based collision resolution algorithm is the de facto MAC protocol for the passive RFID system because of its efficiency, and it is easy to implement. In this section, I give the preliminary knowledge on the two most important Aloha-based collision resolution algorithms, FSA and DFSA, and some well-known tags number estimation algorithms.

2.1. Framed Slotted Aloha (FSA) Algorithm. In FSA algorithm, all the contending frames are with the same length. When a tag sends its data to the reader, it randomly selects a slot in the contending frame to its data, and then the reader will send the tag an acknowledgement if the reader receives the data successfully. Note that, in FSA, the reader uses a fixed contending frame size and the tag will send its data in a cyclic sequence until the data was successfully received by the reader. Figure 3 illustrates the process of the FSA collision resolution algorithm. As shown in the figure, assume there are three tags which need to be recognized, and the contending frame size is four. At first, the reader sends a request command to the tags and it will also tell the tags that the contending frame size is four at the same time. In the first read cycle, tag 2 and tag 3 transmit their data in slot 3 at the same time, and hence collisions occurred. Tag 1 and only tag 1 sends its data in slot 2, and hence tag 1 can be successfully recognized by the reader. This process terminated until all the tags were recognized or there is no collision occurring in one contending frame.

FSA algorithm is inefficient since this strategy might incur a high collision rate if the reader selects a small contending frame. However, if the reader selects a large contending frame size, it is not good either since this could create a lot of idle slots, and this also implies bandwidth wastage.

2.2. Dynamic Framed Slotted Aloha (DFSA) Algorithm. Since a small contending frame size could cause a large number of collisions and bring bandwidth wastage in dense environments and a large contending frame size could cause a large number of idle slots and also bring bandwidth wastage in sparse environments, the reader should have the ability of

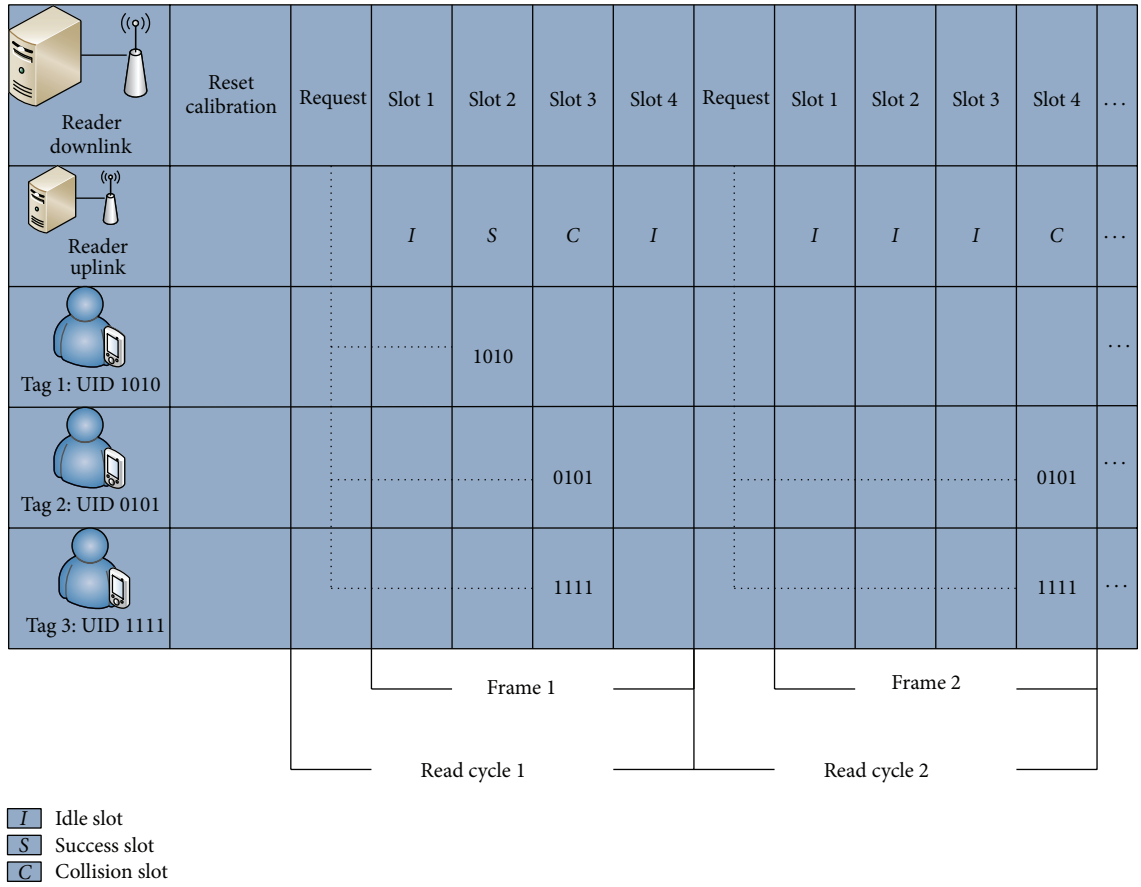


Figure 3: FSA collision resolution algorithm.

dynamically adjusting its contending frame size according to the number of unidentified tags. An enhanced FSA version, DFSA algorithm, was proposed to improve the performance of RFID systems. In general, DFSA algorithm has better performance than FSA algorithm since, in DFSA algorithm, the reader can dynamically adjust its contending frame size according to the number of unidentified tags. Figure 4 illustrates the process of DFSA collision resolution algorithm. As shown in the figure, assume there are three tags which need to be recognized, and, initially, the contending frame size is 4. At first, the reader sends a request command to the tags and indicates the contending frame size. In the first read cycle, tag 2 and tag 3 transmit their data in slot 3 simultaneously, and hence collision occurred. Tag 1 and only tag 1 sends its data in slot 2, and hence the reader can successfully recognize tag 1 because slot 2 is singly occupied. Since there is only one tag recognized in the first read cycle, this implies there are still two tags that need to be recognized. Hence, the reader might determine a new contending frame size, three, for the next read cycle. This process terminated until all tags were recognized or there is no collision occurring in one read cycle. In general, DFSA algorithm is more efficient than FSA algorithm because DFSA algorithm can dynamically adjust the contending frame size based on the current number of unidentified tags in the system [17].

Obviously, the performance of DFSA algorithm largely depends on a proper choice of contending frame size, and of course the contending frame size should be related to the number of unidentified tags in the system. However, when the reader is unable to know the exact number of unidentified tags, it is difficult for the reader to make a correct decision.

Regarding how to estimate the number of unidentified tags, there have been adequate discussions on this issue in the literatures. In [19–22], the authors developed the following equations to estimate the number of unidentified tags in RFID systems. The parameters a_0, a_1, and a_m are the expected values of idle, success, and collision slots in a contending frame, and E, S, and C are the observed idle, success, and collision slots. According to the theory of Chebyshev's inequality which indicates that in any probability distribution almost all random variables are close to the expected value, we can model the tags number, n, as follows:

$$\varepsilon_{vd} = \min \left| \begin{pmatrix} a_0 \\ a_1 \\ a_m \end{pmatrix} - \begin{pmatrix} E \\ S \\ C \end{pmatrix} \right|. \quad (1)$$

The expected values of number of idle, success, and collision slots are

$$a_0 = N \left(1 - \frac{1}{N}\right)^n,$$

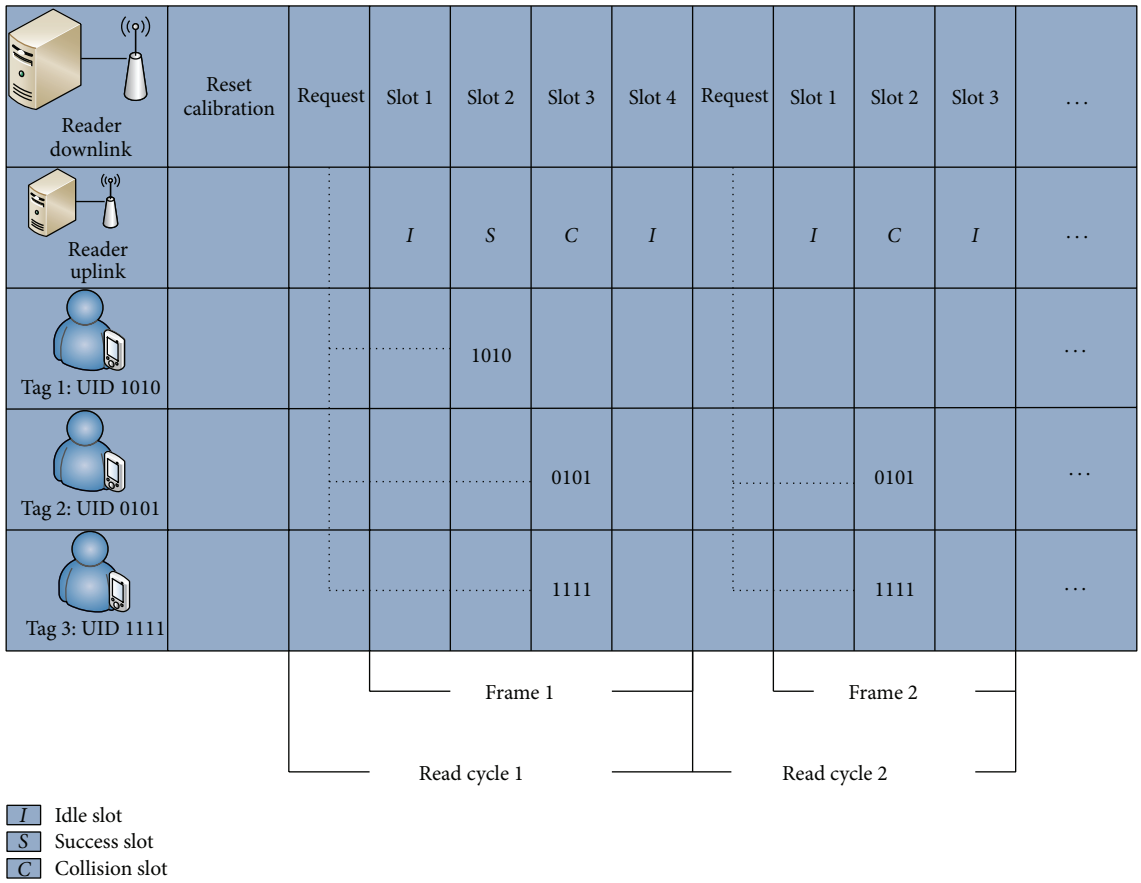


Figure 4: DFSA collision resolution algorithm.

$$a_1 = n\left(1 - \frac{1}{N}\right)^{n-1},$$
$$a_m = 1 - a_0 - a_1, \tag{2}$$

where n is the number of contending tags and the contending frame size is N.

In [20], the author also pointed out that it is intuitive to observe that the number of collided slots provides a lower bound of the actual number of unidentified tags trying to access the channel during the last read cycle because some tags might transmit their data in the same slot, and thus we have

$$N = R + 2C, \tag{3}$$

where C is the number of collision slots, R is the number of successful slots, and N is the estimated number of tags.

In paper [23] the authors proposed a modified equation to estimate the number of unidentified tags. Unlike (3), the authors proposed that the expected number of tags involved in each collided slot can be solved by the following equation:

$$C_{\text{tags}} = \frac{1}{C_{\text{rate}}} = \lim_{n\to\infty} \frac{1 - P_s}{P_c} = 2.3922, \tag{4}$$

where P_s and P_c denote probability of successful slot and collision slot. Hence, the estimated number of unidentified tags is

$$N = R + 2.39C. \tag{5}$$

2.3. Enhanced Tag Estimation Method. This method [24] only utilizes the probability for the collision slot in tag estimation but does not make use of the probabilities for the idle and success slots. Hence, it fully utilizes the mentioned information to increase the estimation accuracy, and it also needs to compute the average number of the tags included in a collision slot, but it does not suppose any condition on the tags number and the frame length. That is, the number of collision slots can be expressed as follows:

$$C_{\text{tag}} = \frac{n - P_{\text{succ}}N}{P_{\text{coll}}N} = \frac{n - n(1 - 1/N)^{n-1}}{N - N(1 - 1/N)^n - n(1 - 1/N)^{n-1}}. \tag{6}$$

Hence, the number of tags n can be obtained as follows:

$$n = S + C_{\text{tag}}C. \tag{7}$$

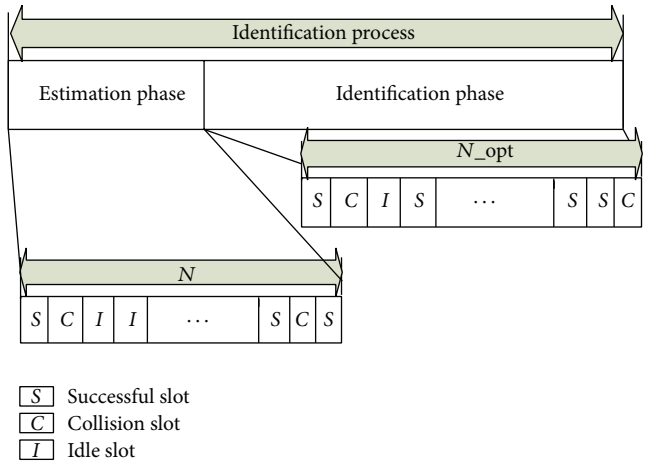


Figure 5: Identification process of the proposed scheme.

Table 1: Notations and variables used in analysis.

Notations and variables	Meaning and explanation
p_{idle}	Probability of idle slot
$p_{\text{successful}}$	Probability of successful slot
$p_{\text{collision}}$	Probability of collision slot
T_{req}	Time for request frame
N	Contending frame size
n	Number of unidentified tags

Finally, we can estimate the number of tags by solving the following equation:

$$C\frac{n - n(1 - 1/N)^{n-1}}{N - N(1 - 1/N)^{n} - n(1 - 1/N)^{n-1}} + S - n = 0. \quad (8)$$

3. Proposed Scheme

The basic idea behind the proposed scheme is simple. That is, in practice, when the reader monitors the channel status, low collision rate and low successful rate imply that the contending frame might be too large. On the other hand, high collision rate and low successful rate imply that the contending frame might be too small. Before we start to discuss the issues of interest, important notations and variables are defined in Table 1, and they will be used throughout this paper.

3.1. Run-Time Estimation of Number of Tags. As mentioned above, in a RFID system, a low collision rate implies that the number of unidentified tags is low, and the contending frame size should be set small. On the other hand, consecutive collisions indicate that there are numerous unidentified tags competing for the channel in the system. In such a case, the size of contending frame should be set considerably large to avoid collisions in the next read cycle. Hence, in order to exploit the information of the channel status, we define the probability of idle slot, p_{idle}, to be the probability that an idle slot occurs in a time slot in the contending frame. Recall that the channel status can be generally divided into three states: idle, successful, and collision. Therefore, the probability of idle slot can be defined as follows:

$$p_{\text{idle}} = \frac{\text{number of idle slots}}{\text{contending frame length}}. \quad (9)$$

As the channel statuses are ever-changing, hence, in practice, the value of p_{idle} has to be updated for every ready cycle to reflect the actual state of the channel, as shown in Figure 5.

Pseudocode 1, C language-like pseudocode, describes the proposed scheme.

Finally, in [24] the authors developed an analytical model to study the system throughput of DFSA based RFID systems. Hence, once I get the information of how many existing tags are in reader's surrounding, I will use the model proposed in [24] to search for an optimal frame size that maximizes the system throughput based on the estimated results of number of unidentified tags.

4. Simulation Results

In this section, I evaluate the performance of the proposed scheme. I carry out the experiments with four different collision resolution algorithms, namely, DFSA with $S + 2C$, DFSA with $S + 2.39C$, DFSA with minimum distance, and the proposed scheme. To focus on the system throughput related issues and to reduce the complexity of simulations, what follows are the basic assumptions used in our simulation. First, no tags leave or enter reader's interrogation zone during simulation. Second, the default values used in the simulations are listed in Table 2, and each simulation runs at least 1,000 times. Finally, in DFSA, optimal frame size means that the contending frame size equals number of existing tags. As for the proposed scheme, frame size N is set according to the research results described in [25]. Also, if the tag identification process is not complete yet, both DFSA and

```
Proposed_Scheme ()
{
// initialization
N = 64; // initial value of contending frame size
tags[n] = 0; // initial value of number of un-identified tags
idle = 0; // initial value of collision rate
estimate_counter = 0; // initial value of estimated counter
do
{
        for (i = 0; i ≤ n − 1; i++) {
        tags[i] = (rand()%N) + 1;
        } // each tag randomly selects an available slot from 1 to N
        for (j = 0; j ≤ frame − 1; j++) {
        for (k = 0; k ≤ n − 1; k++) {
        If (tags[k] == (j + 1)) {
        collision[j]++; } }
        }// investigating the total occupied slot number
        for (int l = 0; l ≤ frame − 1; l + +) {
        if (collision [l] == 0)
        idle ++;
        }//investigating the total idle slot number
        estimate_counter ++;
        p_idle = idle/N;
        If (p_idle == 0){
                N = N × 2;
                estimate_counter −−;
        }
        If (estimate_counter == 1){
        n_estimate = log10(p_idle)/log10(1 − (1/N)); }
  }while (estimate_counter ! = 1)
    // estimation process terminated
}
```

Pseudocode 1

Table 2: Default attribute values used in the simulation.

Attribute	Value/setting	Meaning and explanation
σ	4.9 ms	Time needed for each time slot
T_{req}	52 ms	Time needed for request command
G_{tag}	−29 dBm	Transmitting power of tag
G_{reader}	36 dBm	Transmitting power of reader
N	64 slots	Initial contending frame size

the proposed scheme will adjust the contending frame size in next ready cycle.

Figure 6 shows the estimated error versus the number of tags. Simulation results indicate that the proposed method has the lowest estimated error rate (less than 6%). However, $S + 2C$ and $S + 2.39C$ usually have higher estimated error rate, especially when the number of tags becomes large. As for the performance of Vogt's method [19–23], this method also can achieve an estimated error rate lower than 6%, in average.

Figure 7 depicts the total delay time as the number of tags increases. Total delay time is defined as the total time for the process of tags identification to identify all tags within reader's interrogation zone. In general, for all algorithms, the total time needed to identify all the tags increases as the number of tags increases. Although there is not much difference in the measured values when the number of tags is large, please note that the proposed scheme reaches a better performance than the DFSA algorithm when the number of tags is small. This is because the slot time is 4.9 ms and the time for request command is 52 ms. Hence, if there is small number of tags in reader's interrogation zone, the length of request command occupies higher ration of each read cycle for DSFA algorithm, and this causes longer total delay time since request command is an overhead. However, as shown in Figure 7, the proposed scheme still reaches a better performance in most cases.

In Figure 8, we compare the number of read cycles for each algorithm as the number of tags gradually increases. As shown in the figure, the proposed scheme has the least number of read cycles because its contending frame size is optimal. On the contrary, the DFSA algorithm has the highest number of read cycles because its frame size is usually the smallest among all the collision resolution algorithms, and this leads to large number of read cycles.

Figure 9 depicts the collision rate as the number of tags increases. As we expected, the proposed scheme has the lowest collision rate because its contending frame size is optimal, and it can effectively reduce the collision rate.

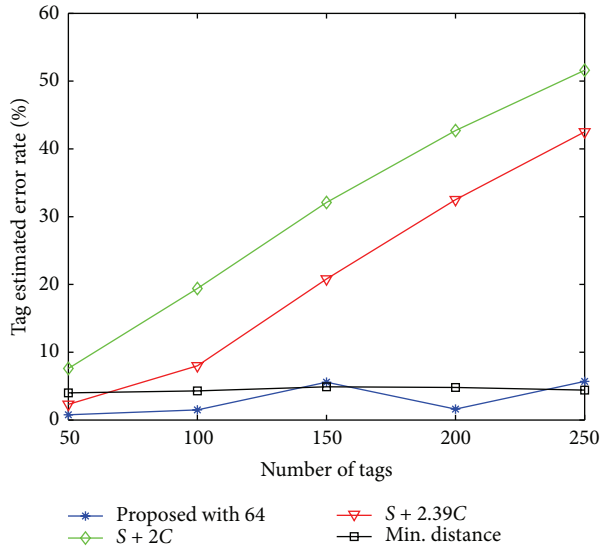


Figure 6: Simulation results for tag estimated error.

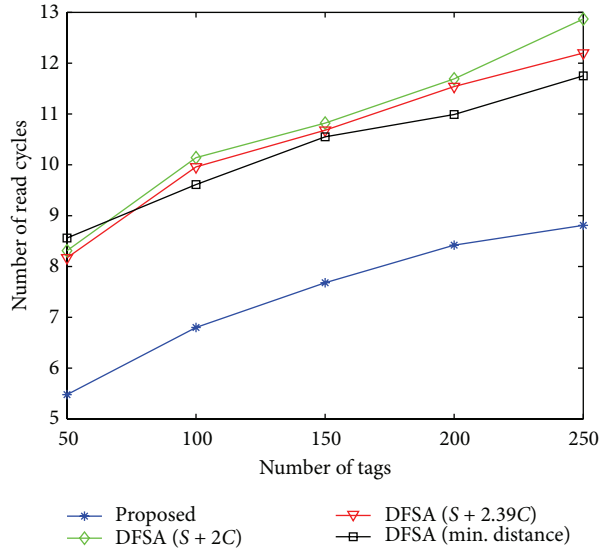


Figure 8: Read cycle versus number of tags.

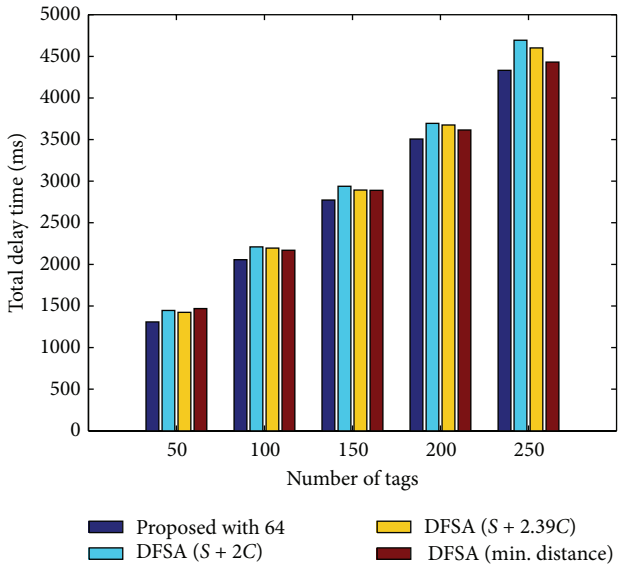


Figure 7: Total delay time versus number of tags.

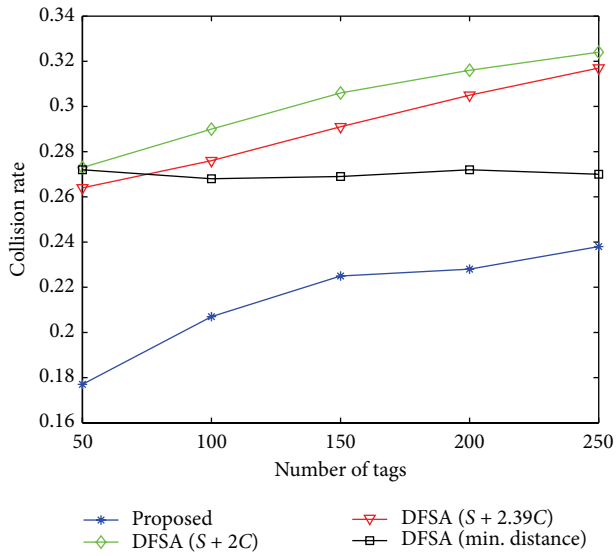


Figure 9: Collision rate versus number of tags.

However, as shown in Figure 9, the DFSA algorithm has the highest collision rate because it usually has the smallest contending frame size among all of the collision resolution algorithms. Comparing this with the simulation results shown in Figure 7, we can observe that the gain of reduced collision rate is at the cost of increased total delay time.

In Figure 10, I compare the utilization factor of each collision resolution algorithm. Utilization factor is defined below:

$$\text{Utilization factor} = \frac{\text{Number of successful slots}}{\text{Number of total slots}}. \quad (10)$$

Please note that, in the above equation, total time slot is the sum of successful slots, idle slots, collision slots, and request command. Obviously, the proposed scheme can get better performance than the DFSA algorithm as it can dynamically adjust its contending frame size according to the current channel status. Furthermore, the proposed scheme obviously outperforms the other three collision resolution algorithms in most cases. This is because the proposed scheme takes the time of request command into account, but DFSA does not consider the time of request command, and, however, request commend is a fixed overhead in each read cycle.

Finally, in Figure 11, we depict the simulation results of power consumption versus the number of tags. As shown in the figure, the DFSA algorithm sends the most signals to identify all tags and, on the contrary, the proposed scheme sends the least signals to identify all tags. Based on the simulation results, it is confirmed that the proposed scheme is

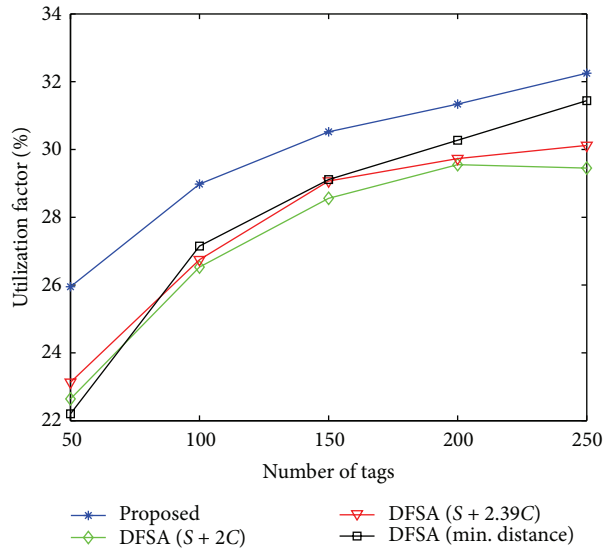


Figure 10: Utilization factor versus number of tags.

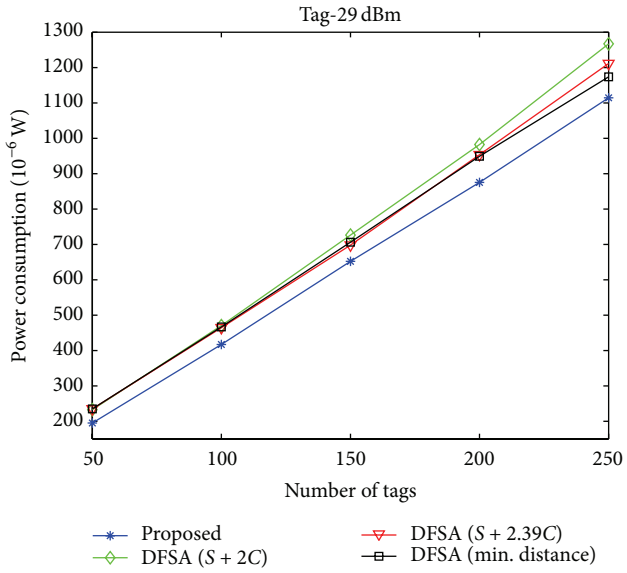


Figure 11: Power consumption versus number of tags.

effective because it achieves lower power consumption, lower reading time, and higher system throughput.

5. Conclusions

In construction of IoT network, the RFID technologies play the role of the front-end data collection, as the basis of IoT. However, in the RFID system, the contending frame size of traditional DFSA based collision resolution algorithm is far from being optimal, and this brings more power consumption, longer reading time, and degraded system throughput. In this paper, I propose a simple and easy-to-implement but yet well-performing pragmatic solution to find out the number of unidentified tags and then I use this result to search for an optimal contending frame size for DFSA based collision resolution algorithm in RFID systems. According to the simulation results, the proposed tags number estimation algorithm has an estimated error rate less than 5%, and it is a strong evidence that the proposed scheme is very efficient. Also, through extensive simulations, important performance metrics such as total delay time, collision rate, and system utilization are thoroughly investigated. Comparing with traditional DFSA algorithm, the simulation results show that the proposed scheme can reduce 8% of collision rate. That is, it proves that the proposed scheme is very energy-saving and cost-effective. In the future, I am planning to apply the same technique in the next generation WiFi protocol and its application in business [26] and try to develop new technique and solutions to deal with the potential weakness in RFID-based Internet of Things systems [27].

Disclosure

A preliminary version of this paper was published in the conference proceedings of ISWPC 2011.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

The author would like to thank Hsuan-Wei Tsao for his help with the simulation.

References

[1] H. W. Wang, "An explorative study of continuance intention on cloud learning with applying social networks blended in new era of internet technology applications," *Journal of Internet Technology*, vol. 16, no. 3, pp. 563–570, 2015.

[2] Y.-C. Chang and H.-W. Wang, "Mobile business via cross layer approach toward intelligent RFID purchasing system," *Journal of Internet Technology*, vol. 11, no. 7, pp. 965–974, 2010.

[3] D.-J. Deng, L.-W. Chang, H.-W. Wang, D.-C. Huang, and Y.-M. Huang, "Is RTS/CTS mechanism effective for WLANs?" *Journal of Internet Technology*, vol. 11, no. 7, pp. 955–964, 2010.

[4] Z. G. Prodanoff, "Optimal frame size analysis for framed slotted ALOHA based RFID networks," *Computer Communications*, vol. 33, no. 5, pp. 648–653, 2010.

[5] K. Finkenzeller, *RFID Handbook: Radio Frequency Identification Fundamentals and Applications*, John Wiley & Sons, 2010.

[6] International Organization for Standardization, *ISO/IEC 18000 Part 6: Information Technology Automatic Identification and Data Capture Techniques-Radio Frequency Identification for Item Management Air Interface*, International Organization for Standardization, 2003.

[7] M. Tajima, "Strategic value of RFID in supply chain management," *Journal of Purchasing and Supply Management*, vol. 13, no. 4, pp. 261–273, 2007.

[8] L. Castro and S. F. Wamba, "An inside look at RFID technology," *Journal of Technology Management Innovation*, vol. 2, no. 1, 2007.

[9] M. Nuce, "Five frequently asked questions about RFID," *Retail-intelligence*, pp. 18–20, 2015, http://www.apparelmag.com/.

[10] D.-J. Deng, C.-H. Ke, H.-H. Chen, and Y.-M. Huang, "Contention window optimization for ieee 802.11 DCF access control," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 5129–5135, 2008.

[11] D.-J. Deng, H.-C. Chen, H.-C. Chao, and Y.-M. Huang, "A collision alleviation scheme for IEEE 802.11p VANETs," *Wireless Personal Communications*, vol. 56, no. 3, pp. 371–383, 2011.

[12] H.-H. Chin, C.-C. Lin, and D.-J. Deng, "E-BEB: enhanced binary exponential backoff algorithm for multi-hop wireless ad-hoc networks," *Wireless Personal Communications*, vol. 76, no. 2, pp. 193–207, 2014.

[13] D.-J. Deng, "PSSB: priority enforced slow-start backoff algorithm for multimedia transmission in wireless ad-hoc networks," *Journal of Network and Computer Applications*, vol. 34, no. 5, pp. 1468–1473, 2011.

[14] Auto-ID Center, *Draft Protocol Specification for a 900 MHz Class 0 Radio Frequency Identification Tag*, Auto-ID Center, 2003.

[15] H. W. Tsao, D. J. Deng, H. W. Wang, and J. H. Chang, "Runtime optimization of framed slotted ALOHA based RFID Systems," in *Proceedings of the 6th International Symposium on Wireless and Pervasive Computing (ISWPC '11)*, pp. 1–6, Hong Kong, February 2011.

[16] ISO, "Information technology—radio frequency identification for item management—part 6: parameters for air interface communications at 860 MHz to 960 MHz," ISO/IEC 18000-6(E), ISO, 2004.

[17] F. C. Schoute, "Dynamic frame length ALOHA," *IEEE Transactions on Communications*, vol. 31, no. 4, pp. 565–568, 1983.

[18] J.-R. Cha and J.-H. Kim, "Novel anti-collision algorithms for fast object identification in RFID system," in *Proceedings of the 11th International Conference on Parallel and Distributed Systems (ICPADS '05)*, pp. 63–67, IEEE, Fukuoka, Japan, July 2005.

[19] H. Vogt, "Efficient object identification with passive RFID tags," in *Pervasive Computing: First International Conference, Pervasive 2002 Zurich, Switzerland, August 26–28, 2002 Proceedings*, vol. 2414 of *Lecture Notes in Computer Science*, pp. 98–113, Springer, Berlin, Germany, 2002.

[20] H. Vogt, "Multiple object identification with passive RFID tags," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC '02)*, Hammamet, Tunisia, October 2002.

[21] H. Vogt, "Efficient object identification with passive RFID tags," in *Pervasive Computing*, vol. 2414 of *Lecture Notes in Computer Science*, pp. 98–113, Springer, Berlin, Germany, 2002.

[22] H. Huo, J. Q. Liu, and Y. J. Wang, "Flood diversion algorithm for anticollision in RFID system," *International Journal of Distributed Sensor Networks*, vol. 2015, Article ID 245914, 11 pages, 2015.

[23] L. Jia and Z. Youguang, "The analysis of anti-collision algorithm based on timeslot in RFID system," in *Communication and Network*, 2006.

[24] Z. Li, C. He, and H.-Z. Tan, "An enhanced tag estimation method applied to tag anti-collision algorithm in RFID systems," in *Proceedings of the International Conference on Information Science and Technology (ICIST '11)*, pp. 703–708, Nanjing, China, March 2011.

[25] D.-J. Deng and H.-W. Tsao, "Optimal dynamic framed slotted ALOHA based anti-collision algorithm for RFID systems," *Wireless Personal Communications*, vol. 59, no. 1, pp. 109–122, 2011.

[26] D.-J. Deng, K.-C. Chen, and R.-S. Cheng, "IEEE 802.11ax: Next generation wireless local area networks," in *Proceedings of the 10th International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness (QSHINE '14)*, pp. 77–82, Rhodes, Greece, August 2014.

[27] I. Erguler, "A potential weakness in RFID-based Internet-of-things systems," *Pervasive and Mobile Computing*, vol. 20, pp. 115–126, 2015.

Hindawi Publishing Corporation
Mobile Information Systems
Volume 2015, Article ID 521093, 14 pages
http://dx.doi.org/10.1155/2015/521093

Research Article

Power Aware Mobility Management of M2M for IoT Communications

Awais Ahmad,[1] Anand Paul,[1] M. Mazhar Rathore,[1] and Seungmin Rho[2]

[1]*School of Computer Science and Engineering, Kyungpook National University, Daegu 702-701, Republic of Korea*
[2]*Department of Multimedia, Sungkyul University, Anyang 431-003, Republic of Korea*

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

Received 28 May 2015; Accepted 29 July 2015

Academic Editor: Neeraj Kumar

Machine-to-Machine (M2M) communications framework is evolving to sustain faster networks with the potential to connect millions of devices in the following years. M2M is one of the essential competences for implementing Internet of Things (IoT). Therefore, various organizations are now focusing on enhancing improvements into their standards to support M2M communications. Thus, Heterogeneous Mobile Ad Hoc Network (HetMANET) can normally be considered appropriate for M2M challenges. These challenges incorporated when a mobile node (MN) selects a target network in an energy efficient scanning for efficient handover. Therefore, to cope with these constraints, we proposed a vertical handover scheme for handover triggering and selection of an appropriate network. The proposed scheme is composed of two phases. Firstly, the MNs perform handover triggering based on the optimization of the Receive Signal Strength (RSS) from an access point/base station (AP/BS). Secondly, the network selection process is performed by considering the cost and energy consumption of a particular application during handover. Moreover, if there are more networks available, then the MN selects the one provided with the highest quality of service (QoS). The decision regarding the selection of available networks is made on three metrics, that is, cost, energy, and data rate. Furthermore, the selection of an AP/BS of the selected network is made on five parameters: delay, jitter, Bit Error Rate (BER), communication cost, and response time. The numerical and experimental results are compared in the context of energy consumption by an MN, traffic management on an AP/BS, and QoS of the available networks. The proposed scheme efficiently optimizes the handoff related parameters, and it shows significant improvement in the existing models used for similar purpose.

1. Introduction

Adaptation of heterogeneous access network and efficient use of available resources are one of the key confront for the next generation of mobile communication. M2M practices miscellaneous models in terms of the electronic smart grid, connected cars, body area network, and Android communication using vehicular technologies [1]. M2M communication is an important facilitating technology for Internet of Things (IoT), which facilitates an outsized number of implicated escorts to a large number of research challenges [2]. In 3GPP terminology, M2M communication is usually referred to as Machine-Type Communication (MTC) [3–5].

Over the last few years, the numbers of heterogeneous networks available at a particular location were increased noticeably [6]. Various communication networks show inherent characteristics in terms of handover failure and consumption of energy and cost, which offer higher communication performance. To highlight the eventual function of M2M communication for developing wide-ranged connections among various devices, the potential of HetMANET cannot be neglected in this regard.

A handover process starts from the machine when it experiences a weak received RSS from a base BS/AP. When RSSI reached a predefined threshold, machine (MN) starts to search for available networks. The handover time is mainly dependent on the scanning delay of the available networks. Furthermore, an optimum network can be selected for effective handover process among available networks on the basis of price, security, transmission rate, and quality of service (QoS).

Employing the available technologies for MTC leads to various challenges, including a selection of the best network for handover, incompatibility among different networks, and handover delay. To address these challenges in the HetMANET, an efficient, organized handover management scheme is required, which can switch communication data from one network to another with the minimum packet loss and delay, respectively. When a device is moving from one BS/AP to another, it executes a discovery mechanism for searching nearby BSs/APs and then establishes a connection with higher QoS. The selection of inappropriate network introduces high handover time and delay in a handover process. This handover delay can be minimized by adjusting different factors like RSS, data rate, available bandwidth, and Signal to Interference and Noise Ratio (SINR) from a BS [7, 8].

In 2008, International Telecommunication Union-Radio Sector (ITU-R) defined new specifications for 4G standard called International Mobile Telecommunication Advanced (IMT-Advanced). IMT-Advanced supports 100 Mbits/s for high mobility connections and 1 Gbits/s for low mobility connections [9]. With the increase in data rates, new technologies such as WiFi and WiMAX participate actively to develop new technology modification. Development of new technologies is urgently needed since these networks face key constraints of compatibility issues.

The IEEE 802.21 is the MIH standard in 2008 for seamless handover between networks of the same and different type [10]. Recently, much research has been performed to improve the currently available MIH standard [11–13]. The MIH standard is still facing many challenges that need to be addressed; that is, (i) long handover time is required for the MIIS server, which is located many hops away, (ii) time needed for handover process is very short when the handover number is frequent in a handover region, and (iii) failure of a hop requires alternate routes to connect MIIS server that can increase handover time. In the MIH standard, MN initiates handover upon receiving RSS below the predefined threshold. The time required for handover is constant even if the MIIS server is located many hops away. When MIIS server is located many hops away, a longer time is required for MN to get the information of the available networks. If an alternate route is available for MIIS server, which consists of several hops in the case of route failure, then the time required for handover will also increase that may cause the breaking of connection during handover process in the worst condition. MIH standard utilize RSS for handover initiation. RSS suffered from different problems such as wrong network selection and too early and late handover, as shown in Figure 1.

Therefore, to address the challenges above, this paper introduces QoS based efficient handover scheme in which a BS collects information, in advance, of the available networks for MN. In this scheme, MN does not wait for long time in a handover region for the collection of information of the available network. Each time when MN needs information of available networks, it is available one hop away from BS. This straightforward and distinct scheme has many advantages including the following: (i) MN search for nearby BS/AP is done in an efficient manner, (ii) the proposed scheme is not affected by the number of incoming connections, and (iii) MN does not consume excess energy when it scans for nearby network.

The remainder of this paper is organized as follows. In Section 2, we give a brief background of the existing schemes and their shortcomings faced by handover for M2M communications in HetMANET. In Section 3, we propose an energy efficiency vertical handover scheme from M2M communication. In Section 4, a detailed analytical and simulation analysis is discussed in detail. Finally, Section 5 offers a conclusion.

2. Related Work

In the last decade, various schemes have been proposed for the improvements of handover management in HetMANET. Most of these schemes are based on the optimization of different parameters necessary for handover. The optimization of these parameters reduces the handover time and latency. With the passage of time, the number of the new access networks was rapidly increased, producing signaling overhead and other issues related to handover phenomenon. Similarly, the access new technologies such as LTE Advanced and Bluetooth 4.0 low energy were introduced to save communication time and energy. All of the recent technologies try to provide its customer with the best QoS. The QoS of a network can be enhanced if a consumer is provided with a continuous connection to different networks.

To upgrade the QoS of a network, a scheme is proposed, based on optimizing parameters such as Bit Error Rate, delay, jitter, and data rate in [14]. The decision of handover is performed by using fuzzy logic and analytic hierarchy approaches. The proposed scheme receives the context information like networks related information, user preferences, and service requirements for an efficient handover process. MNs periodically check the RSS level with the current AP/BS; if the RSS drops below a particular level, then the MN initiates network selection phase. The network quality scoring function is defined to evaluate the QoS of a network. The network with the highest QoS is selected for the handover. The proposed scheme takes handover initiation decision on the basis of available RSS, due to which the number of false handover indications increases considerably [15].

Considering the challenges to minimize the handover number, a scheme based on self-optimizing is proposed, which extracts velocity of the MN on the basis of location information followed by the selection of an appropriate RSS level for the handover [16]. The optimization of a single parameter is not a generic solution for handover management. A complete set of handover parameters should be optimized for an efficient handover process. The efficiency of handover can be maximized if each phase of the handover process is optimized for the best performance. The MIH standard does not provide an optimized solution for a handover management because this standard supports only one MIIS server for all of the available access networks. In future, the number of access networks and MNs will be increased due to advancement in network access technologies. Thus,

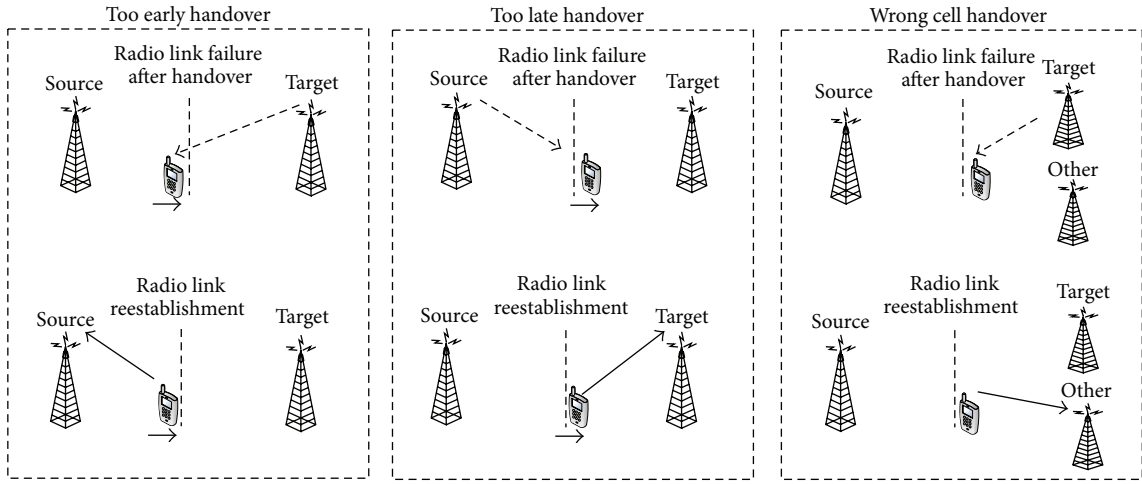


Figure 1: Handover problems due to RSS.

a single MIIS server will not be appropriate for all of the access networks.

In the next generation network, MN will be provided with multiple optimized routes to send data from one end to another. During a handover process, selection of an optimized route for data transfer is a challenging task. A scheme has been proposed to provide MN with an optimized route after the handover has been processed [12]. The proposed scheme efficiently reduced the handover latency and achieved fast recovery of the optimized path. A similar scheme has also been proposed in [13]. The scheme optimizes the route optimization for tunnel establishment to buffer packets during handover. The scheme efficiently solves buffer overflow problem in proactive handover techniques. The packet loss and handover delay due to buffer overflow are significantly minimized. However, still many issues are yet to be answered in terms of the selection of the optimal network during handover. The route optimization not only helps in balancing traffic on a particular AP/BS, but also maximizes the probability of new connections on an optimized route.

The energy consumption during the selection of networks is a major factor in a handover process in HetMANET. The energy consumption by an MN directly depends on the application running during a handover process. A scheme based on the energy efficient handover for multimedia based applications is proposed, which utilizes the concept of adapt or handover for balancing the multimedia traffic during a handover process [17]. The proposed scheme saves energy consumed due to the insignificant degradation in QoS. A single objective handover management cannot be adopted for a generic solution. However, the energy could also be reserved by the real-time power managements scheme in M2M communications, mobility management, probabilistic modeling, and graph based on M2M communications, [18–21]. Furthermore, the same parameters are also achieved by optimizing data transmission in device-to-device communication and WSN, which is based on the advanced clustering scheme [22, 23]. Such scheme is based on the received signal strength of the sensor node.

The selection of the less expensive network with best QoS during handover leads to the smooth transition of an ongoing session from one network to another. Therefore, the cost optimization must be considered during a handover process. A scheme based on cost aware handover decision is proposed which uses two cost functions, that is, triggering and priority decision [24]. Both functions are optimized for the best values of signal transmission quality, handover signaling cost, handover latency, and estimate interference. The proposed scheme efficiently transfers an ongoing session from one cell to another after checking the cost of the adjacent cells. However, still some of the parameters such as data rates and data rate based costs are not addressed in the current schemes. Therefore, we propose a solution that considers possible multiple parameters that affect the quality of a handover process in M2M.

3. Proposed Scheme

This section presents our proposed handover scheme for M2M in detail. Figure 2 delineates the architecture of handover scheme that M2M practices. Multiple BSs and APs are deployed in the large geographic area having different networks (HetMANET), in which an MN moves from one network to another, performing some handover.

3.1. Assumptions and Definitions. In this section, we present assumptions made during the design of our network and simulation model. Some of the scenario related definitions are also given.

Assumption 1 (heterogeneous devices). All the MNs have different configuration; that is, their battery requirements are different from each other.

Assumption 2 (communication radius model). In the communication range of a BS, "A" has the radius "R" that is centered at "c." It can be defined as $\mathrm{CR}(c, R) = \{A, q \in S : D(A-q) \leq R_A\}$, where CR represents communication radius,

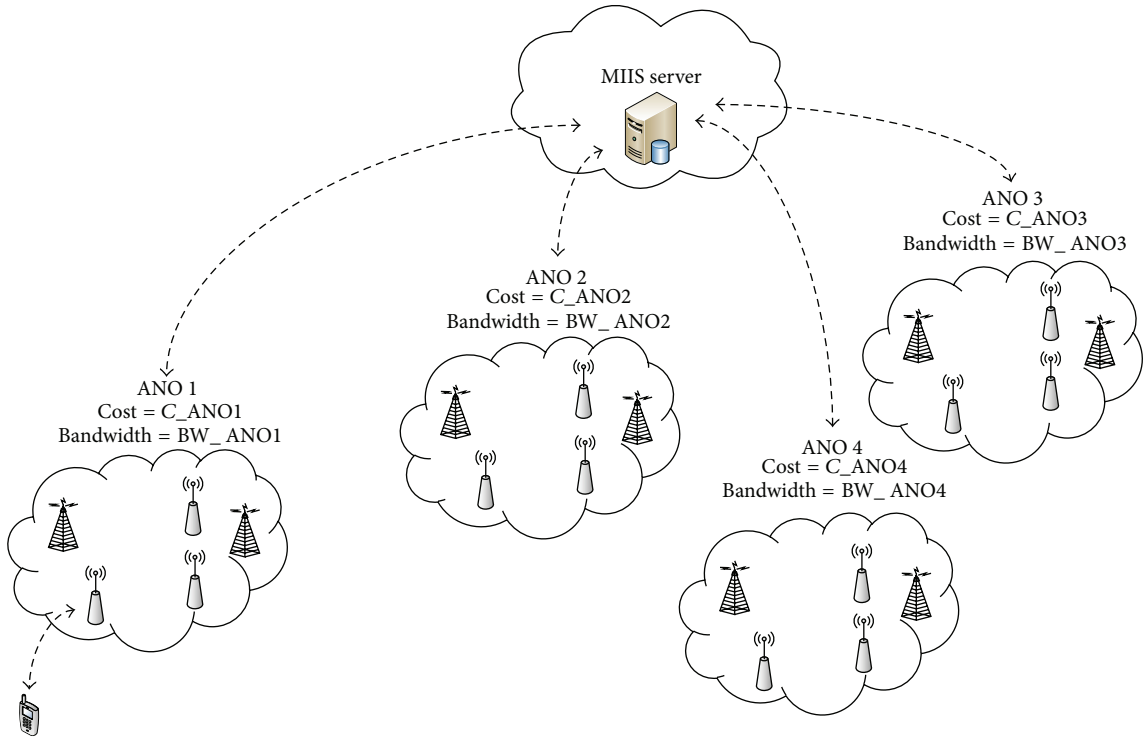


Figure 2: Proposed M2M communication handover scheme.

S represents the set of deployed nodes, and $D(A - q)$ is the distance between BS and q in the M2M network.

Definition 3 (Medium Scale Network). If all the MNs have direct communication access to the BS/AP, then the network is considered to be Medium Scale Network (MSN). Suppose that, in any environment, the M2M network comprising of 100 MNs deployed in the area of 100 m × 100 m is considered as MSN. This definition can be modeled as $\forall p \Lambda p \in S$, $|D(p - \mathrm{BS})| < R_p$, where p is an MN among the set "S" of deployed MNs and $D(p - \mathrm{BS})$ is the distance between any of the deployed network nodes say, p and the BS. R_p is the communication radius of node p.

Definition 4 (Large Scale Network). If any of the deployed MN does not have direct communication access to the BS/AP, then the network is considered to be Large Scale Network (LSN). Suppose that, in any environment, the M2M network comprising of 100 MNs deployed in the area of 200 m × 200 m is considered to be Large Scale Network (LSN). This definition can be modeled as $\exists p \Lambda p \in \mathrm{S}$, $|D(p - \mathrm{BS})| > R_p$, where p is a node among the set "S" of deployed nodes and $D(p-\mathrm{BS})$ is the distance between any of the deployed network nodes say, p and the base station. R_p is the communication radius of node p.

3.2. Overview. An MN can perform a handover from one Access Network Operator (ANO) to another upon weak link connectivity. The MN obtains the information of cost and data rate of available networks from the MIIS server during handover to select the target network. The MIIS server stores the information of geographical locations of point of attachments (PoAs) of an ANO. Every ANO needs to send information regarding cost packages and data rates to the MIIS server. If an ANO updates either the cost model or the data rate information, it will also update this information in the MIIS server. Figure 2 delineates the fundamental idea of the proposed scheme. The proposed scheme consists of three phases: (1) handover triggering phase (2), network selection, and (3) handover execution phase.

3.2.1. Handover Triggering Phase. In the proposed scheme, we have used a threshold mechanism for handover triggering. This mean that handover is triggered if the RSS from the current network drops below a predefined threshold. An optimal threshold mechanism reduces the number of false handover indications as well as the number of handover failures to a network with overloaded APs/BSs. We set a threshold of RSS level on the boundary of the coverage area. Let d represent the radius of the coverage area of AP and BS. According to the signal propagation model [25], the threshold should be set based on the distance $(1-\delta) \times d$ from AP or BS, where δ represents the fluctuation that is produced due to the variation in network data rate dynamics. The value of δ is taken between 0 and 1. The threshold is given by the following equation:

$$\theta_T = K_1 - K_2 \log\left((1 - \delta)\, d\right), \tag{1}$$

where K_1 represents the antenna gain and signal wavelength and K_2 represents the path loss factor. Most of the traditional

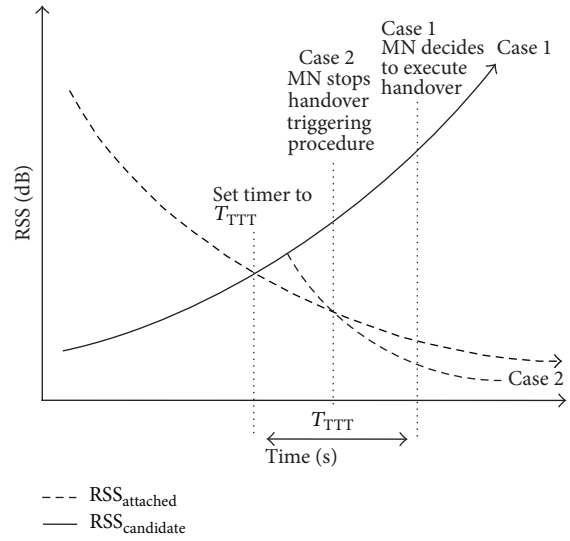


Figure 3: RSS based handover in M2M Communication.

approaches have used RSS for triggering handover. RSS based triggering techniques have several limitations: (i) selection of inappropriate network, (ii) an overloaded AP/BS selected for handover, and (iii) increases of false handover indications. To avoid these problems, we optimize the RSS value using (1), and it considerably reduces the false handover indications as shown as in Figure 3.

With a view to elaborating the proposed handover triggering phase, we consider a reference example based on Figure 3 [26]. In 3GPP, various handover measurements techniques are defined which supports mobility [27, 28]. However, handover triggering based on RSS and Time-to-Trigger (TTT) techniques are usually used in a horizontal handover in LTE system since its simplicity and efficiency made it easy to implement [29]. As shown in Figure 3, MN periodically measures the RSS of the neighboring AP/BS. If the RSS of the candidate MN is greater than the RSS from the current MN, the timer is set to T_{TTT} T_{TTT} seconds by the MN and starts to observe $RSS_{candidate}$ and $RSS_{attached}$. Apparently, if $RSS_{candidate} > RSS_{attached}$ is incessantly following during the T_{TTT} seconds, the MN performs handover to the candidate MN (case 1). However, if the $RSS_{candidate} < RSS_{attached}$ follows during the T_{TTT} seconds, the MN stops observation and come back to its initial state (case 2).

3.2.2. Network Selection. The network selection phase is further divided into the following subsections.

(i) Cost. To select the new network for optimal handover, we need to consider the cost of applications used by an MN. The target network assigns the same cost as that of the old network to continue the movement of the MN in HetMANET. If a target network is not providing the cost that is equal to the old network, then the MN selects new cost from the target network, which is acceptable to the MN. Otherwise, the MN experiences long delay and even breaking of connection during handover. The MN normally uses different types of applications like real-time streaming applications, voice, elastic applications (web browsing, chatting, etc.), and interactive audio and video call.

Table 1: Application weight table.

Application type	Weight
Interactive	$W1$ (0.1~0.3)
Elastic application	$W2$ (0.3~0.5)
Voice	$W3$ (0.5~0.7)
Real-time streaming	$W4$ (0.7~0.9)
Demagnetizing factor	$1 \rightarrow 1/(4\pi)$

Multimedia applications require more cost as compared to elastic applications. We assign a particular weight of cost to each category of application as listed in Table 1.

Assuming that there are M applications, the weights of all the applications running on the MN, denoted by C, are as follows:

$$C = \sum_{i=1}^{M} w_i C_i, \tag{2}$$

where w_i represents the weight of each application. The value of various weights is taken from 0 to 1, depending on the priority of an application. An application with the highest priority is assigned the largest weight. For instance, if MN's device is running a real-time streaming application, it will be assigned the highest weight since the streaming application can tolerate a handover delay of only 150 ms and a packet loss of 3%, respectively [30]. Therefore, the MN selects a network with less possible cost, which has the potential to run a particular application during switching from one network to another.

(ii) Energy. In HetMANET, the MN consumes a significant amount of energy for scanning the available PoAs. In particular, the application with high priority needs much energy, which is normally required for fast scanning. Depending on the density of the medium (APs and BSs) of the network, the interface for a particular network is periodically switched to sleep and active states. The energy (denoted by E) required by MN for the scanning of a particular PoA of a network is given by

$$E = \sum_{i=1}^{n} P_i \times t_s, \tag{3}$$

where P_i is the power required by MN for the scanning of a PoA of an access network and t_s represent the time taken for interface scans.

The energy required for scanning during handover depends on the application used by MN. If the request is of a high priority, then the MNs perform fast scanning. In this case, the energy required for scanning will be high. In traditional approaches, the scanning procedure is throughout the uniform and mainly depends on the RSS from an AP/BS, which leads to a high packet loss. Thus, we restrict

the scanning energy consumption depending on the applications running by the MN during scanning. The proposed energy efficient scanning procedure significantly reduces the energy consumption by the MN during handover.

(iii) QoS Computation. The network selection phase is considered an important factor in a handover management scheme. When the MN is moving across the HetMANET, it performs some handover switching from one network to another. For efficient handoff, we should choose the target network that provides acceptable cost and sufficient data rate for the applications running on the MN. Also, we need to minimize the energy for scanning a PoA for handoff. When the MN is moving away from the current ANO where RSS drops below a predefined threshold, it selects the target network for handoff using three metrics: cost, data rate, and energy. In our scheme, we introduce a QoS function to select an optimal network by integrating three metrics, which is given by

$$Q = W_c * \ln(C) + W_R * \ln\left(\frac{1}{R}\right) + W_E * \ln(E), \quad (4)$$

where C, R, and E represent cost, data rate, and energy. Similarly, W_c, W_R, and W_E are the weights of cost, data rate, and energy, respectively. The MN obtains the values of parameters C and R of available networks from MIIS server. The energy consumption by an interface using a particular application is computed by the MN. The weights are assigned randomly depending on the priority of application. This mean that the application with high data rate requires extra cost as compared to the application with fewer data. Similarly, every application requires different data rates depending on the nature of the application. For instance, real-time application requires more data rate as compared to an elastic application. Therefore, a target network is selected on the basis of cost, data rate, and energy which requires an interface of the MN during handover using (4).

Algorithm 5 (handover triggering and network selection).

(1) Computing Threshold θ_T

Get distance d: MN $\longleftrightarrow$ AP/BS
Compute δ
Normalized distance $\leftarrow (1 - \delta) \times d$
$K_1 \leftarrow$ path loss factor at the edge of the cell
$K_1 \leftarrow$ path loss factor at the distance d
$\theta_T = K_1 - K_2 \log((1 - \delta)d)$

Get RSS at distance $d + i$,

$$\text{where } i = \begin{cases} i + 5, \text{ where } i = 1 & \text{if PoA = AP} \\ i + 10, \text{ where } i = 1 & \text{if PoA = BS.} \end{cases} \quad (5)$$

(2) Network Selection

Cost computation:
$N_i \rightarrow C_i \rightarrow$ MIIS server
$M \leftarrow$ Number of Applications
Assigning weight to each application
$C = \sum_{i=1}^{M} w_i C_i$
Energy computation:
$P \leftarrow$ power require by interface for scanning
t_s = scanning slot
$E = \sum_{i=1}^{n} P_i \times t_s$
QoS Computation:

$$Q_i = W_c * \ln(C_i) + W_R * \ln\left(\frac{1}{R_i}\right) + W_E * \ln(E_i) \quad (6)$$

where $i = 1, \ldots, n$.

(iii) Optimal AP/BS Selection. Once a particular network is selected for handover, the next step is to select an appropriate AP/BS for handover. As we know that the competition among different networks is increasing every day. Each network is trying to provide best QoS with low cost and high data rate to their user. Similarly, every network is attempting to deploy AP/BS everywhere to provide an MN with "always best connected" functionality. To achieve similar functionality in our scheme, we implant a handover decision-making model in our proposed scheme to provide an MN with the best AP/BS in HetMANET environment. There are several decision-making schemes available in the literature. The TOPSIS decision model has remarkable applications in handover management [31–33]. Therefore, we used TOPSIS decision-making scheme to select one of the APs/BSs for handover. There are two types of criteria available for the selection of the AP/BS. The first type directly affects the performance of an AP/BS and the second type increases the performance of the AP/BS. To minimize the imbalance effect of both of these parameters, we choose only those parameters that directly affect the performance of an AP/BS. We choose five different criteria for the selection of the network. These criteria include delay (α), jitter (β), Bit Error Rate (BER) (γ), communication cost (c), and response time (σ). The decision-making matrix (M) is represented as follows:

$$M(x_{ij}) = \begin{bmatrix} \alpha_1 & \beta_1 & \gamma_1 & c_1 & \sigma_1 \\ \alpha_2 & \beta_2 & \gamma_2 & c_2 & \sigma_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_m & \beta_m & \gamma_m & c_m & \sigma_m \end{bmatrix}. \quad (7)$$

The maximum and minimum value of a parameter in a network is represented through $P_j^* = \max_{1 \le i \le m}(x_{ij})$ and $P_j^\circ = \min_{1 \le i \le m}(x_{ij})$, respectively, where P represent a parameter. It is also important to normalize the decision-making matrix. Therefore, we perform linear scaling by checking the distance of each criterion from minimum ($x_{ij}^* = x_{ij}/P_j^*$) and

maximum ($x_{ij}^{\circ} = x_{ij}/P_j^{\circ}$) values. The normalized decision-making matrix (M^*) becomes as follows:

$$M^*\left(x_{ij}\right) = \begin{bmatrix} \alpha_1^* & \beta_1^* & \gamma_1^* & c_1^* & \sigma_1^* \\ \alpha_2^* & \beta_2^* & \gamma_2^* & c_2^* & \sigma_2^* \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_m^* & \beta_m^* & \gamma_m^* & c_m^* & \sigma_m^* \end{bmatrix}. \tag{8}$$

The superscript $(*)$ is used to represent criteria after normalization.

The proposed approach is purely user based decision handover scheme. Therefore, we give the user the option to assign each criterion a particular weight. These weights help us in calculating negative and positive ideal situation of a network. In particular, an AP/BS with the more positive ideal situation is closer for the selection of handover. The weighted normalized matrix (Z) is represented as follows:

$$Z = \begin{bmatrix} w_1\alpha_1^* & w_2\beta_1^* & w_3\gamma_1^* & w_4c_1^* & w_5\sigma_1^* \\ w_1\alpha_2^* & w_2\beta_2^* & w_3\gamma_2^* & w_4c_2^* & w_5\sigma_2^* \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ w_1\alpha_m^* & w_2\beta_m^* & w_3\gamma_m^* & w_4c_m^* & w_5\sigma_m^* \end{bmatrix}. \tag{9}$$

After calculating the weighted normalized decision matrix, the next step is to compute the ideal situations, as we choose those parameters that directly affect the performance of an AP/BS. Therefore, the maximum and minimum value in each column of the matrix Z is represented through negative (S^-) and positive (S^+) ideal situations, respectively, using the following relations:

$$\begin{aligned} S^+ &= \left[\left(\min_i z_{ij} \mid j \epsilon J\right)\right] = \left[p_1^*, p_1^*, \ldots, p_5^*\right], \\ S^- &= \left[\left(\max_i z_{ij} \mid j \epsilon J\right)\right] = \left[p_1^{\circ}, p_2^{\circ}, \ldots, p_5^{\circ}\right]. \end{aligned} \tag{10}$$

To check whether these ideal situations fulfill the requirements of an appropriate AP/BS, we compare them with the reference ideal situation. Similarly, TOPSIS also ranks the available AP/BSs by comparing the ideal situations with reference situations. Therefore, we check the distance of each criterion from (S^-) and (S^+) using the followings relations:

$$\begin{aligned} H_i^+ &= \sqrt{\sum_{k=1}^{5}\left(z_{ik} - p_k^*\right)^2}, \quad i = 1, 2, \ldots, n, \\ H_i^- &= \sqrt{\sum_{k=1}^{5}\left(z_{ik} - p_k^{\circ}\right)^2}, \quad i = 1, 2, \ldots, n, \end{aligned} \tag{11}$$

where H_i^+ and H_i^- represent the degree of negative and positive ideal situations. To elaborate the structure of these situations, we illustrate it in Figure 4.

Finally, the optimal AP/BS is selected by computing the relative degree approach (R) of each AP/BS as follows:

$$R_i^* = \frac{H_i^-}{H_i^+ + H_i^-}. \tag{12}$$

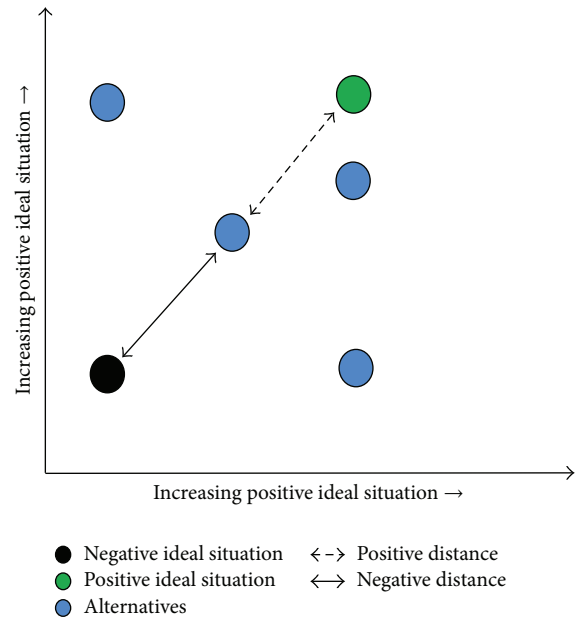


Figure 4: Explanation of ideal situation.

If there are multiple AP/BSs available in a HetMANET environment, then one can compute the degrees of each AP/BS and then sort them to select the one with the highest degree. In general, the working of the TOPSIS decision model is illustrated in Algorithm 6.

Algorithm 6 (working of TOPSIS).

(1) Initialization and normalization of matrix M:

$$M_{ij} = \frac{f_{ij}}{\sqrt{\sum_{i=1}^{k} {f_{ij}}^2}} \quad i = 1, 2, \ldots, m; \; j = 1, \ldots, n. \tag{13}$$

(2) Computing the weighted normalized matrix whose elements are

$$Z_{ij} = w_j * M_{ij}, \quad i = 1, 2, \ldots, m; \; j = 1, \ldots, n. \tag{14}$$

(3) To determine the ideal situations (positive and negative):

$$\begin{aligned} S^+ &= \min Z_{ij} \\ &\quad j \in J \mid i = 1, \ldots, \; n = \left[z_i^+, z_{i+1}^+, \ldots, z_n^+\right], \\ S^- &= \max Z_{ij} \\ &\quad j \in J \mid i = 1, \ldots, \; n = \left[z_i^+, z_{i+1}^+, \ldots, z_n^+\right]. \end{aligned} \tag{15}$$

(4) Computing the separation measure of each situation:

$$\begin{aligned} H_i^+ &= \sqrt{\sum_{k=1}^{n}\left(z_{kj} - z_k^+\right)^2}, \quad k = 1, 2, \ldots, m, \\ H_i^- &= \sqrt{\sum_{k=1}^{n}\left(z_{kj} - z_k^-\right)^2}, \quad k = 1, 2, \ldots, m. \end{aligned} \tag{16}$$

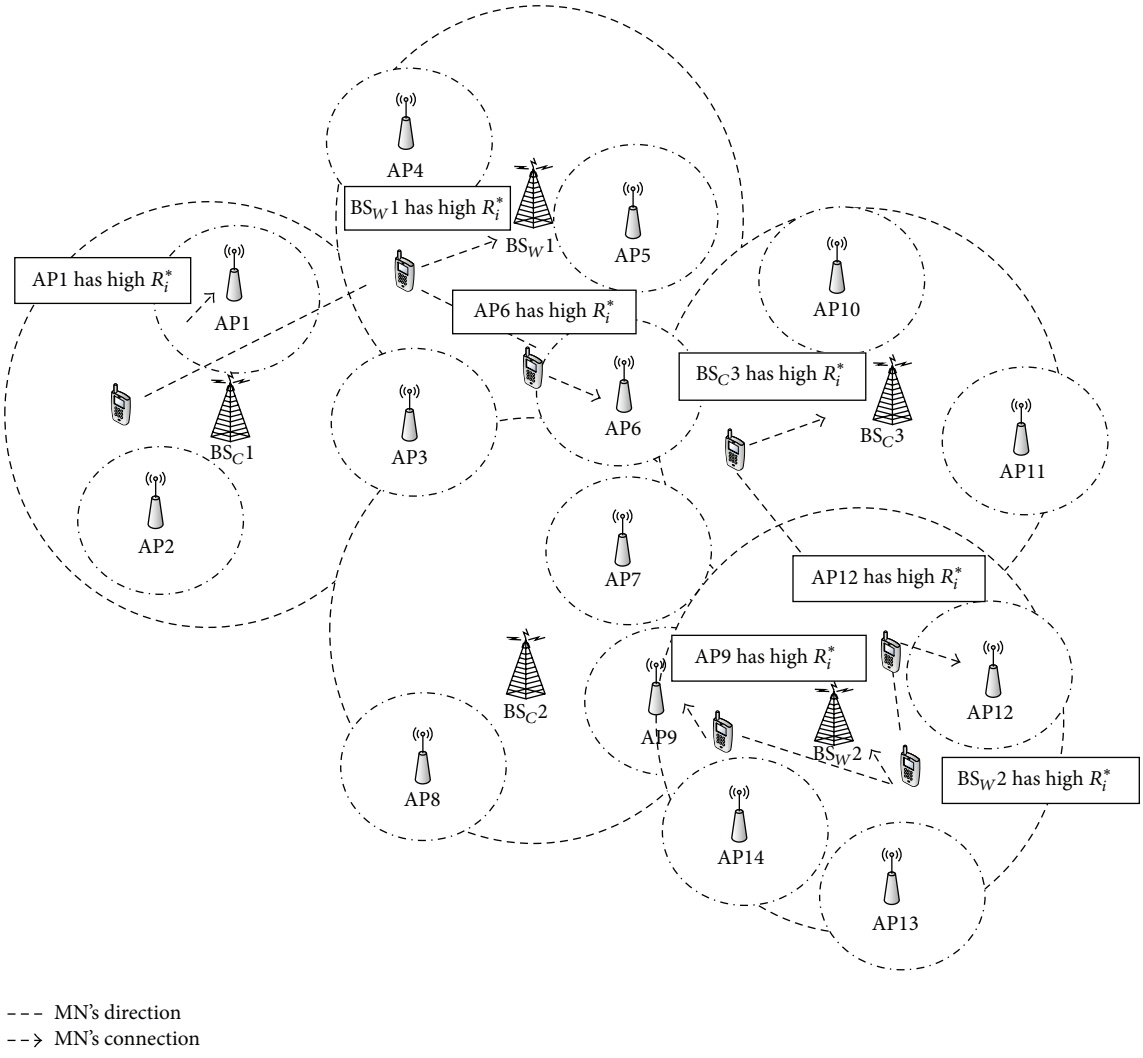


Figure 5: Proposed simulation scenario.

(5) Computing relative closeness of each criterion to ideal situation.

3.2.3. Handover Execution. The MN performs handover execution after selecting the network with the highest QoS. The MN requests the serving AP/BS to connect the network. The AP/BS forwards this request to the MIIS server. The new network sends a connection response to the MN, and the MN performs handover to the new AP/BS. The MN releases the resources and terminates the connection with the old network.

4. Performance Evaluation

In this section, we present the simulation results to highlight the benefits of the proposed handover triggering and network selection scheme. First, we show the advantages of the QoS aware network selection scheme. Second, we performed some experiments to check the handover decision model in dense and low coverage HetMANET environment. Furthermore, we evaluate the working of the proposed scheme in C programming language. The proposed approach is tested on three different networks, that is, WIFI, UMTS, and WiMAX. Different numbers of mobile nodes are tested in the proposed scenario with a speed ranging from 10 to 100 km/h. The number of applications is assigned randomly to each MN during initialization. In Figure 5, we are using only two BSs of the WiMAX network and three BSs of the cellular network because of the availability of space. In the actual simulation scenario, we used around 15 and 20 WiMAX and cellular BSs, respectively. The simulation time is set differently with the number of nodes. We test four sets of nodes, that is, 25, 50, 75, and 100, with a simulation time of 30, 60, 90, and 120 minutes, respectively. The MIH standard does not implement MIIS server in NS 2.29 V3. Therefore, we implement the MIIS server to store the cost and data rate information of the available networks. Moreover, the proposed scheme is tested

for a longer duration of time to check its performance and quality in high speed and congested scenarios. The proposed M2M communication scenario in HetMANET scenario is shown in Figure 5.

In Figure 5, the MN is initially connected with the AP1. After moving away from the AP1, the MN has found three different types of networks, that is, BSc, BSw, and AP3. In general, an MN has three different networks to decide handover to one of them. Our proposed approach enables the MN to scan the available networks and compute the QoS of each network. The MN found that BSw (WiMAX) provided the highest QoS. Therefore, the MNs choose WiMAX network for handover. The MN also uses the proposed decision model to check the available BSs of the WiMAX network. Similarly, the MN found the BSw1 with the highest degree and, therefore, the MN performs handover to it. Furthermore, the MN continues its movement in the proposed scenario. The handover is shown on the label attached to the MN in Figure 5.

The cost for UMTS network is fixed, and the costs for WIFI and WiMAX networks are generated randomly. We used two types of cost values for each network, that is, cost per minute and cost per data volume. The value of selection criteria for the network as well AP and BS is given in Table 2.

The energy consumption values by each interface are taken randomly from the ranges present in Table 2. Similarly, the RSS values are generated depending on the data rate. During the simulation, we periodically check the relation between data rate and RSS. When the data rate is increasing, the RSS is decreasing as they are in indirect proportion to each other.

We used five different parameters for the selection of AP/BS of the target network. As previously discussed, these parameters are indirectly eruptional to the performance of an AP/BS. The values of all of these parameters depend on the distance of MN and AP/BS. If the MN is away from the AP/BS, the values are high, and as the MNs move closer to the AP/BS, the values are decreasing. Initially, we do not set any particular values for these parameters. The values are changing with the distance of MN from the target AP/BS. Therefore, to achieve the optimal changing of these parameters, we implant a location management system using the coordinate geometry. The location management is simulated, and we obtain remarkable results. Finally, we do not simulate the handover execution phase and left this to the network operator.

An interface requires high energy if the MN is running an application that requires higher data rate. For example, the streaming application requires higher data rate compared to the elastic application. Therefore, a streaming application consumes more energy than an elastic application. Similarly, the MN consumes more energy on scanning if the numbers of available networks are high. Sometimes, the MN consumes unnecessary energy on scanning available networks that are far away from the MN. Therefore, in the proposed approach, we perform a dynamic sort of scanning based on the density (number of APs and BSs) of the medium. The results obtained from simulating the density based scanning are shown in Figure 6. The result shows that the energy consumption

Table 2: Simulation values in the proposed scheme.

Column heading	Column heading two	Column heading three
Data rate (kbps)	0–1200	0–1500
Energy (W)	4.0	3.0
Cost	0–30	10–50
RSS (dB)	1–20	1–20

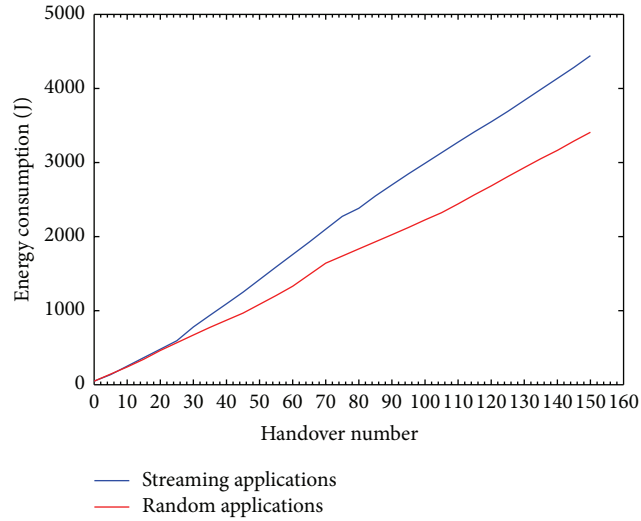


Figure 6: Application based energy consumption.

is significantly optimized by scanning a particular set of APs and BSs. Similarly, the energy that is consumed on unnecessary scanning is highly reduced. The performance of the proposed scheme shows that the energy consumed on scanning of all of the available AP/BSs of all networks is now reduced to one particular network.

We compute the device lifetime by running the simulation for a longer duration of time with the different application running on the MN's device. The MN is assigned periodically various applications. Similarly, the MN performed several handover ranging from hundreds to thousands. The device lifetime is recorded in both with and without proposed schemes. The efficient selection of the target network and AP/BS highly reduced the energy required for scanning. Therefore, the device lifetime is also increased. As shown in Figure 7, the device is consuming two types of energies (1) on Scanning and (2) running different applications.

We evaluate the energy consumption of the MN against the scanning number of APs/BSs. The MN scans a particular set of APs/BSs. Therefore, it requires less amount of energy. In Figure 8, we compared the performance of the proposed energy optimization with the scheme presented in [34]. We compute the scanning time of an interface against the number of APs/BSs. Furthermore, we compute the average energy consumption by an application using all of the three interfaces. Finally, the results are calculated and drawn in Figure 8. If an application used by the MN has high priority, then the MN needs fast scanning, which requires higher energy as compared to the application with lower priority.

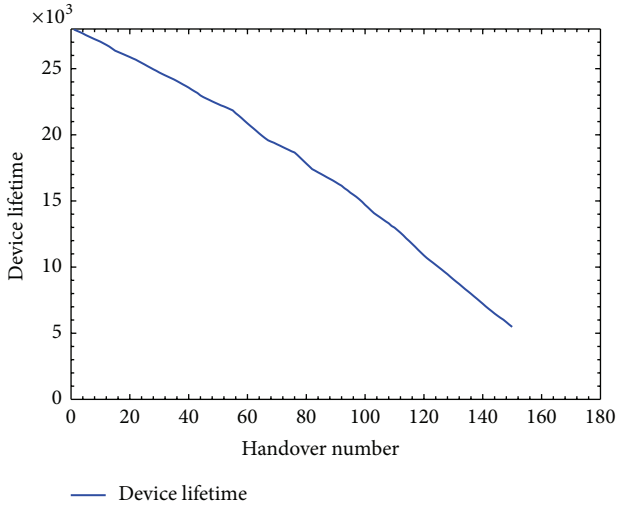


Figure 7: Device lifetime.

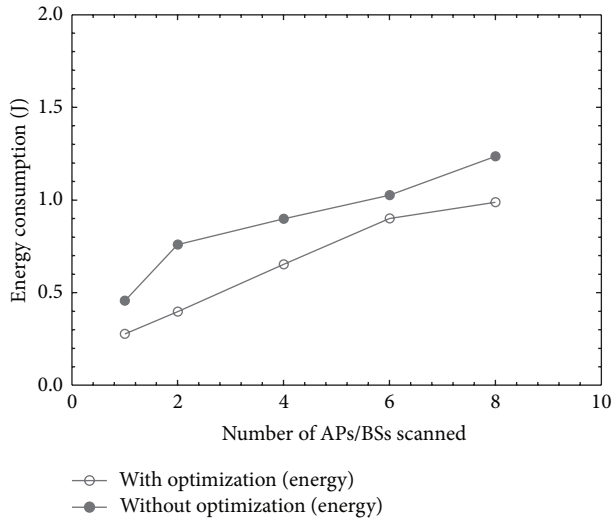


Figure 8: Performance analysis of energy optimization.

Therefore, we assign the MN the highest weight if the applications running on it have high priority and vice versa. The range of weights for energy is taken from 0 to 1.0. The proposed optimization of energy required less energy than the existing scheme due to the new power aware interface management scheme.

The AP/BS can provide services to a limited number of MNs. As the number of MNs increases on AP/BS, the QoS decreases. Therefore, we need traffic management that can efficiently balance the number of connections on a particular AP/BS. If the number of connections on AP/BS exceeds a given threshold (φ), the AP/BS should not accept any more connections. If the MN requests a connection to an overloaded AP/BS, the AP/BS rejects the connection application for the particular MN. To address the issue above,

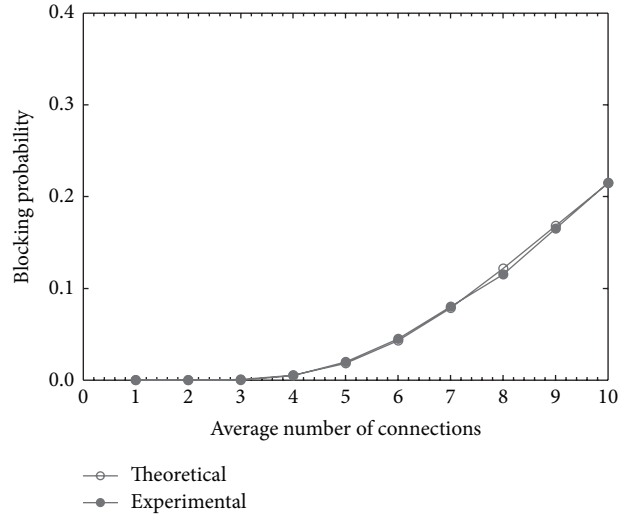


Figure 9: Theoretical and experimental blocking probability.

we modeled the traffic on an AP/BS. Let λ be the number of connections on AP/BS computed as follows:

$$\lambda = \lambda + \sum_{i=1}^{N} \tau_i \quad \text{where } \lambda < \varphi, \tag{17}$$

where γ represents the number of connections already present on an AP/BS and the τ is a new connection arriving on AP/BS. We define a two-state Markov chain model for AP/BS. In the first state, the AP/BS accepts new connection since it has vacant channels available for new connections, while in the second state the AP/BS does not accept new connections since it has no available channels for incoming connections. We called the first state as an open state, whereas the other one is a close state. The probability of a close and open state depends on the leaving and joining of new MNs, respectively.

If an AP/BS reaches a close state, then it blocks any incoming connections. The probability of blocking a new connection can be represented as follows [24, 35]:

$$P_b = \sum_{i=0}^{C} \left(1 - \beta_{i+1}\right) \times P_i, \tag{18}$$

where P_i is the probability of a channel that is either busy or available and β is the state of an AP/BS whether it is in the open or closed state. We restrict the boundaries of β to either 1 (open state) or 0 (close state). The computation of the blocking probability on an AP/BS is summarized in Algorithm 7.

We performed theoretical and experimental analysis of the proposed handover blocking probability results in Figures 9 and 10. As shown in Figure 9, theoretical and experimental results are almost similar demonstrating that increasing the number of new connections consequently increases the blocking probability. Similarly, in Figure 10, the theoretical and experimental analysis is giving similar results, with the increase in a mean number of connections, affecting the total

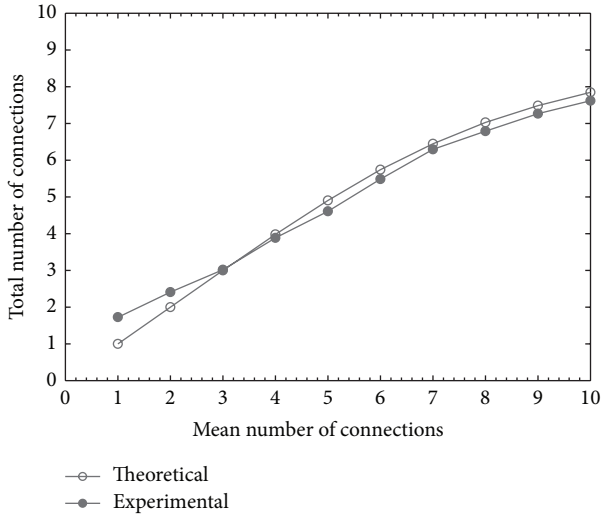


Figure 10: Mean number of connections.

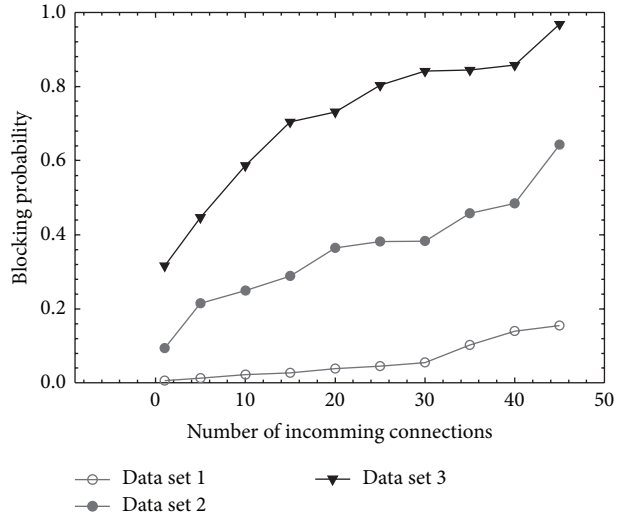


Figure 11: Performance analysis of traffic management.

number of connections. The experimental analysis has been carried in "C programming" language, while the numbers of MNs are distributed randomly using Poisson distribution.

Algorithm 7 (computation of blocking probability).

$M \leftarrow$ Maximum number of connections

$\varphi \leftarrow$ Connections threshold
$P \leftarrow$ Probability of busy or available state
$P_b \leftarrow$ blocking probability
$\beta \leftarrow$ 0 or 1 (close or open state)

While $(\lambda < \varphi)$ and (index $\leq M$) do

$\lambda = \lambda + \sum_{i=1}^{N} \tau_i$;
if $(\lambda > \varphi)$
$P_b = \sum_{i=0}^{C} (1 - \beta_{i+1}) \times P_i$
if $(P_b == 1)$
block new connections
else
accept new connection
index++

end while

The experimental and theoretical results are very close to each other which shows the accuracy of the proposed approach. The employing of blocking probability approach significantly models the traffic on an AP/BS. Moreover, the MNs do not scan those APs/BSs, which are already in a close state. Thus, using this way, the proposed approach always provides the MN with only those APs/BSs, which has channels available for the new connections.

Similarly, we test every possible probability of incoming connection on AP/BS. The probabilities are divided into three different data sets; that is, the first set is 0.1~0.3, the second set is 0.4~0.7, and the third set contains 0.8~1.0 with high probability of close state in data set 3, average close state probability in data set 2, and, similarly, low probability in data set 1. On each channel of an AP/BS, we test all the possible probabilities from set 1. Furthermore, we compute the average blocking probability of AP/BS in the meantime when there is no particular channel available for incoming connection, or it is already in use. In Figure 11, we have shown that the blocking probability is high for data set 3 since most of the channels are already occupied. Similarly, for 1 and 2 data sets, the close state is considerably small as compared to the first data set. It is also shown that the increase in the number of incoming connections increases the blocking probability. In fact, upon arrival of new connections on the AP/BS, the channels are occupied and are switched to the close state. The close state probability is evaluated by using the following equation:

$$P_{\text{Close}} = \frac{P_C \times P_N}{1 - P_H \times (1 - P_C)}, \tag{19}$$

where P_C, P_H, and P_N are the probabilities of a channel availability or unavailability, holding time distribution, and new call attempt that has been accepted, respectively.

Moreover, we investigated the handover initiation process on different distances from a target network. We gradually increase the velocity of an MN to check the variation in data rate dynamics. Each application has assigned a particular weight based on data rate requirements. The range of weights is taken from 0 to 1. An application that requires high data rate is assigned high weight (nearly equal to 1.0) and the application which requires less data rate is assigned (nearly equal to 0) lower weight. Table 2 shows the different values used in the performance evaluation of data rate optimization phase.

The velocity of an MN is checked against the application's weight. The proposed scheme performed efficiently since both the velocity of the MN and the weight of the application

Table 3: Simulation parameters used in data rate optimization.

σ	Distance from new network	Data rate in current network (Mbit/s)	Velocity of MN (km/h)
0.1	80	2	10
0.2	75	2.5	15
0.3	70	3	20
0.4	65	3.5	25
0.5	80	2	10

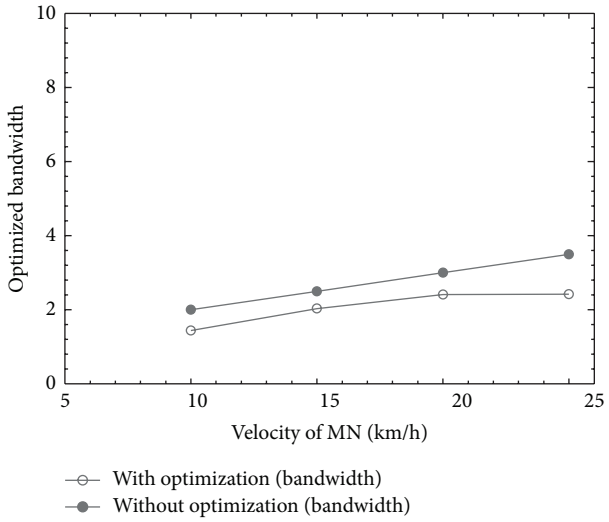


Figure 12: Performance analysis of data rate optimization.

are increased. It is due to the optimization of parameters used in Table 3 above. In Figure 12, the proposed with optimization data rate is employed against the without optimization data rate. The proposed optimization efficiently solves the high MN's velocity problem. The velocity of the MN is gradually increased, and we observed that the application that requires high data rate was shifted to the optimized data rate. Our proposed solution significantly optimized the data rate problem for the MN during the handover process.

Finally, we compute the quality of each network available in the vicinity of the MN's on the current AP/BS. The MN computes the QoS of each network using (4). The MN selects the network with the highest QoS and proceeds in contact with it. Figure 13 delineates the performance selection of an optimal network using different weights of data rate, cost, and energy ranging from 0 to 1. The different values obtained in the performance evaluation Sections 4.1, 4.2, and 4.3 are classified into three different data sets on the basis of weights assigned to data rate, cost, and energy. In data set 1, the weight of the data rate is less compared to the cost and energy, data set 2 has less weight of energy compared to the data rate and cost, and, in data set 3, the weight of the cost is smaller compared to the data rate and energy. The QoS of a network is tested against the user preference (random selection of applications from Table 1) in terms of the cost, data rate, and energy required during a handover process. The selection of

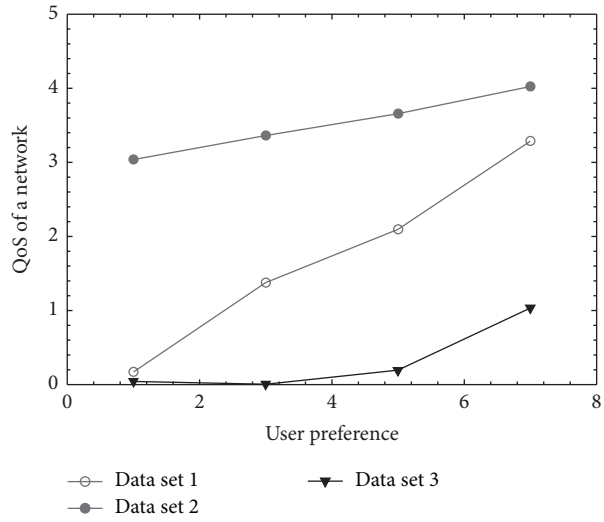


Figure 13: Optimization of QoS of a network.

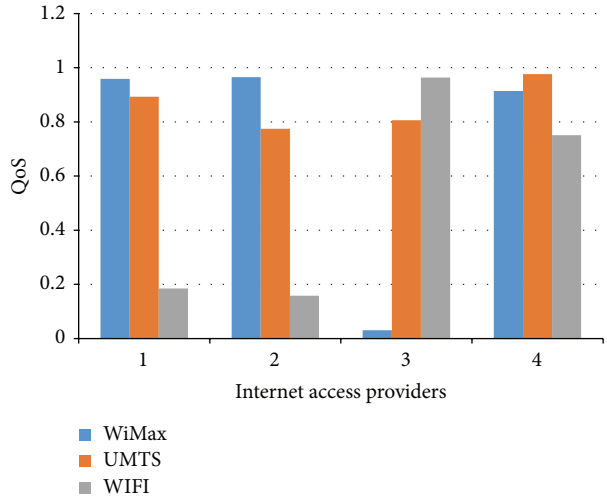


Figure 14: QoS during handover.

QoS of a network significantly achieves good results in the case of data set 3, which shows that most of the users preferred data rate compared to the cost and energy.

Moreover, we also compute the QoS of each network after each handover. The comparison of the technologies against the QoS is illustrated in Figure 14. The proposed scheme selects the available networks on the basis of the applications running on the MN's device.

5. Conclusion

In this paper, we proposed a QoS based vertical handover scheme for M2M communications in HetMANET, which represents multiparameters optimization technique for a handover process. The proposed scheme efficiently obtained the information of the communication cost of all the available networks. The scanning of the available network is performed based on the density of the AP/BS. Moreover,

the MN optimized the energy required by an interface for scanning and making a connection to the new network. The AP/BS optimized the traffic on AP/BS for providing the best connectivity and QoS to the users. The handover initiation phase is triggered by using the proposed optimal threshold scheme due to which the numbers of failed handover are significantly minimized. The optimizations of parameters above are quantified in a QoS function network. The QoS function returns the suitable network against the application used by the MN during handover. The quantitative analysis shows the accuracy and strength of the proposed scheme. For future work, we are planning to develop an optimization technique based on the decision modeling as well as fuzzy logic technique.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the IT R&D Program of MSIP/IITP (10041145, Self-Organized Software Platform (SoSp) for Welfare Devices), and it was supported by the Brain Korea 21 Plus Project (SW Human Resource Development Program for Supporting Smart Life) funded by the Ministry of Education, School of Computer Science and Engineering, Kyungpook National University, Korea (21A20131600005).

References

[1] R.-H. Gau and C.-P. Cheng, "Optimal tree pruning for location update in machine-to-machine communications," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2620–2632, 2013.

[2] M. Zorzi, A. Gluhak, S. Lange, and A. Bassi, "From today's intranet of things to a future Internet of things: a wireless- and mobility-related view," *IEEE Wireless Communications*, vol. 17, no. 6, pp. 44–51, 2010.

[3] Study on facilitating machine to machine communications in 3GPP systems, 3GPP TR 22.868.

[4] 3GPP, "System improvements for machine-type communications," 3GPP TR 23.888, V1.0.0, 2010.

[5] Service requirements for machine-type communications, 3GPP TS 22.368, V10.2.0, 2010.

[6] J. Kellokoski, J. Koskinen, R. Nyrhinen, and T. Hamalainen, "Efficient handovers for machine-to-machine communications between IEEE 802.11 and 3GPP evolved packed core networks," in *Proceedings of the IEEE International Conference on Green Computing and Communications (GreenCom '12)*, pp. 722–725, IEEE, Besançon, France, November 2012.

[7] *IEEE Standard for Local and Metropolitan Area Networks- Part 21: Media Independent Handover*, IEEE Standard, 2008.

[8] S. Thomson and T. Narten, "IPv6 stateless address autoconfiguration," IETF, RFC 2462, 1998.

[9] H. Lin and H. Labiod, "Hybrid handover optimization for multiple mobile routers-based multihomed NEMO networks," in *Proceedings of the IEEE International Conference on Pervasive Services (ICPS '07)*, pp. 136–144, Istanbul, Turkey, July 2007.

[10] M. I. Sanchez, M. Urueña, A. De La Oliva, J. A. Hernandez, and C. J. Bernardos, "On providing mobility management in WOBANs: integration with PMIPv6 and MIH," *IEEE Communications Magazine*, vol. 51, no. 10, pp. 172–181, 2013.

[11] F. Buiati, L. J. G. Villalba, D. Corujo, J. Soares, S. Sargento, and R. L. Aguiar, "Hierarchical neighbor discovery scheme for handover optimization," *IEEE Communications Letters*, vol. 14, no. 11, pp. 1020–1022, 2010.

[12] J. S. Choi, B. G. Jung, and T.-I. Kim, "LMA initiated route optimization protocol for improving PMIP handover performance," *IEEE Communications Letters*, vol. 13, no. 11, pp. 871–873, 2009.

[13] W. Yumei, F. Yufei, and Z. Lin, "Coordinating fast handover and route optimization in proxy mobile IPv6," in *Proceedings of the 5th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '09)*, pp. 1–4, IEEE, Beijing, China, September 2009.

[14] M. Zekri, B. Jouaber, and D. Zeghlache, "Context aware vertical handover decision making in heterogeneous wireless networks," in *Proceedings of the 35th Annual IEEE Conference on Local Computer Networks (LCN '10)*, pp. 764–768, IEEE, Denver, Colo, USA, October 2010.

[15] P. P. Bhattacharya and P. K. Banerjee, "Signal to interference ratio based fuzzy logic controlled call handover," in *Proceedings of the Annual IEEE India Conference*, pp. 1–5, IEEE, New Delhi, India, September 2006.

[16] Y. Yang, P. Yu, and W. Li, "Handover self-optimization mechanism based on velocity for cellular networks," in *Proceedings of the 7th International ICST Conference on Communications and Networking in China (CHINACOM '12)*, pp. 606–610, IEEE, August 2012.

[17] R. Trestian, O. Ormond, and G.-M. Muntean, "Energy-quality-cost tradeoff in a multimedia-based heterogeneous wireless network environment," *IEEE Transactions on Broadcasting*, vol. 59, no. 2, pp. 340–357, 2013.

[18] A. Paul, "Real-time power management for embedded m2m using intelligent learning methods," *ACM Transactions on Embedded Computing Systems*, vol. 13, article 148, 2014.

[19] A. Ahmad, S. Jabbar, A. Paul, and S. Rho, "Mobility aware energy efficient congestion control in mobile wireless sensor network," *International Journal of Distributed Sensor Networks*, vol. 2014, Article ID 530416, 13 pages, 2014.

[20] A. Paul and S. Rho, "Probabilistic model for M2M in IoT networking and communication," *Telecommunication Systems*, 2015.

[21] A. Paul, "Graph based M2M for IoT environment," in *Proceedings of the Research in Adaptive and Convergent Systems (RACS '13)*, pp. 45–46, ACM, Montreal, Canada, October 2013.

[22] A. Ahmad, A. Dainel, and P. Anand, "Optimized data transmission using cooperative devices in clustered D2D communication," in *Proceedings of the Conference on Research in Adaptive and Convergent Systems (RACS '14)*, ACM, Towson, Md, USA, October 2014.

[23] A. Ahmad, M. M. Rathore, A. Paul, and B.-W. Chen, "Data transmission scheme using mobile sink in static wireless sensor network," *Journal of Sensors*, vol. 2015, Article ID 279304, 8 pages, 2015.

[24] T. Wu, J. Huang, X. Yu, X. Qu, and Y. Wang, "Cost-aware handover decision algorithm for cooperative cellular relaying networks," in *Proceedings of the 67th Vehicular Technology*

Conference (VTC '08), pp. 2446–2450, IEEE, Singapore, May 2008.

[25] Q.-T. Nguyen-Vuong, N. Agoulmine, and Y. Ghamri-Doudane, "A user-centric and context-aware solution to interface management and access network selection in heterogeneous wireless environments," *Computer Networks*, vol. 52, no. 18, pp. 3358–3372, 2008.

[26] S. J. Bae, M. Y. Chung, and J. So, "Handover triggering mechanism based on IEEE 802.21 in heterogeneous networks with LTE and WLAN," in *Proceedings of the International Conference on Information Networking (ICOIN '11)*, pp. 399–403, IEEE, Barcelona, Spain, January 2011.

[27] 3GPP, "EUTRA radio resource control," 3GPP TS 36.331 V9.2.0, 3GPP, 2010.

[28] 3GPP, "Service Requirements for Machine-Type Communications," 3GPP TS 22.368 V11.0.0, 2010.

[29] D. Aziz and R. Sigle, "Improvement of LTE handover performance through interference coordination," in *Proceedings of the 69th IEEE Vehicular Technology Conference (VTC Spring '09)*, pp. 1–5, IEEE, Barcelona, Spain, April 2009.

[30] ITU, 2014, http://www.itu.int/pub/R-QUE-SG07.

[31] E. Patouni, N. Alonistioti, and L. Merakos, "Modeling and performance evaluation of reconfiguration decision making in heterogeneous radio network environments," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 1887–1900, 2010.

[32] A. Ahmed, L. M. Boulahia, and D. Gaïti, "Enabling vertical handover decisions in heterogeneous wireless networks: a state-of-the-art and a classification," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 2, pp. 776–811, 2014.

[33] H.-H. Choi, "An optimal handover decision for throughput enhancement," *IEEE Communications Letters*, vol. 14, no. 9, pp. 851–853, 2010.

[34] G. Kuhn, J. Eisl, and H. Becker, "Co-operative handover in 3G system architecture evolution," in *Proceedings of the 32nd IEEE Conference on Local Computer Networks (LCN '07)*, pp. 643–650, IEEE, Dublin, Ireland, October 2007.

[35] R. Ramjee, R. Nagarajan, and D. Towsley, "On optimal call admission control in cellular networks," in *Proceedings of the IEEE 15th Annual Joint Conference of the IEEE Computer Societies. Networking the Next Generation*, vol. 1, pp. 43–50, San Francisco, Calif, USA, March 1996.