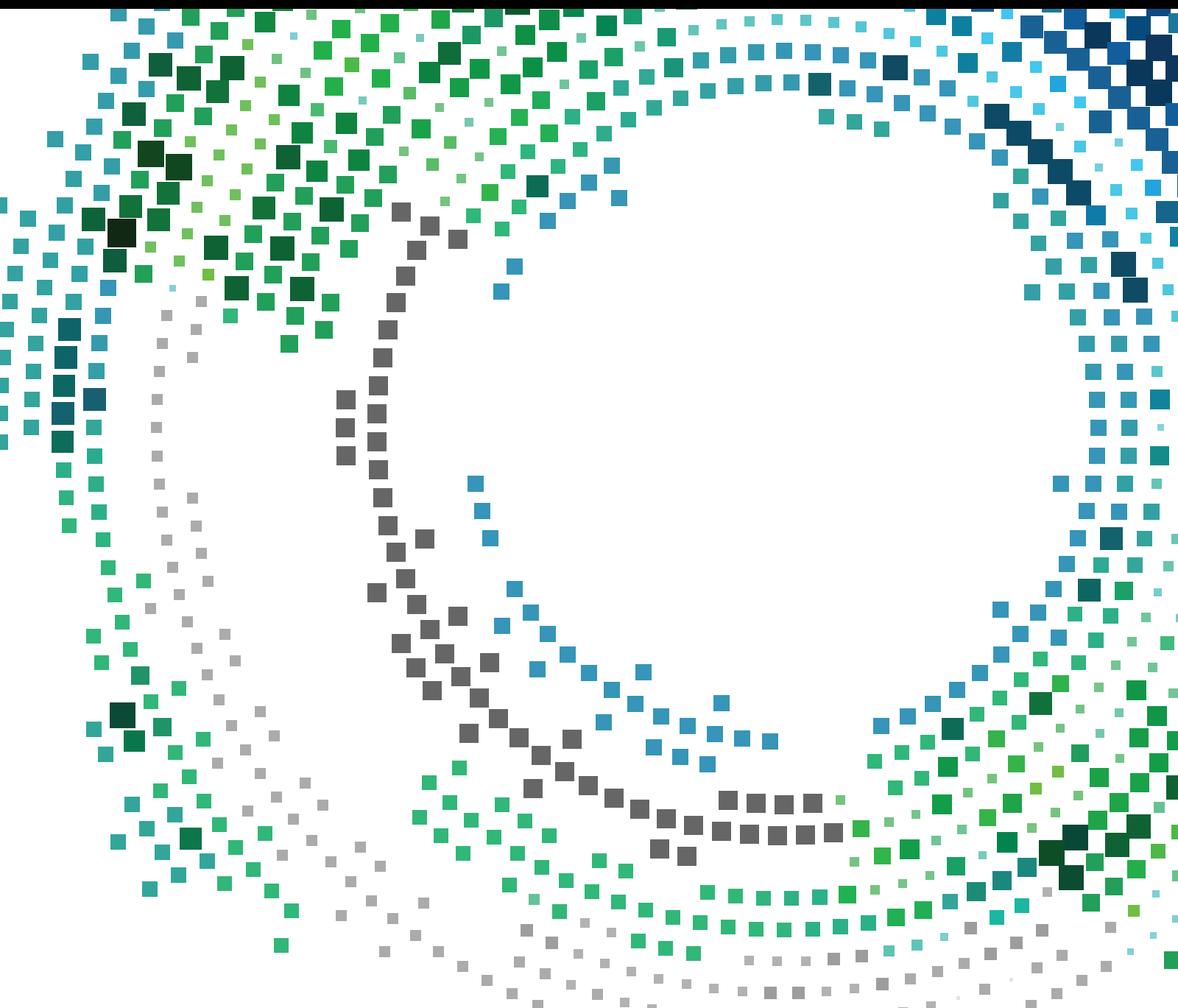# Self-Organization in Mobile Networking Systems

Guest Editors: Hyun-Ho Choi, Seok-Chul Kwon, Youngwook Ko, and Jung-Ryun Lee

Self-Organization in Mobile Networking Systems

Self-Organization in Mobile Networking Systems

Guest Editors: Hyun-Ho Choi, Seok-Chul Kwon, Youngwook Ko, and Jung-Ryun Lee

Editor-in-Chief

David Taniar, Monash University, Australia

Editorial Board

Contents

Editorial
Self-Organization in Mobile Networking Systems

Hyun-Ho Choi,[1] Seok-Chul Kwon,[2] Youngwook Ko,[3] and Jung-Ryun Lee[4]

[1]*Department of Electrical, Electronic and Control Engineering, Hankyong National University, Anseong, Republic of Korea*
[2]*Intel Corporation, Santa Clara, CA, USA*
[3]*School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast, UK*
[4]*School of the Electrical Engineering, Chung-Ang University, Seoul, Republic of Korea*

Correspondence should be addressed to Jung-Ryun Lee; jrlee@cau.ac.kr

Current mobile communication and networking systems, such as cellular network, wireless local area network, ad hoc network, and vehicular network, require manual configuration and centralized control for both deployment and operation. Such a paradigm would not be so feasible in future mobile networking systems, where a tremendous number of embedded devices and everyday items will be interconnected by various wireless technologies. One of the important design features for these future systems is the so-called self-x functions including self-configuration, self-optimization, self-adaption, self-healing, and self-protection. These self-organizing functions can reduce the network deployment time, save the operational costs, enhance the network performance, and facilitate network management without human involvement. For this reason, the self-organization in networks has become an interesting research topic and still needs to be investigated for future mobile networking systems.

This special issue contains 11 papers selected from submissions to the open call for papers. These papers highlight some of the current research interests and achievements in the area of self-organizing algorithms, protocols, and applications in various mobile networking systems.

The paper by J. Park and Y. Lim is entitled "Adaptive Access Class Barring Method for Machine Generated Communications." This paper proposes an adaptive access barring method for machine type communication (MTC) in 3GPP cellular networks. Since an eNB does not know the number of MTC devices in its coverage area, it is difficult to control the barring factor by predicting the number of MTC devices in a service area of a cell. Therefore, the authors have controlled the barring factor based on the prediction of access intensity which can be measured at an eNB. Because the proposed method can manipulate the barring factor autonomously according to the access intensity, it is superior to the original method in terms of the access success probability and the collision probability.

The paper by K. Lee and J. Ko is entitled "Adaptive Power Allocation and Splitting with Imperfect Channel Estimation in Energy Harvesting Based Self-Organizing Networks." The authors have introduced an adaptive power allocation and splitting (APAS) scheme which takes imperfect channel estimations in consideration, whereas previous researches in RF-based information and energy transfer build up on the assumption that perfect channel estimation is easily achievable. The proposed APAS scheme achieves near-optimal performances for transferring energy and data over a single RF transmission.

The paper by J. M. Arana et al. is entitled "Random-Access Technique for Self-Organization of 5G Millimeter-Wave Cellular Communications." This paper proposes two different types of random access preambles (RAPs) to reduce the processing time in the random access stage. Analyses of the correlation property, false-alarm probability, and detection probability verify that the proposed RAPs are suitable for random access in mm wave cellular systems with directional beams because of the smaller processing time and high detection probability in multiuser environments.

The paper by W. Wang et al. is entitled "An Adaptive Information Quantity-Based Broadcast Protocol for Safety Services in VANET." This paper proposes an information quantity-based broadcast protocol to ensure the efficiency

of safety messages dissemination. The presented emergency-degree-based broadcast (EDCast) protocol differentiates each packet's priority for accessing the channel based on its emergency degree so as to provide vehicles with more safety information timely and accurately. In addition, an adaptive scheme is presented to ensure fast dissemination of messages in different network condition. The proposed EDCast achieves higher broadcast efficiency and less redundancy with less delivery delay so that it is feasible and necessary for incorporating information quantity of messages in designing an efficient safety message broadcast protocol.

The paper by F. Zhou et al. is entitled "Utility Maximization for Load Optimization in Cellular/WLAN Interworking Network Based on Generalized Benders Decomposition". In this paper, a utility maximization (UTMAX) optimization model and an ASRAO algorithm based on generalized Benders Decomposition is presented to investigate load optimizing from a perspective of system utilization maximization. The authors show a tradeoff between improvements in user throughput fairness and system total throughput. Since UTMAX requires a mixed integer nonlinear programming, which is intractable intuitively, the authors proposed ASRAO to solve this problem optimally and effectively. An optional phase for expediting ASRAO is proposed by using relaxation and approximation techniques, which reduces nearly 10% iterations and time needed by normal ASRAO from simulation results. The results also show UTMAX's good effects on improving WLAN usage and edge user throughput.

The paper by J. A. Fernández-Segovia et al. is entitled "A Fast Self-Planning Approach for Fractional Uplink Power Control Parameters in LTE Networks". In this work, the authors presented the self-planning of Uplink Fractional Power Control (FPC) settings. For this purpose, the authors formulated the FPC planning problem in a cell basis through the combination of multiple regular scenarios built on a per-adjacency basis. A detailed inspection on the FPC parameter values to identify the most important variables in the scenario impacting optimal FPC settings follows. And regression equations are built based on key variables for a simple FPC parameter calculation. Results show that network performance with proposed FPC parameter settings shows better performance compared with typical FPC configurations from operators.

The paper by E. Cheng et al. is entitled "A TDoA Localization Scheme for Underwater Sensor Networks with Use of Multilinear Chirp Signals." This paper takes the Multilinear Chirp (MLC) signals as the location signal to improve the anticollision ability and proposes a fast efficient detection method called Mixing Change Rate-Fractional Fourier Transform (MCR-FrFT) in order to increase the detection efficiency of MLC. The proposed method transforms the combined rates of MLC into symmetry triangle rates and then separates the multiuser signals based on the transformed rates by using FrFT. The proposed method can detect the locations signals, estimate the time difference of arrival, reduce the multiple access interference, and improve the location performance.

The paper by J. Paek et al. is entitled "A Measurement Study of BLE iBeacon and Geometric Adjustment Scheme for Indoor Location-Based Mobile Applications." This paper performs empirical measurements on three different Bluetooth Low-Energy (BLE) iBeacon devices for indoor location-based mobile applications. This measurement result shows that the signal strength readings from iBeacon vary significantly over different vendors, mobile platforms, environmental or deployment factors, and usage scenarios. From this measurement result, the authors proposed a simple class attendance checking application by performing a simple form of geometric adjustments to compensate for the natural variations in beacon signal strength readings.

The paper by J. Shim and Y. Lim is entitled "WSN-Based Height Estimation of Moving Object in Surveillance Systems." This paper considers the physical height to differentiate an intruder from detected objects in the wireless sensor network-based surveillance system. Using the measured information from sensors, the authors estimated the height of the detected object. Based on the height, if the detected object is decided as an intruder, an alarm is given to a control center. The proposed mechanism correctly and fast estimates the height of the object without complex computation.

The paper by O. Jo et al. is entitled "Intergroup Joint Scheduling for Mitigating Asymmetric Uplink Interference in Self-Organizing Virtual Cell Networks." This paper introduces the concept of self-organizing Virtual Cell Network (VCN). In this self-organizing VCN, the authors proposed an efficient scheduling algorithm that considers the asymmetry of interference between downlink and uplink to mitigate intercell interference with little computing overhead. The basic concept is to construct scheduling groups that consist of several users. Each user in a scheduling group is affiliated with a different cell. Then, the intercell groups are managed efficiently in the proposed VCNs. The proposed scheduling method does not require the exchange of a lot of information among base stations to schedule the users over the entire network.

The paper by W. Lee and B. C. Jung is entitled "Improving Energy Efficiency of Cooperative Femtocell Networks via Base Station Switching Off." This paper proposed BS switching-off technique for a cooperative femtocell network in which multiple femtocell BSs (FBSs) simultaneously send packets to the same mobile station. The authors formulated the optimization problem to find the optimal set of FBSs to be turned off and proposed a suboptimal scheme operating in a distributed manner in order to reduce the computational complexity of the optimal scheme. Results showed that the energy consumption is reduced compared with conventional schemes and the suboptimal scheme achieved the near-optimal performance with lower computational complexity.

Acknowledgments

Research Article

Improving Energy Efficiency of Cooperative Femtocell Networks via Base Station Switching Off

Woongsup Lee[1] and Bang Chul Jung[2]

[1]*Department of Information and Communication Engineering, Gyeongsang National University, Jinju, Republic of Korea*
[2]*Department of Electronics Engineering, Chungnam National University, Daejeon, Republic of Korea*

Correspondence should be addressed to Bang Chul Jung; bangchuljung@gmail.com

Recently, energy efficiency (EE) of cellular networks has become an important performance metric, and several techniques have been proposed to increase the EE. Among them, turning off base stations (BSs) when not needed is considered as one of the most powerful techniques due to its simple operation and effectiveness. Herein, we propose a novel BS switching-off technique for cooperative femtocell networks where multiple femtocell BSs (FBSs) simultaneously send packets to the same mobile station (MS). Unlike conventional schemes, cooperative operation of FBSs, also known as coordinated multipoint (CoMP) transmission, is considered to determine which BSs are turned off in the proposed technique. We first formulate the optimization problem to find the optimal set of FBSs to be turned off. Then, we propose a suboptimal scheme operating in a distributed manner in order to reduce the computational complexity of the optimal scheme. The suboptimal scheme is based on throughput ratio (TR) which specifies the importance of a particular FBS for the cooperative transmission. Through simulations, we show that the energy consumption can be greatly reduced with the proposed technique, compared with conventional schemes. Moreover, we show that the suboptimal scheme also achieves the near-optimal performance even without the excessive computations.

1. Introduction

In recent days, environmental pollution has become one of the biggest threats to mankind. The generation of electricity is one of the major sources of environmental pollution due to the use of fossil fuel or nuclear energy [1]. Given that the amount of electricity consumed by telecommunication systems (TSs) is considerably huge, a lot of efforts have been made to reduce the electricity usage in TSs such as the use of renewable electricity generation in base station (BS) [2, 3]. The TSs with less electricity usage, that is, green cellular networks (GCNs), will be more important in the near future because the amount of electricity consumption is expected to increases nearly 20% per year [4].

One way to reduce the electricity consumption by TSs is to improve the energy efficiency of TS. Particularly, it is crucial to increase the energy efficiency of BS, since it consumes much more electricity compared to other network entities (more than half of total electricity consumption in TSs is caused by BSs), for example, mobile stations (MSs). For this reason, various methods have been taken into account in order to improve the energy efficiency of BS [5, 6].

Most powerful yet simple way to improve the energy efficiency of BS is to halt the operation of redundant BSs. However, the set of BSs which are switched off should not be selected randomly but it should be carefully chosen because the shutdown of BSs can severely deteriorate the performance of cellular network (CN), and it can also cause the outage of mobile services. Therefore, the shutdown of BSs has recently been the subject of extensive investigation [3, 7–14].

First, in [7, 8], the shutdown of macro BSs was considered. Specifically, in [7], the shutdown of macro BSs which are located at rural area during night time, that is, the statistical shutdown scheme where network operators manually decide which BSs to be switched off, was considered and in [8] the energy efficiency and coverage probability of macro
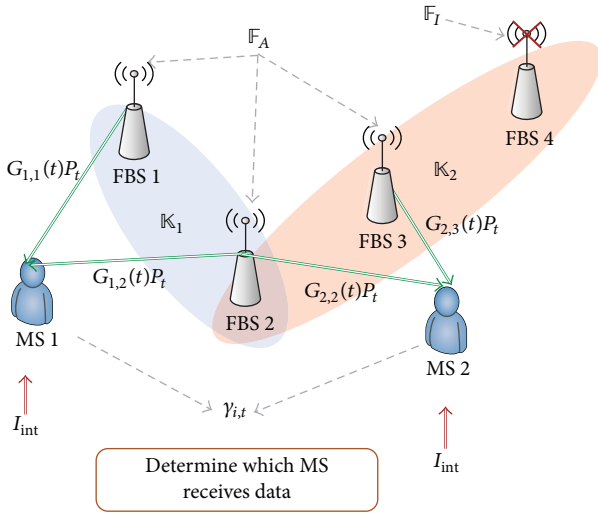
FIGURE 1: System model that shows MSs and FBSs utilizing cooperative transmission.

BS shutdown were investigated. However, the operation of macro BS is hard to be terminated in an adaptive way, according to real-time network condition, for example, traffic load, because the coverage of macro BS is wide such that the impact of switching off will be enormous. Accordingly in this paper we only take into account the shutdown of femto-BSs (FBSs) which have relatively small coverage area such that the impact of shutdown is rather limited to narrow area.

The shutdown of FBSs has been extensively investigated in recent days [9–13]. The author of [9] proposed the adaptive shutdown of indoor FBSs in which the operation of redundant FBSs is stopped by examining the achievable throughput. In [10], the operation of FBSs which are assigned small number of subcarriers was halted since the traffic load of these FBSs is low such that the shutdown of these FBSs is unlikely to deteriorate the performance of CN severely. Finally, the distance between the FBS and MSs is taken into account in the determination of the set of FBSs to shut down; that is, FBSs which are far from MSs are turned off since the contribution of these FBSs to the performance of CN is likely to be minor.

Usually, the traffic load on each FBS is used as a metric in determining FBSs to be shut down; that is, FBS with low traffic load is more likely to be turned off [8–10] because the traffic load indicates the importance of FBS in network. However, the use of traffic load can be inappropriate when multiple FBSs cooperate in data transmission; that is, cooperative transmission is considered, especially when coordinated multipoint (CoMP) [15, 16] is used. To be more specific, in the cooperative transmission, the incurred traffic load does not necessarily indicate the significance of FBSs in network.

For example, let us assume the scenario in which three FBSs (in Figure 1, FBS 4 is not participating in data transmission) (FBS 1, FBS 2, and FBS 3) are cooperatively transmitting data to two MSs (MS 1 and MS 2); compare Figure 1. Moreover, assume that FBS 1 and FBS 3 are close to MS 1 and MS 2, respectively, and FBS 2 is in between two FBSs such that FBS 1 and FBS 2 cooperatively transmit data to MS 1 and FBS 2

and FBS 3 cooperatively transmits data to MS 2. Furthermore, assume that FBS 2 is farther away from MS 1 and MS 2 compared with FBS 1 and FBS 3, respectively. In this scenario, although the traffic load of FBS 2 is larger than that of other two FBSs, it is more beneficial to turn off the FBS 2 because it unlikely deteriorates the throughput of network severely.

Therefore, new metric is needed in the determination of FBSs which utilize cooperative transmission. Given that FBSs are likely to cooperate with each other due to low transmit power, it is crucial to devise the efficient shutdown of FBS with cooperative transmission. Although [14] considered the shutdown of FBSs which use CoMP, FBSs which belong to the same cooperative transmission set cannot be selectively turned off, such that the network performance cannot be optimized. Moreover, the joint use of shutdown and CoMP was considered in [13]; however, BSs to be switched off are determined without taking into account the CoMP.

(A) Our Contributions. We herein propose an efficient shutdown scheme of FBSs considering cooperative transmission. The main contributions are as follows:

(1) In our proposed shutdown scheme, the operation of FBSs which uses cooperative transmission is efficiently halted with low computational complexity in distributed manner. Unlike previous approaches which mainly focus on the use of traffic load, we propose to use throughput ratio (TR) in determining the set of FBSs to be shut down such that the data rate of cooperative transmission is not overly degraded. To the best of our knowledge, this paper is the first study to use TR in the FBS shutdown.

(2) We investigate the performance of our proposed scheme using simulation. The results show that our proposed scheme achieves almost the same performance as the scheme based on the optimization problem, while the number of computations can be significantly reduced which makes our proposed scheme more practical. It is also shown that the power consumption of our proposed scheme is far below compared to that of the conventional scheme.

The remainder of the paper is organized as follows. In Section 2, we describe our system model and the optimization problem to find optimal set of FBSs to be turned off is formulated in Section 3. The proposed heuristic scheme is explained in Section 4 which is followed by the simulation results in Section 5. Finally, in Section 6, we provide our conclusions.

2. System Model

Our system model is depicted in Figure 1. In this paper, we consider the downlink of CN in which multiple FBSs cooperatively transmit the same data to MS on single channel, for example, CoMP [15, 16]. We assume that multiple FBSs and MSs coexist. The sets of all FBSs, active FBSs, and inactive FBSs are denoted as $\mathbb{F}$, $\mathbb{F}_A$, and $\mathbb{F}_I$, respectively. For example, in Figure 1, FBS 1, FBS 2, and FBS 3 are in $\mathbb{F}_A$ while FBS 4 belongs to $\mathbb{F}_I$.

We assume that the set of MSs is denoted as $\mathbb{M}$ and just represents the set of FBSs which can cooperate in data transmission to MS i; that is, not all FBSs are able to participate in data transmission to MSs. For example, in Figure 1, only FBS 1 and FBS 2 are in $\mathbb{K}_1$ and hence FBS 3 cannot transmit data to MS 1. Note that a specific clustering algorithm to determine $\mathbb{K}_i$ is not taken into account in this work; that is, any clustering algorithm can be used. Moreover, we let P_t be the transmit power of FBSs, where the transmit power control is not considered. Furthermore, the time-varying channel gain is denoted as $G_{i,j}(t)$, where i is the index of MS, j is the index of FBS, and t is the time. It should be noted that $G_{i,j}(t)$ comprises both path-loss and fast fading. In addition, the bandwidth and the noise density are denoted as W and N_0, respectively. In this paper, we assume that only one MS can receive data at the same time (this assumption is reasonable because FBSs within close proximity are unlikely to transmit data to multiple MSs simultaneously on the same channel, due to the excessive interference among them [15]), such that the interference among cooperative transmission is not considered. In addition, the amount of interference caused by other communication entities, for example, macro BS, is modeled as I_{int}; compare Figure 1. Finally, we assume that the power consumption of active FBS is $P_A > 0$ and that of turned-off FBS is 0.

We take into account a simple cooperative transmission scheme in which collaborating FBSs simply transmit the same signal to the target MS [15]. Then, the average throughput of MS$_i$, $R_i(\mathbb{F}_A)$, can be formulated as follows:

$$R_i\left(\mathbb{F}_A\right) = \mathbb{E}\left[\gamma_{i,t} W \log_2\left(1 + \frac{\sum_{j \in \{\mathbb{F}_A \cap \mathbb{K}_i\}} G_{i,j}(t) P_t}{N_0 W + I_{\text{int}}}\right)\right], \quad (1)$$

where $\gamma_{i,t}$ is the indicator (given that multiple MSs cannot receive data at the same time in our system model, $\sum_{i=1}^{|\mathbb{M}|} \gamma_{i,t} = 1$) which shows whether MS$_i$ receives data at time t; that is, $\gamma_{i,t} = 1$ when FBSs transmit data to MS$_i$ at time t and $\gamma_{i,t} = 0$ otherwise. Moreover, $\mathbb{E}[\cdot]$ is the expectation. In (1), it is worth noting that only the signal power from active FBSs is accumulated.

Moreover, the average energy consumption of CN when the set of active FBSs is $\mathbb{F}_A$, which we denote as $E(\mathbb{F}_A)$, can be written as follows:

$$E\left(\mathbb{F}_A\right) = P_A \left|\mathbb{F}_A\right|, \quad (2)$$

where $|\cdot|$ denotes the cardinality of the set; for example, $|\mathbb{M}|$ corresponds to the number of all MSs. Note that when none of FBSs is turned off, the average throughput of MS$_i$ and the average energy consumption of CN will become $\mathbb{E}[\gamma_{i,t} W \log_2(1 + (\sum_{j \in \mathbb{K}_i} G_{i,j}(t)P_t)/(N_0 W + I_{\text{int}}))]$ and $P_A |\mathbb{F}|$, respectively, because $\mathbb{F}_A$ becomes $\mathbb{F}$.

Finally, we can derive the average traffic load of FBS$_j$, when the active set of FBSs is $\mathbb{F}_A$, which we denote as $U_j(\mathbb{F}_A)$, as follows:

$$U_j\left(\mathbb{F}_A\right) = \mathbb{E}\left[\sum_{i \in \mathbb{M}} R_{i,t}\left(\mathbb{F}_A\right) \mathbb{1}_{j \in \mathbb{K}_i} \mathbb{1}_{j \in \mathbb{F}_A}\right], \quad (3)$$

where $\mathbb{1}(x)$ is the indicator function whose value is 1 when x is true and 0 otherwise.

3. Problem Formulation

The optimization problem to find the optimal set of switched-off FBSs can be formulated based on our system model. The objective of the formulated problem is the minimization of the total electricity consumption of FBSs, $E(\mathbb{F}_\mathbb{A})$, while guaranteeing the performance degradation of CN to be less than the predefined threshold. To be more specific, we assume that the total network throughput should be larger than the predefined threshold, R_{thr}. Then, the optimization problem to find $\mathbb{F}_A$ can be formulated as follows:

$$\begin{aligned} \underset{\mathbb{F}_A \subset \mathbb{F}}{\text{minimize}} \quad & E\left(\mathbb{F}_\mathbb{A}\right) \\ \text{s.t.} \quad & \sum_{i \in \mathbb{M}} R_i\left(\mathbb{F}_A\right) \geq R_{\text{thr}}. \end{aligned} \quad (4)$$

It is worth noting that problem (4) is a binary programming such that it is nonconvex and the solution cannot be found within short computation time, especially, when the number of FBSs is large; that is, $|\mathbb{F}| \gg 0$. In order to solve this excessive computational complexity problem, we propose a heuristic way to find the optimal $\mathbb{F}_A$ within polynomial time in Section 4.

Note that the optimization problem to find $\mathbb{F}_A$ with conventional shutdown scheme, where FBSs with low traffic load are turned off, can be formulated as follows:

$$\begin{aligned} \underset{\mathbb{F}_A \subset \mathbb{F}}{\text{minimize}} \quad & E\left(\mathbb{F}_\mathbb{A}\right) \\ \text{s.t.} \quad & \sum_{i \in \mathbb{M}} R_i\left(\mathbb{F}_A\right) \geq R_{\text{thr}} \\ & U_a\left(\mathbb{F}\right) \geq U_b\left(\mathbb{F}\right), \quad \forall a \in \mathbb{F}_A, \ b \in \mathbb{F}_\mathbb{I}. \end{aligned} \quad (5)$$

In (5), the last constraint guarantees the FBSs with low traffic load to be switched off. It should be noted that, due to this last constraint, the optimal solution of (5) can be found within polynomial time unlike optimization problem (4). To be more specific, the optimal solution can be found by turning off FBS in ascending order of traffic load until the constraint on data rate is satisfied. However, the solution will work poorly in practice when the cooperative transmission is considered, as shown in our performance evaluation.

4. Proposed Scheme

Herein, we have proposed a heuristic scheme based on TR, which indicates the contribution of certain FBS in the cooperative transmission, to solve the intractability problem of optimization problem (4). To be more specific, the TR value of FBS$_j$, which we denote as TR$_j$, can be defined as follows:

$$\begin{aligned} & \text{TR}_j \\ & = \sum_{i \in \mathbb{M}} \frac{\mathbb{E}\left[\gamma_{i,t} W \log_2\left(1 + G_{i,j}(t) P_t / \left(N_0 W + I_{\text{int}}\right)\right)\right]}{R_i\left(\mathbb{F}\right)}. \end{aligned} \quad (6)$$

Note that, in (6), numerator corresponds to the expected achievable data rate of MS$_i$ when FBS$_j$ is the only FBS

```
(1)   Set $\mathbb{F} = \mathbb{F}_A$ and $\mathbb{F}_I = \emptyset$
(2)   Calculate $\text{TR}_j$ based on (6)
(3)   repeat
(4)      Turn-off $\text{FBS}_j$ after delay $D(\text{TR}_j)$
(5)      $\mathbb{F}_A = \mathbb{F}_A \setminus \{j\}$ and $\mathbb{F}_I = \mathbb{F}_I \cup \{j\}$
(6)   until $\sum_{i \in \mathbb{M}} R_i(\mathbb{F}_A) < R_{\text{thr}}$
(7)   $\mathbb{F}_A = \mathbb{F}_A \cup \{j\}$ and $\mathbb{F}_I = \mathbb{F}_I \setminus \{j\}$
```

ALGORITHM 1: FBS shutdown based on TR.

TABLE 1: Simulation parameters.

Parameters	Values
Number of MSs	2
Path-loss	$34.5 + 38 \log_{10}(d_{i,j})$ (IMT-R M.2135 urban macromodel [17])
I_{int}	-70 dBm
N_0	-174 dBm/Hz
Bandwidth (W)	10 MHz
$\gamma_{1,t}$	0.5
$\gamma_{2,t}$	0.5
Transmit power of FBS	23 dBm

FIGURE 2: Data rate versus ϵ when $|\mathbb{F}| = 10$.

to be activated. Accordingly, if the operation of FBS_j is important in view of cooperative transmission, the value of TR_j will be high. For example, if FBS_j is the only FBS which transmits data to MS_i, then $\sum_{i \in \mathbb{M}}(\mathbb{E}[\gamma_{i,t}W \log_2(1 + G_{i,j}(t)P_t/(N_0W + I_{\text{int}}))]/R_i(\mathbb{F})) = 1$; however, if FBS_j is not participating in transmission to MS_i, that is, $i \notin \mathbb{K}_i$, then $\sum_{i \in \mathbb{M}}(\mathbb{E}[\gamma_{i,t}W \log_2(1 + G_{i,j}(t)P_t/(N_0W + I_{\text{int}}))]/R_i(\mathbb{F})) = 0$.
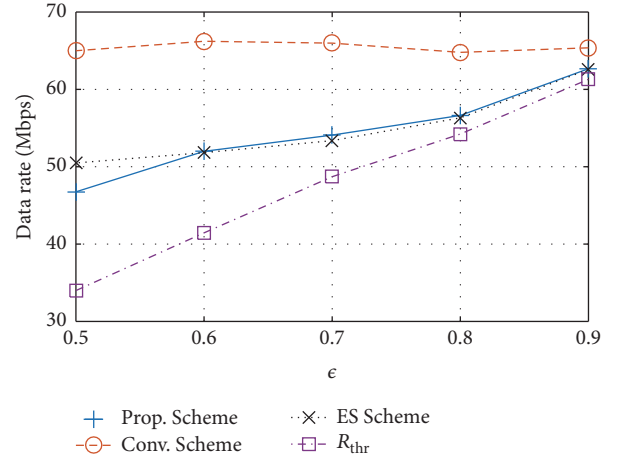
In our proposed scheme, the TR value of FBSs is used to decide which FBS to be switched off, such that FBS_j with low TR_j will be turned off. More specifically, FBSs are switched off based on the ascending order of TR_j. The shutdown of FBSs can be performed in distributed manner because each FBS can calculate its own TR_j based on its own channel condition and information from central management unit (CMU) which manages the operation of FBSs. Let $D(x)$ be the increasing function of x; for example, $D(x) = x^2$. Then, the operation of FBS_j can be halted after the delay of $D(\text{TR}_j)$ and CMU can cease the shutdown of remaining active FBSs when $\sum_{i \in \mathbb{M}} R_i(\mathbb{F}_A) < R_{\text{thr}}$. The last turned-off FBS should be immediately turned on to guarantee our constraint. It is worth noting that the overhead of our proposed scheme will be negligible compared to that of the optimal scheme which can be found by directly solving (4). In summary, the overall procedures of the proposed algorithm are described in Algorithm 1.

5. Performance Evaluation

In this section, we investigate the performance of our proposed FBS shutdown scheme using simulation. In the performance evaluation, we assume that N FBSs are randomly deployed and they transmit data using CoMP to two MSs which are located at the $(0, 0)$ and $(50\,\text{m}, 0)$; that is, $|\mathbb{M}| = 2$. The distribution of FBSs is uniform, and $10\,\text{m} < x < 40\,\text{m}$ and $-15\,\text{m} < y < 15\,\text{m}$, where (x, y) is the position of FBS; that is, FBSs are uniformly distributed within a square area. We assume that FBS whose value of x is less than 30 m belongs to set $\mathbb{K}_1$ and FBS whose value of x is larger than 20 m is within set $\mathbb{K}_2$, such that only FBSs whose x are within $20\,\text{m} < x < 30\,\text{m}$ can transmit data to both MSs (we have used this simulation scenario with only two MSs, in order to focus on the difference between conventional switching-off scheme and our proposed scheme; performance evaluation in more general environment, where more MSs coexist, will be an interesting extension of this work). IMT-R M.2135 urban macromodel (therefore, the path-loss becomes $34.5 +$

$38 \log_{10}(d_{i,j})$, where $d_{i,j}$ is the distance between MS_i and FBS_j) is used for path-loss [17]. Furthermore, we assume that $I_{\text{int}} = -70$ dBm, $N_0 = -174$ dBm/Hz, W is 10 MHz, $\gamma_{1,t} = \gamma_{2,t} = 0.5$, and the transmit power of FBS, P_t, is 23 dBm. In addition, we assume that the channel experiences the fast fading which is modeled as a circularly symmetric complex Gaussian random variable with mean 0 and variance 1. Simulation parameters are summarized in Table 1.

In the performance evaluation, we examine the performance of our proposed scheme based on Algorithm 1 (Prop. Scheme) and also consider the performance of optimal scheme which can be found by solving (4) with exhaustive search (ES Scheme). Moreover, the conventional shutdown scheme, in which FBSs are switched off based on their traffic load, is also taken into account (Conv. Scheme), where FBS to be shut down is randomly chosen when multiple FBSs have the same traffic load. Note that the shutdown of Conv. Scheme is adjusted such that $\sum_{i \in \mathbb{M}} R_i(\mathbb{F}_A) \geq R_{\text{thr}}$ is satisfied. Finally, in the simulation, we assume that $R_{\text{thr}} = \epsilon \sum_{i \in \mathbb{M}} R_i(\mathbb{F})$, where $0 \leq \epsilon \leq 1$, such that ϵ indicates the degree of throughput deterioration that the system can tolerate. In other words, low value of ϵ means that decreasing power consumption is more important such that more FBSs will be turned off even if the throughput of the system is decreased significantly.

In Figures 2 and 3, the data rate and power consumption ratio, by varying ϵ, are shown when the number of FBSs, $|\mathbb{F}|$,
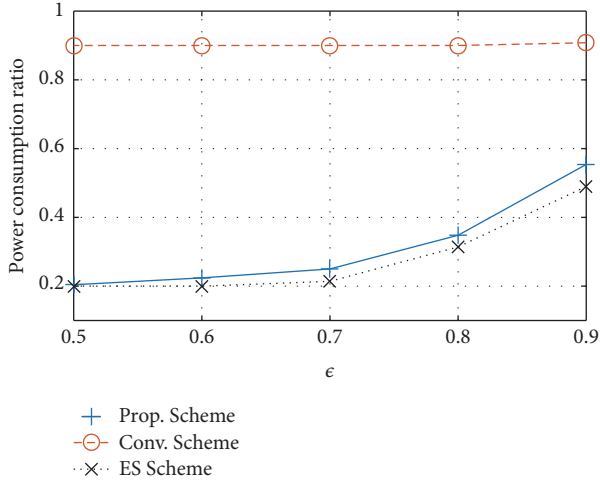
FIGURE 3: Power consumption ratio versus ϵ when $|\mathbb{F}| = 10$.

FIGURE 5: Power consumption ratio versus number of FBSs when $\epsilon = 0.7$.

FIGURE 4: Data rate versus number of FBSs when $\epsilon = 0.7$.
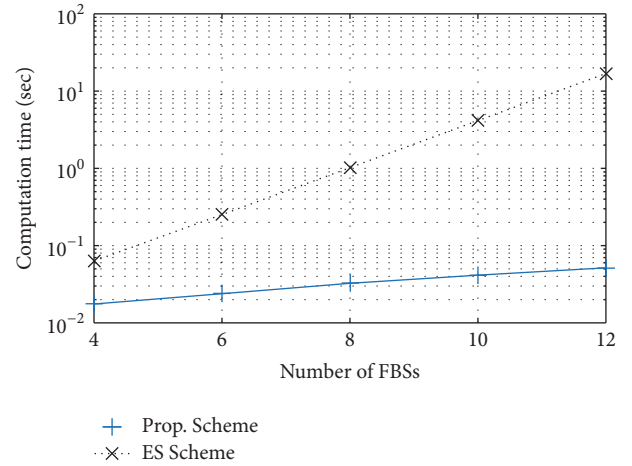
FIGURE 6: Computation time versus number of FBSs when $\epsilon = 0.7$.

is 10, where the power consumption ratio is the ratio of power consumption of the scheme and the power consumption without FBS shutdown, that is, $|\mathbb{F}_A|/|\mathbb{F}|$. For example, when the power consumption ratio is 0.5, 50% of energy can be saved. We can find that the data rate of all considered schemes is larger than R_{thr}; that is, constraints on data rate are satisfied in all cases. However, the data rate of Prop. Scheme and ES Scheme decreases as ϵ decreases while that of Conv. Scheme remains the same because the number of FBSs to be turned off is unchanged in Conv. Scheme while that of Prop. Scheme and ES Scheme decreases greatly, as can be verified from Figure 3. Specifically, in Prop. Scheme and ES Scheme, nearly 80% of FBSs can be turned off without violating the constraint on data rate such that energy consumption can be reduced significantly, which validates the benefit of our proposed scheme. Moreover, we can verify that our proposed scheme achieves near-optimal performance in terms of power consumption.

Figures 4 and 5 show the data rate and power consumption ratio by varying the number of FBSs when ϵ is 0.7. We can observe that the data rate increases as the number

of FBSs increases because received signal power of MSs is likely to increase. It is worth noting that the data rate of all considered schemes exceeds R_{thr}; that is, the constraint on data rate is satisfied even when $|\mathbb{F}|$ increases. Moreover, we can find that the power consumption ratio of our proposed scheme is much lower than that of Conv. Scheme, such that more energy can be saved in our proposed scheme compared to conventional FBS shutdown scheme. Particularly, we can observe that the gap between the power consumption ratio of our scheme and that of conventional scheme increases as $|\mathbb{F}|$ increases because the number of redundant FBSs is more when $|\mathbb{F}|$ is large; that is, the operation of more FBSs can be halted without hurting the data rate significantly. Therefore, we can conclude that our proposed scheme works better when the number of FBS is large. It should be noted that the power consumption ratio of Prop. Scheme and ES Scheme is almost the same; that is, the performance of our proposed scheme is nearly optimal, although Prop. Scheme is a heuristic scheme that works in distributed manner.

Finally, in Figure 6, the computation time of Prop. Scheme and ES Scheme by varying $|\mathbb{F}|$ is depicted. As can be

seen from the results, the computation time of our proposed scheme is much lower than that of ES Scheme because the number of computations required for ES Scheme increases exponentially with $|\mathbb{F}|$. Therefore, our proposed scheme is much better to be implemented in practice, especially in large CN where large number of FBSs is considered.

6. Conclusions

In this paper, the efficient way to halt the operation of redundant FBSs in cooperative transmission was exploited. Given that existing FBS shutdown schemes are inappropriate for FBS with cooperative transmission, we proposed new scheme by formulating the optimization problem. In order to solve the excessive computational complexity problem, a heuristic scheme based on TR was devised, which can operate in distributed manner. Through simulations based on realistic parameters, it was shown that our proposed scheme achieves near-optimal performance without excessive computations, such that the power consumption of FBSs can be greatly reduced without overly degrading the data rate of CN. The consideration of more generalized system model, in which FBSs transmit with different transmit power on multiple channels, might be an interesting extension of this work.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

References

[1] X. Fang, S. Misra, G. Xue, and D. Yang, "Smart grid—the new and improved power grid: a survey," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 4, pp. 944–980, 2012.

[2] K. Kusakana and H. J. Vermaak, "Hybrid renewable power systems for mobile telephony base stations in developing countries," *Renewable Energy*, vol. 51, pp. 419–425, 2013.

[3] S. Zhang, N. Zhang, S. Zhou, J. Gong, Z. Niu, and X. Shen, "Energy-aware traffic offloading for green heterogeneous networks," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1116–1129, 2016.

[4] A. Mukherjee, S. Bhattacherjee, S. Pal, and D. De, "Femtocell based green power consumption methods for mobile network," *Computer Networks*, vol. 57, no. 1, pp. 162–178, 2013.

[5] R. Mahapatra, Y. Nijsure, G. Kaddoum, N. Ul Hassan, and C. Yuen, "Energy efficiency tradeoff mechanism towards wireless green communication: a survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 686–705, 2016.

[6] F. K. Shaikh, S. Zeadally, and E. Exposito, "Enabling technologies for green internet of things," *IEEE Systems Journal*, 2015.

[7] E. Oh, B. Krishnamachari, X. Liu, and Z. Niu, "Toward dynamic energy-efficient operation of cellular network infrastructure," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 56–61, 2011.

[8] Y. S. Soh, T. Q. S. Quek, and M. Kountouris, "Dynamic sleep mode strategies in energy efficient cellular networks," in *Proceedings of the 2013 IEEE International Conference on Communications (ICC '13)*, pp. 3131–3136, Budapest, Hungary, June 2013.

[9] E. Yaacoub and A. Kadri, "Green operation of LTE-A femtocell networks benefiting from centralized control," in *Proceedings of the 11th International Wireless Communications and Mobile Computing Conference (IWCMC '15)*, pp. 408–412, Dubrovnik, Croatia, August 2015.

[10] H. Nabuuma, E. Alsusa, and W. Pramudito, "A load-aware base station switch-off technique for enhanced energy efficiency and relatively identical outage probability," in *Proceedings of the 81st IEEE Vehicular Technology Conference (VTC Spring '15)*, Glasgow, UK, May 2015.

[11] A. Bousia, A. Antonopoulos, L. Alonso, and C. Verikoukis, "'Green' distance-aware base station sleeping algorithm in LTE-Advanced," in *Proceedings of the IEEE International Conference on Communications (ICC '12)*, pp. 1347–1351, Ottawa, Canada, June 2012.

[12] W. Vereecken, M. Deruyck, D. Colle et al., "Evaluation of the potential for energy saving in macrocell and femtocell networks using a heuristic introducing sleep modes in base stations," *Eurasip Journal on Wireless Communications and Networking*, vol. 2012, article 170, 2012.

[13] G. Cili, H. Yanikomeroglu, and F. R. Yu, "Cell switch off technique combined with coordinated multi-point (CoMP) transmission for energy efficiency in beyond-LTE cellular networks," in *Proceedings of the 2012 IEEE International Conference on Communications (ICC '12)*, pp. 5931–5935, Ottawa, Canada, June 2012.

[14] A. Al-Dulaimi, A. Anpalagan, M. Bennis, and A. V. Vasilakos, "5G green communications: C-RAN provisioning of CoMP and femtocells for power management," in *Proceedings of the IEEE International Conference on Ubiquitous Wireless Broadband (ICUWB '15)*, pp. 1–5, Montreal, Canada, October 2015.

[15] W. Lee and D. Cho, "Simultaneous RTS and sequential CTS considering multiple cooperative relays," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 5, pp. 2369–2374, 2013.

[16] W. Lee and H. Lee, "Performance evaluation of coordinated multi-point transmission and reception in indoor mobile communication systems," *Journal of Information and Communication Convergence Engineering*, vol. 11, no. 3, pp. 167–172, 2013.

[17] Report ITU-R M.2135-1, Guidelines for evaluation of radio interface technologies for IMT-Advanced, December 2009.

Research Article

An Adaptive Information Quantity-Based Broadcast Protocol for Safety Services in VANET

Wenjie Wang, Tao Luo, and Ying Hu

Beijing Key Laboratory of Network System Architecture and Convergence, Beijing Laboratory of Advanced Information Networks, Beijing University of Posts and Telecommunications, Beijing 100876, China

Correspondence should be addressed to Wenjie Wang; isa_guet@163.com

Vehicle-to-vehicle communication plays a significantly important role in implementing safe and efficient road traffic. When disseminating safety messages in the network, the information quantity on safety packets changes over time and space. However, most of existing protocols view each packet the same to disseminate, preventing vehicles from collecting more recent and precise safety information. Hence, an information quantity-based broadcast protocol is proposed in this paper to ensure the efficiency of safety messages dissemination. In particular, we propose the concept of emergency-degree to evaluate packets' information quantity. Then we present EDCast, an emergency-degree-based broadcast protocol. EDCast differentiates each packet's priority for accessing the channel based on its emergency-degree so as to provide vehicles with more safety information timely and accurately. In addition, an adaptive scheme is presented to ensure fast dissemination of messages in different network condition. We compare the performance of EDCast with those of three other representative protocols in a typical highway scenario. Simulation results indicate that EDCast achieves higher broadcast efficiency and less redundancy with less delivery delay. What we found demonstrates that it is feasible and necessary for incorporating information quantity of messages in designing an efficient safety message broadcast protocol.

1. Introduction

The *Vehicular ad hoc Network* (VANET) is a specific application for wireless communication technology implementing in *Intelligent Transportation System* (ITS) [1] to improve the safety and efficiency of road traffic. In the United States, a 75 MHz bandwidth at the 5.9 GHz band [2] has been allocated to support the *Dedicated Short-Range Communication* (DSRC) [3] for supporting various applications in VANET. Moreover, the 802.11p standard defines specifications of the physical layer and the *Medium Access Control* (MAC) layer of *Vehicle*-to-*Vehicle* (V2V) or *Vehicle*-to-*Roadside* (V2R) communications. Lots of safety applications are available for satisfying the requirement of ITS, such as accident warning, collision detecting, and intelligent driving [4]. These safety services guide drivers to take some action, so as to improve the safety and quality of road transportation. Since the communication range of vehicles is limited between 300 m and 400 m, life-critical safety messages need to be broadcasted by multihop to cover the *Region of Interest* (RoI), for example, 1000 m.

Different from transmitting nonsafety messages in unicast or geomulticast, broadcasting safety messages requires stricter *Quality of Service* (QoS), for example, small latency and high reliability. Besides, because of lack of Request to *Send/Clear to Send* (RTS/CTS) scheme in the current scheme of IEEE802.11p, the failure of transmission often occurs because of weak links or frequent collisions with hidden terminals. What is more, there is no *acknowledgment* (ACK) scheme to detect the transmission failure, and no backoff mechanism is used to alleviate the newly coming collisions even in the sparse condition, which will significantly degrade the broadcast performance. Moreover, VANET is characterized by uneven vehicle distribution and highly dynamic topology. Therefore, how to broadcast the safety messages quickly and reliably is a challenging work, attracting increasing attention from industrial and academic communities.

To disseminate safety-related messages, rich broadcast protocols [4, 5] have been proposed, such as waiting-based

protocols [6–10], probability-based protocols [11, 12], and contention-based protocols [13, 14]. Their common ideas are farthest-first, allowing the farthest vehicles to rebroadcast packets first. Take waiting-based protocols as an instance, vehicles that locate farther away from themselves to sender are assigned higher priority to rebroadcast packets in terms of less waiting time, so as to achieve greater hop progress and less end-to-end delay.

However, most of existing broadcast protocols view all messages equally and overtake their different information quantity. For example, a newly produced emergency message contains more information quantity than an old message. However, they are assigned with the same size of *Contention Window* (CW) by most contention-based protocols. Another example is that waiting-based protocols give higher priority to the message from a farther location while vehicles actually prefer the message generated in a nearby station.

In this work, we define the concept of information quantity of messages in VANET and explore the benefits of integrating information quantity into the design of safety data broadcast protocol. Information quantity is an index of emergency level of messages, which indicates the safety status of road and traffic. The more emergent event a message represents, the more information quantity it contains. Since the messages with more information quantity are more useful to the vehicles driving towards where the messages take place, the higher priority should be assigned to them for channel access. Therefore, we propose the concept *emergency-degree (ED)* to vividly evaluate the information quantity of each message. Then an emergency-degree-based safety messages broadcast protocol, EDCast, is designed. Packets with higher ED values are assigned higher priorities in terms of a higher broadcast probability and a smaller size of CW. Besides, an adaptive scheme based on estimated vehicle density is adopted to optimize delay performance.

The main contributions of this paper are as follows:

(1) We propose the concept of *emergency-degree* (ED) to quantify the information quantity of different messages. ED incorporates three message attributes, which are temporal information quantity, spatial information quantity, and type information quantity. Spatial information quantity depends on cumulative number of *hops* (*hop* in italic means the number of hops in this paper) and the overall covering distances from generation to current time. To the best of our knowledge, this is the first attempt to mathematically evaluate the dynamic information quantity of messages in VANET.

(2) EDCast, an adaptive emergency-degree-based broadcast protocol, is proposed, which gives greater opportunities to more urgent messages to disseminate quickly. Compared to three other representative broadcast protocols in a typical highway scenario using NS2 simulator, EDCast performs better in terms of broadcast efficiency and delivery delay. It also reaches higher emergency packet delivery ratio while ensuring fast dissemination of emergency messages.

The rest of this paper is organized as follows. We review the related works in Section 2. We discuss our motivation and challenges in Section 3. We present metrics of information quantity in Section 4. We propose EDCast protocol in detail in Section 5 and discuss its performance in Section 6. Finally the conclusions are drawn in Section 7.

2. Related Works

In our previous work [15], we have summarized by simulation that receiver-based protocols outperform sender-based protocols due to high reliability, little overhead, and completely distributed manner. Mflood [16] is one of the most common receiver-based protocols. Although it provides fast and reliable dissemination, the broadcast storm [6] in dense network would significantly degrade broadcast performance. To mitigate broadcast storm, many valuable protocols have been proposed, including waiting-based, probability-based, and contention-based schemes.

The basic idea of waiting-based protocols is to differentiate candidates' waiting time, which is inversely proportional to the distance from receivers to last forwarder. Slotted-1 [7] assigned vehicles in farthest segments waiting less time before rebroadcast. Similarly, a binary-partition-assisted broadcast (BPAB) [8] was proposed to improve the latency performance using a binary-partition scheme. To obtain hop progress and reduce redundancy, eSBR [9] were present to help receivers make forwarding decisions according to the one-hop covering range. In addition, a two-phase broadcast scheme OppCast was proposed [17]. One phase satisfies fast dissemination while the other phase ensures reliability. Recently, ROFF [10] was proposed to mitigate collision by considering the short difference of waiting time of adjacent vehicles.

Probability-based schemes allow parts of receivers to forward packets to reduce redundancy. In slotted-p and weighted-p [7], vehicles that are farther away from the previous forwarder have higher probabilities to rebroadcast packets, so that the hop progress would be maximized and the end-to-end delay would be minimized. The forwarding probability of packets in [11] was given based on the number of duplicates received by vehicles, while in [12] the probability relies on the location and the density of network. In this way, the collision is reduced significantly.

Different from waiting-based schemes, contention-based schemes differentiate packets' priorities by adjusting the size of CW rather than using a timer. In [13], a priority setting scheme and a collision avoidance mechanism which is different from CSMA/CA were proposed. Nonzero sizes of IEEE 802.11 MAC backoff window were given to class-three service while zero window size was given to class-one emergency messages. In this way, emergency message gets the channel before beacon even though they are ready to broadcast at the same time because beacon has to wait for a DIFS before transmission. In [14], vehicles located in the farthest narrow segment were assigned smaller size of CW to rebroadcast packets first.
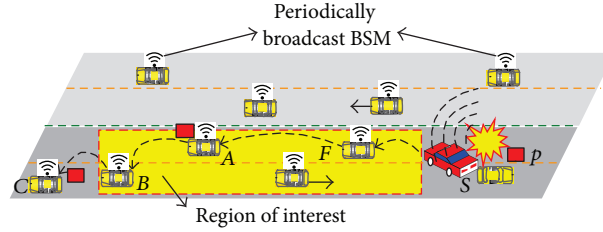
FIGURE 1: A typical highway scenario in VANET.

The common drawback of these schemes is that they view all packets the same and ignore the difference of the information quantity of messages, which is of great significance in providing vehicles with real-time and precise safety information. Some efforts [18, 19] have been present to differentiate packets' priority, such as EDCA scheme in 802.11p; however, they only consider only one aspect of information quantity (e.g., distance or type of messages). In addition, PVCast [20] integrates data preference into broadcast protocol. Although more data coverage and throughput can be achieved, lots of redundancy would be produced to satisfy the data preference of all vehicles in the network.

3. Safety Information Quantity of Message: Motivation and Challenges

In real vehicular networks, there are many kinds of safety messages to broadcast for satisfying different applications. Take a typical highway scenario as an example, as shown in Figure 1. There are two types of safety messages in the network. One is the *Basic Safety Message* (BSM) [21], which is broadcasted periodically to announce vehicle's status information (i.e., position, direction, velocity, neighbor number, etc.) to its neighbors for collision detecting, for instance. Another one is *emergency message* (EM), which is disseminated by multihop to warn the vehicles driving towards the given place where the emergency happened. We can see from the figure that an emergency message in terms of packet p is forwarded hop by hop until it covers the whole RoI. With the increase of time or as the extension of space in the form of cumulative *hops* and distances, the emergency level of p decreases. Thus its information quantity gets less and even turns to be zero when it reaches the end or out of RoI. A message with zero information quantity is useless for any vehicle. To this end, we can conclude that the information quantity of messages in vehicular network changes temporally and spatially. Besides, EM contains more information quantity than BSM in the same condition.

We then use two examples to show the necessity to consider the information quantity when designing a broadcast protocol. In Figure 2(a), two packets $p1$ and $p2$ are generated at the same place but at different time. Farthest-first protocols assign them the same priority to access the channel due to the same one-hop coverage. However, the newer packet $p2$ contains more recent condition of traffic, which is more useful to the coming vehicle F. Therefore, to obtain more

information quantity for vehicle F, vehicle B should have a higher priority to broadcast $p2$.

In Figure 2(b), two packets are generated at the same time but at different locations. Since $p1$ covers longer one-hop distance from current receiver to *last forwarder*, most existing protocols, for example, waiting-based protocols, give higher priority to vehicle A to broadcast $p1$ because A can disseminate a packet farther away from its source than B does. However, vehicle F prefers to receive the nearby packet $p2$ compared to $p1$ because $p2$ contains the status of a vehicle closer to F. The earlier F receives $p2$, the more information quantity F can get. To obtain more accurate information of road and traffic, therefore, vehicle B should have higher priority to broadcast $p2$.

Motivated by these observations, we think it is necessary to give much attention to study the difference of information quantity of messages in VANET and to incorporate it into the design of safety data broadcast protocol. We attempt to analyze the information quantity of messages in the following three aspects.

Temporal Information Quantity. The longer duration from generation to current time a message experiences, the less information quantity it contains. New message is more important than outdated message.

Spatial Information Quantity. The larger space in terms of cumulative *hop* and distances a message covers from generation location to current location, the less information quantity it contains. A nearby message with fewer *hops* is more urgent than a farther one with more *hops*.

Type Information Quantity. EMs are life-critical messages while BSMs are background beacons. Hence, EM contains more information quantity than BSM.

We face the following challenges when we integrate the information quantity into the design of safety data broadcast protocol.

Challenge 1. A comprehensive model is needed to quantify the information quantity of messages. The information quantity depends on several attributes of messages, such as temporal attribute, spatial attribute, and type attribute. All these attributes should be considered completely. Ignoring any one of them will lead to severe decrease of broadcast performance, such as high delivery delay [20].

(a) Temporal information quantity

(b) Spatial information quantity
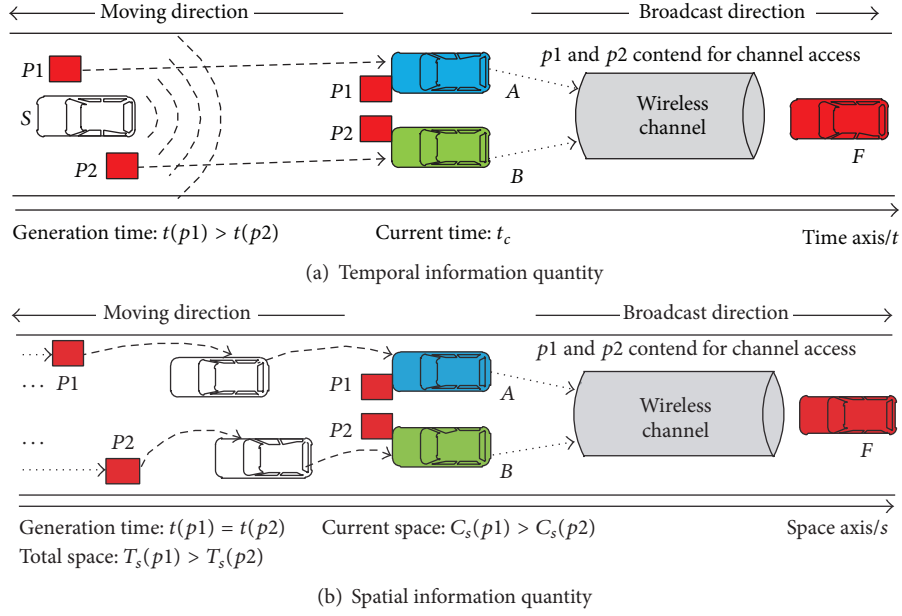
FIGURE 2: Two aspects of information quantity.

Challenge 2. Other than ensuring vehicles to receive more recent and more accurate emergent messages in time, the newly proposed protocol should address the common problems of broadcast protocols such as reliability, timeliness, redundancy, overhead, and scalability. Only in this way are the increasing kinds of safety applications likely to be supported better.

4. Metric of Information Quantity

In this section, we will model the information quantity to address Challenge 1 discussed in Section 3. Since the messages are disseminated in the unit of packet, the following discussion about emergency-degree modeling will be per-packet level.

Definition 1. Given a packet p arriving at vehicle v to be disseminated, its *emergency-degree* $ED_v(p)$ indicates the value of information quantity, which depends on temporal, spatial, and type information quantity.

We explore the modeling and computation of emergency-degree in the following. As is shown in Figure 1, given an emergency packet p, it is generated by vehicle S with the location (x_0, y_0) at time t_0. After broadcasting by vehicle S, packet p is received by vehicle F at current time t_c with current location (x_c, y_c), where the message is waiting to be handled based on a forwarding rule. At this moment, the cumulative *hop* the packet p experiences turns to be h_c from zero. So long as the packet is not out of RoI, it should be forwarded sequentially. Then each packet can be defined by an 8-tuple $\{t_0, t_c, x_0, y_0, x_c, y_c, h_c, \text{type}_p\}$. For example, once a vehicle receives a packet p, we can say that, during the time $\Delta t = t_c - t_0$, the type_p packet covers a distance

$\text{dis}_c = \sqrt{(x_c - x_0)^2 + (y_c - y_0)^2}$ after broadcasting h_c hop. In the procedure of multihop broadcast, the content of packets' 8-tuple changes, representing the dynamic of information quantity. For a given packet, its emergency-degree value is changing both temporally and spatially, by which the information quantity of packets is quantified. Therefore, the computation of emergency-degree should include all three classes of packets properties. Then the emergency-degree of packet p at location (x_c, y_c) and time instance t_c for vehicle v can be expressed as

$$ED_v(p) = T_v(p) \cdot S_v(p) \cdot TP_p. \tag{1}$$

In formula (1), $T_v(p)$ is the temporal emergency-degree function of duration that packet p experiences. $S_v(p)$ is the spatial emergency-degree function of distances and *hop* that packet p covers. And TP_p is the type emergency-degree function of the type that packet p belongs to. The product of function $T_v(p)$, $S_v(p)$, and TP_p integrates all three packet information quantities into the signal expression of packet emergency-degree. In the following we will define these functions, respectively.

4.1. Temporal Emergency-Degree Function. The newer message can provide vehicles with recent road and traffic condition; thus it should has a higher emergency-degree value. With the increase of duration a message experiences, its emergency-degree decreases. The descent speed is high at first and gets slow gradually. Moreover, no matter how long a message experiences, it is still useful theoretically for some applications such as density estimation of vehicles. Therefore, temporal emergency-degree value is always greater than zero. In addition, since the latency of multihop EM (on the level of millisecond [22, 23]) is much less than the update interval of

BSM (more than 100 ms), EM is always more urgent. Thus the temporal emergency-degree of EM is greater than that of BSM when they are created at the same time. From these observations, we summarize the properties of a temporal emergency-degree function which should include that

(1) $T_v(p) > 0$, $dT_v(p)/dt_c < 0$, $d^2T_v(p)/d^2t_c > 0$;

(2) $T_v(\text{EM})$ decreases slower than $T_v(\text{BSM})$.

Formula (2) is proposed to quantify the temporal emergency-degree of a packet:

$$T_v(p) = e^{-\omega_{tp}(t_c - t_0)}, \tag{2}$$

where ω_{tp} is the fading factor of temporal emergency-degree depending on the type of message. We define $\omega_b > \omega_e > 0$, which represents that the temporal emergency-degree of EM decreases slower than that of BSM. Formula (2) differentiates different types of packets generated at different time, for example, Figure 2(a); then the emergency packets newly generated will have a higher priority.

4.2. Spatial Emergency-Degree Function. Due to channel fading and collision, many emergency packets actually experience different *hop* when they reach adjacent vehicles for competing channel access, although uniform forwarding rule is followed. With the increase of *hops*, packets become useless for the potential receivers, because the intermediate forwarders gradually expose the packets' information every hop. As a result, packets with different *hop* do not stay the same level of emergency, although they are generated at the same time and are covering the same distances. Vehicles prefer to receive packets with fewer cumulative *hops*, because these packets are of more indeterminacy. That is why spatial information quantity depends on both the message's covering distances and cumulative *hop*.

With the extension of space, the emergency-degree of distance and *hop* decreases. On one hand, the distance emergency-degree decreases as the distance increases until the message reaches out of RoI, where it becomes zero. We summarize the properties of a distance emergency-degree function which should include that

(1) $\text{DIS}(p) = 0$, if vehicle v moves away from generation location of packet p;

(2) $\text{DIS}(p) = 0$, if $\text{dis}_c > d_{\text{RoI}}$, where d_{RoI} is the range of RoI;

(3) $\text{DIS}(p) \geq 0$, $d\text{DIS}(p)/d\text{dis}_c < 0$, if $\text{dis}_c \leq d_{\text{RoI}}$.

Formula (3) is proposed to quantify the distance emergency-degree of a packet:

$$\begin{aligned}\text{DIS}(p) &= \max\left(0, 1 - \frac{\text{dis}_c}{d_{\text{RoI}}}\right) \\ &= \max\left(0, 1 - \frac{\sqrt{(x_c - x_0)^2 + (y_c - y_0)^2}}{d_{\text{RoI}}}\right).\end{aligned} \tag{3}$$

On the other hand, the *hop* emergency-degree degrades rapidly as the cumulative *hop* rises. Although *hop* attribute plays a small part on spatial emergency-degree, it should not be ignored, especially in the bad communication condition such as channel fading, obstructions, interference, contention, and collision. For simplicity, we assume that cumulative *hop* and elapsed time characterize the same reduction trend, and the property of *hop* emergency-degree function is similar to that of temporal emergency-degree function. From these observations above, formula (4) is proposed to quantify the *hop* emergency-degree of a packet:

$$\text{HOP}(p) = e^{-w_e h_c}, \tag{4}$$

where w_e is the fading factor of EM.

Since *hop* attribute is mainly affected by communication environment, while distance attribute mainly depends on forwarding rule, no necessary relation exits between them, although they seem to follow a similar varying pattern. From these summaries above, a weight formula is proposed to evaluate a packet's spatial emergency-degree approximately:

$$S_v(p) = A \cdot \text{DIS}(p) + B \cdot \text{HOP}(p), \tag{5}$$

where A and B ($A \geq 0$, $B \geq 0$, and $A + B = 1$) are weight factors of distance emergency-degree and *hop* emergency-degree, respectively. The values of A and B depend on the importance of distance and hop attributes on the spatial emergency-degree in typical scenario of VANETs.

We can see from formulas (3)–(5) that a message that covers longer range and experiences more *hops* becomes less urgent, containing less spatial information quantity. However, the farthest-first schemes assign higher priority to packets that only cover longer distance of current hop, overtaking the whole covered distances from generation to current location, for example, Figure 2(b). On the contrary, we assign lower priority to those packets that undergo broader space. Hence, vehicles will collect more precise safety information.

4.3. Type Information Quantity Function. As we know in EDCA, safety services have higher priority to access the channel in terms of smaller size of CW than nonsafety services. However, the type of safety messages should be further subdivided when considering their information quantity. That is because a number of safety messages [13, 24] would be broadcasted in VANET, acting different levels of emergency to implement all kinds of applications. Apart from affecting the reduction speed of temporal emergency-degree in (2), the type attribute also affects the initial value of packet newly generated. Thus we use different weight to quantify different types of messages:

$$\text{TP}_p = W_{\text{TP}}, \quad \text{TP} \in \{\text{EM}, \text{BSM}\}, \tag{6}$$

in which we define $W_{\text{EM}} > W_{\text{BSM}} > 0$ to indicate that the information quantity of EM is larger than BSM.

To the end, the emergency-degree-based information quantity can be rewritten in

$$
\begin{aligned}
&\mathrm{ED}_v\left(p\right) \\
&= \begin{cases} \left[A \cdot \max\left(0, 1 - \dfrac{\sqrt{(x_c - x_0)^2 + (y_c - y_0)^2}}{d_{\mathrm{RoI}}} \right) + B \cdot e^{-w_e h_c} \right] \cdot e^{-\omega_{tp}(t_c - t_0)} \cdot W_{\mathrm{TP}}, & \vec{v} \in \{\mathrm{RoI}\}, \ \mathrm{dis}_c\left(p\right) < d_{\mathrm{RoI}} \\ 0, & \text{otherwise,} \end{cases}
\end{aligned} \tag{7}
$$

where $\vec{v} \in \{\mathrm{RoI}\}$ means vehicle v moves towards the packet generation location with the opposite direction of broadcast. Other vehicles such as vehicles moving away from the packet generation location or vehicles on the other moving direction will be valued as zero by this model.

Emergency-Degree-Based Information Quantity. Given any two packets at vehicle v, $p1$ will have larger information quantity than $p2$ if $\mathrm{ED}_v(p1) > \mathrm{ED}_v(p2)$. In another way, given two packets at vehicle A and vehicle B, respectively, vehicle A with $p1$ will have higher information quantity than vehicle B with $p2$ if $\mathrm{ED}_A(p1) > \mathrm{ED}_B(p2)$.

With these equations, we have modeled the metric of information quantity with emergency-degree, considering all three attributes. Thus, Challenge 1 discussed in Section 3 can be resolved.

5. EDCast: An Adaptive Emergency-Degree-Based Broadcast Protocol

Having quantified the information quantity of message with emergency-degree metric, we wander the benefits of incorporating it into the design of broadcast protocol. EDCast, an adaptive emergency-degree-based broadcast protocol, is proposed in this section.

5.1. Overview of EDCast. Farthest-first scheme is the most common protocol in VANET, in which farther vehicles are assigned higher priority in terms of less waiting time to access the channel. The waiting time is inversely proportional to the distance from receiver to the previous forwarder, which is shown as follows:

$$
\mathrm{WT} = \mathrm{MaxWT} \cdot \left(1 - \frac{d}{R} \right), \tag{8}
$$

where d is the one-hop distance from itself to the last forwarder, R is the communication range, and MaxWT is the maximum waiting time.

To address Challenge 2 discussed in Section 3, we introduce the farthest-first scheme as the basic principle of EDCast for relay selection and redundancy suppression. Then we integrate information quantity into it to differentiate messages' priority. Figure 3 shows the flow chart of EDCast.

We can see from Figure 3 that once upon a new packet p reaches vehicle A, EDCast will decide whether or not to broadcast the packet and how to disseminate it by calculating the emergency-degree $\mathrm{ED}_v(p)$. After updating network parameters based on the estimated number of vehicles, EDCast assigns packet p appropriate waiting time WT, the size of the minimal CW, and broadcast probability P_r according to its type. The self-adaptation scheme of network status will be described in Section 5.3.

Based on the farthest-first protocols, EDCast is lightweight without increasing any control overhead and fully distributed without the help of RSU. Thus it is feasible and convenient for EDCast to be embedded into the current structure of 802.11p.

5.2. Probability and CW Size Assignment Module. To provide vehicles with the recent and precise information, EDCast configures higher priority to the messages with higher ED values in terms of smaller size of CW. Furthermore, in order to reduce the collision among packets during transmission, EDCast randomly drops the packets with small ED value in certain probability. The probability scheme is also used to alleviate the broadcast storm especially in the dense area. Particularly, if the vehicle is the source of packet, the packet will be disseminated immediately with the initialized network parameters.

We use the following piecewise functions to evaluate the probability and CW size of a packet based on its ED value, which are shown in (9) and (10), respectively.

$$
P_e\left(p\right) = \begin{cases} P_1, & \mathrm{ED}\left(p\right) \geq \overline{\mathrm{ED}_1} \\ P_2, & \overline{\mathrm{ED}_2} \leq \mathrm{ED}\left(p\right) < \overline{\mathrm{ED}_1} \\ P_3, & \text{otherwise,} \end{cases} \tag{9}
$$

where P_i, $i = 1, 2, 3$, is the threshold of forwarding probability, and we predefine $1 > P_1 > P_2 > P_3 > 0$. What is more, we also predefine the threshold of ED value $\overline{\mathrm{ED}_1} > \overline{\mathrm{ED}_2} > 0$, based on which we differentiate the level a packet belongs to.

$$
\mathrm{CW}_{\min}\left(p\right) = \begin{cases} \mathrm{CW}_1, & \mathrm{ED}\left(p\right) \geq \overline{\mathrm{ED}_1} \\ \mathrm{CW}_2, & \overline{\mathrm{ED}_2} \leq \mathrm{ED}\left(p\right) < \overline{\mathrm{ED}_1} \\ \mathrm{CW}_3, & \text{otherwise,} \end{cases} \tag{10}
$$

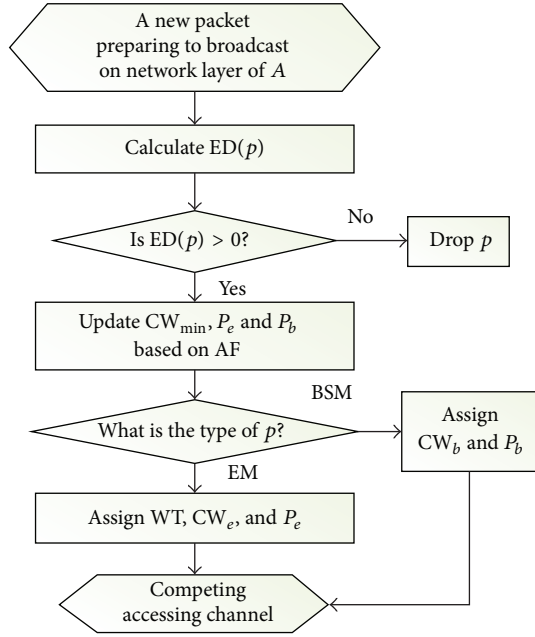where CW_i, $i = 1, 2, 3$, is the threshold of size of CW.

FIGURE 3: The flow chart of EDCast.

The piecewise functions such as (9) and (10) are necessary, for they provide protocol designers with flexibility to adjust the threshold parameters dynamically according to the network condition. In this way, an adaptation scheme could be used to optimize the broadcast performance of EDCast.

5.3. Network Adaptation Module.
As known, VANET characterizes highly dynamic topology. Neither sparse network nor dense network will decrease broadcast performances significantly if the networks parameters remain constant. An adaptation scheme, just as a simple example, is proposed to adjust these parameters roughly. In EDCast, the exponentially weighted moving average (EWMA) [20, 25] traffic density estimator is used to estimate $\widehat{N}_A$, the real-time vehicle number within the one-hop coverage of vehicle A. During the period T (e.g., 5 s), EDCast counts the number of vehicles A can hear. Then $\widehat{N}_A$ can be estimated as the following equation:

$$\widehat{N}_A[T] = \alpha \cdot \widehat{N}_A[T-1] + (1-\alpha) \cdot N_A[T], \qquad (11)$$

where α is the weight factor and $N_A[T]$ is the average number of neighbors A is able to hear during T.

After getting the estimated number of vehicles, we define the *adaptation factor (*AF*)* for each vehicle to adjust broadcasting parameters dynamically when they have packets to disseminate.

Definition 2. Given a vehicle A, the adaptation factor (AF) is the estimated number of vehicles A divided by the maximum number A can hear in theoretical.

The AF value can roughly reflect the changing density around each vehicle, which is also per-packet level:

$$\mathrm{AF}_v = w_{\mathrm{sc}} \cdot \frac{\widehat{N}_v}{\widehat{N}_{\mathrm{MAX}}}, \qquad (12)$$

where $\widehat{N}_{\mathrm{MAX}}$ is the maximum number of neighbors within the communication range of vehicle v, which is calculated by $\widehat{N}_{\mathrm{MAX}} = (2 * R/(L + l_{\mathrm{safe}})) * N_{\mathrm{lane}}$, where R is intended communication range; L is the average length of vehicle; and l_{safe} is the safety length, the average distance between vehicles when they are locating one by one safely in the most heavy traffic. And N_{lane} is the number of lanes. $\widehat{N}_v$ is the estimated number of neighbors of vehicle v. And W_{sc} is the weight factor depending on the chosen simulation scenario, which is used to balance the value of $\mathrm{AF} \in (0, 1)$.

For simplicity, a piecewise function is introduced to adjust the network parameters reasonably according to the estimated adaptation factor of each vehicle. The range of AF is divided into three parts, each of which represents the average level of traffic condition in certain scenario. In our simulation scenario, $\mathrm{CW}_{\mathrm{min}}$ and MaxWT are directly proportional to adaption factor AF, while P_i is inversely proportional to AF. Based on the estimated value of AF, three groups of parameters among parameters pool of CW_{Gi} and P_{Gi} are configured to $\mathrm{CW}_{\mathrm{min}}$ and P_i, respectively, while three single levels of parameters are assigned to MaxWT as well, which are shown in

$$\left(\mathrm{CW}_{\mathrm{min}}^G, P_i^G\right)$$

$$= \begin{cases} (\{\mathrm{CW}_{G1}\}, \{P_{G1}\}), & 0 \leq \mathrm{AF} \leq \mathrm{AF}_1 \\ (\{\mathrm{CW}_{G2}\}, \{P_{G2}\}), & \mathrm{AF}_1 \leq \mathrm{AF} \leq \mathrm{AF}_2 \\ (\{\mathrm{CW}_{G3}\}, \{P_{G3}\}), & \mathrm{AF}_2 \leq \mathrm{AF} \leq 1, \end{cases} \qquad (13)$$

$$\mathrm{MaxWT} = \begin{cases} \mathrm{WT}_1, & 0 \leq \mathrm{AF} \leq \mathrm{AF}_1 \\ \mathrm{WT}_2, & \mathrm{AF}_1 \leq \mathrm{AF} \leq \mathrm{AF}_2 \\ \mathrm{WT}_3, & \mathrm{AF}_2 \leq \mathrm{AF} \leq 1. \end{cases}$$

The detailed configure will be given in the section of simulation. In this way, the greater section AF belongs to, the larger CW_i and MaxWT are and the smaller P_i is.

As is known, the background safety message BSM plays a vital part in providing vehicles' status in realistic VANET, on which many safety applications rely, such as crash detection and density estimation. The more frequent the BSM sends, the more recent and precise status information of vehicles it presents. However, too frequent BSMs will cause server contention for channel access, leading to unreliable transmission. Therefore, there is a trade-off between broadcast performance and the accuracy of vehicle number estimation, which deserves much attention from protocol designers. As a matter of fact, with the increase of vehicle density, vehicles move more slowly, and the relative velocity between vehicles gets smaller, even approaching zero. The status of vehicles in these conditions remains the same for a long duration; thus
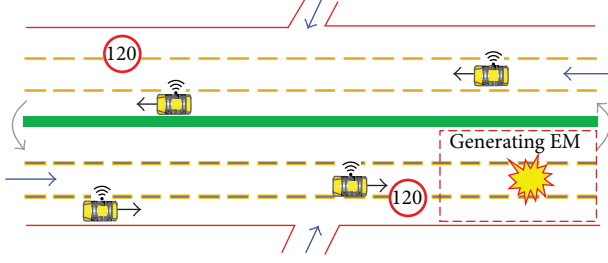
FIGURE 4: A typical highway scenario for simulation.

TABLE 1: Simulation parameters of scenario.

Parameter	Value
Topology scenario	2000 m × 40 m
Lanes, bidirectional	6
Number of vehicles	20, 40, 60, 80, 100
Speed range	30~120 km/h
Communication range	300 m
d_{RoI}	1000 m
EM size	300 bytes
PHY model	TwoRayGround
MAC model	802.11p
Simulation time	200 seconds

it is necessary to adjust the frequency of BSMs dynamically. In addition, the proposed adaption scheme in this paper is only a simple example to inspire better ideas for improving EDCast in the future.

6. Performance Evaluation

The network simulator NS2.35 is used to evaluate the performance of EDCast. EDCast is embedded as a middle-layer protocol between IEEE-802.11p MAC protocol and upper application protocol. When deciding to disseminate a packet, the configured minimum size of CW of this packet is passed from EDCast to MAC layer in the packet header. In this way, we do not need any intrusive change to the current IEEE-802.11p protocol stack. Besides, it is portable for EDCast to be installed into different protocol stacks.

6.1. Simulation Environment

Simulation Scenario. A traffic simulator, Simulation of Urban Mobility (SUMO) [26], is used to generate the trace of vehicles for a typical highway scenario shown in Figure 4. The topology consists of a bidirectional road of 2000 m long and 40 m wide with 6 lanes. There are two entrances in each moving direction: one is at the start and the other one is at the middle. 20–100 vehicles, including four types with different acceleration, deceleration, length, and limited speed, enter the topology successively at a period of 1 second from the four entrances, respectively. All the vehicles move following the modified Krauss Model. A connection file is configured to guide the vehicles driving encircling the bidirectional road, entering the reverse road when exiting from one road. Besides, 5% of the vehicles locate on the road uniformly and move slowly for composing a connected topology.

Communication Configuration. The communication range of each vehicle is set to be 300 m, and the range of RoI of emergency messages is set to be 1000 m. Multihop broadcast is necessary for EMs to cover the whole RoI. Vehicles locating at certain range (e.g., range between 1500 and 2000 m) generate EMs every 1 second with a certain probability (e.g., 50%). Vehicles in the network are both message generators and message forwarders. 10 BSMs are initialized to be generated every 1 second by each vehicle to announce its status. The main parameters for simulation are shown in Table 1.

To illustrate the necessity of considering the information quantity of messages to improve broadcast efficiency, we comparatively study the following protocols in the same scenario.

Mflood. The most original receiver-based protocol is implemented into VANET. Once they receive a packet, vehicles in mflood forward it immediately if the packet is new. Broadcast storm in the scenario of dense network needs to be optimized.

Farthest. The farthest-first protocol is first proposed in [6] for VANET. The vehicles that are farther to sender are assigned higher priority to access the channel in terms of less waiting time. In this way, during the procedure of multihop broadcast, one-hop forwarding can achieve higher geographical progress and less latency. That is why farthest is suggested as the basic idea of many protocols.

Slotted-p. A typical probability-based protocol. Similar to farthest, slotted-p designs a waiting timer to schedule packets based on the distance between the sender and receiver. When the waiting timer expires, the scheduler disseminates the packet with a probability (e.g., 50%) if no duplicates are received. Redundancy is reduced significantly.

EDCast. An adaptive emergency-degree-based protocol is presented in Section 5, which assigns more urgent messages higher priority by configuring higher probability and smaller size of CW, for the purpose of providing vehicles with the recent and precise safety information. The main communication parameters are listed in Table 2.

The following metrics are evaluated for comprehensively understanding the benefits of EDCast.

Delivery Delay. It is the average duration per hop per packet from generation to reception.

Redundancy. It is the total number of packets which are forwarded per vehicle divided by the total number of packets received.

Broadcast Efficiency. It is the total number of packets received by nodes divided by the total number of packets which are generated by sources during the simulation.

TABLE 2: Communication parameters of vehicles.

Parameter	Value
Average length of vehicle	4 m
Safety length between vehicles	2 m
Temporal fading factors w_b, w_e	0.5, 0.8
Spatial weight factors A, B	0.4, 0.6
Type weight factors W_{BSM}, W_{EM}	2, 10
MaxWT$_i$	{5, 15, 25} ms
Emergency-degree threshold ED$_i$	{4, 8}
Broadcast probability P_{Gi}	P_{BSM} = {0.02, 0.05, 0.1, 0.2, 1} P_{EM} = {0.1, 0.3, 0.5, 0.7, 1}
Minimum size of CW$_{Gi}$	{31, 23, 15, 7, 3}
Adaptation factor threshold AF$_i$	{0.3, 0.7}
Weight factor w_{sc}	6

Packet Delivery Ratio of Emergency (e-PDR). It is the percentage of packets covering the whole RoI successfully among the total packets generated by sources.

Broadcast Efficiency of Emergency. It is the total number of emergency packets received by nodes divided by the total number of packets which are generated by sources during the simulation.

Delivery Delay of Emergency. It is the average delivery delay of the first three hops of emergency message.

The first three metrics are regularly known in other literatures while the other three metrics are special for emergency messages in this paper.

6.2. Simulation Results. We simulate every scenario 10 times with different initialized values and get the average values of each metric during the 200-second simulation time. In the following, we present the simulation results of all the six metrics.

Figure 5 shows the average one-hop delivery delay of all packets. We can see that it increases with the density of vehicles due to more intense contention among vehicles to access the shared channel. But the increasing speed of EDCast is smaller than that of three other protocols. When there are 20–40 vehicles in the network, the difference of delivery delay of all protocols is small, while in the scenario of 60–100 vehicles existing, the delivery delay of EDCast is 3 times less than the second best one. The improvement of latency performance of EDCast mainly comes from two aspects: (1) the size of CW of packets is differentiated according to their emergency-degree value other than remaining the same; therefore, the contention and collision among packets are reduced significantly. (2) With the help of adaptation scheme, the maximal waiting time is changing along with the density of vehicles, which reduces the chance of unnecessary waiting. EDCast enables fast dissemination, providing vehicles with more recent safety information.

We then study the average redundancy and broadcast efficiency of all packets, as shown in Figures 6 and 7, respectively.
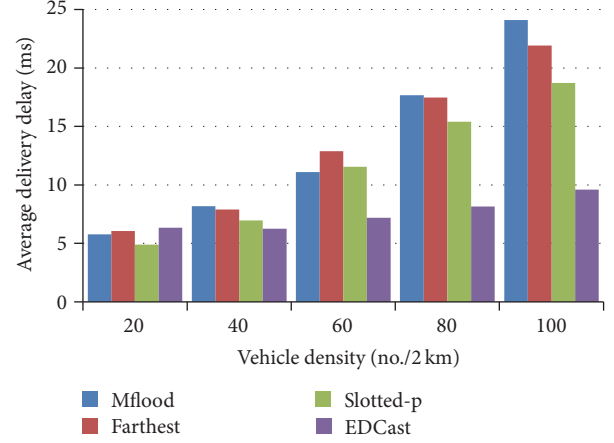
FIGURE 5: Delivery delay of all packets.
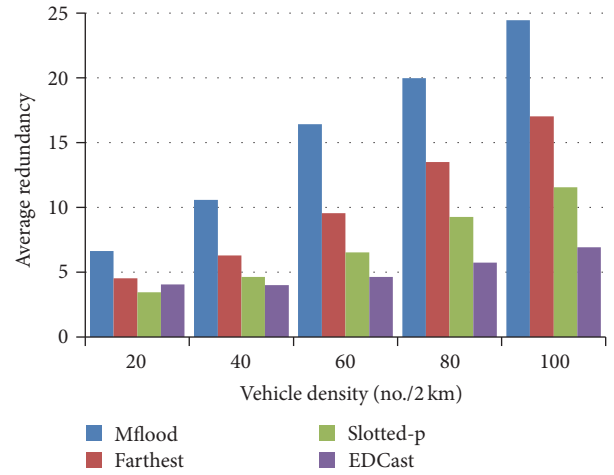
FIGURE 6: Redundancy.

Figure 6 shows that redundancy of all protocols increases with the number of vehicles due to more and more duplicates taking part in forwarding. In the sparse network, for example, 20 vehicles in the highway, EDCast ensures the transmission reliability by increasing dissemination probability, yielding a little more redundancy than the best one, slotted-p. But the redundancy of EDCast is still less than that of farthest and mflood. Moreover, when more than 20 vehicles exist in the highway, the redundancy of EDCast is $1x$-$2x$ less than that of three other protocols. We then plot the average broadcast efficiency of all packets in Figure 7. We can see that with the increase of vehicles the broadcast efficiency of protocols such as mflood, farthest, and slotted-p increases at first, reaches its peak when there are 40–60 vehicles, and then decreases gradually due to significant contention and collision, leading to frequent packet loss. On the contrary, by adjusting the broadcast probability and the size of the minimal CW adaptively based on the emergency-degree, EDCast disseminates packets orderly and quickly. Hence, broadcast efficiency of EDCast is much higher than that of other protocols especially in the 80–100 vehicles cases. What Figures 5–7 show illustrates that EDCast satisfies fast
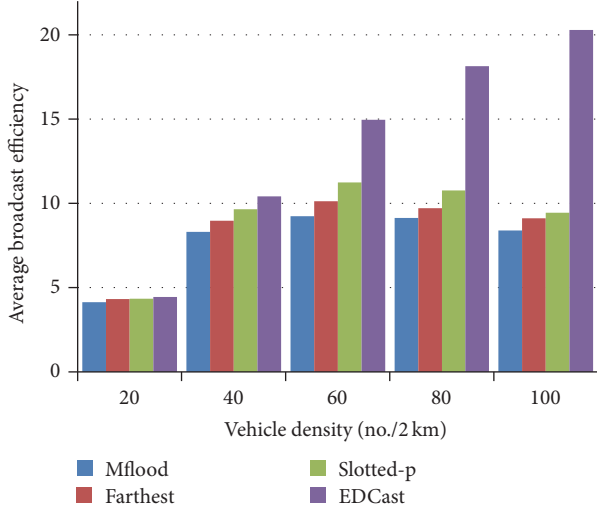
FIGURE 7: Broadcast efficiency.
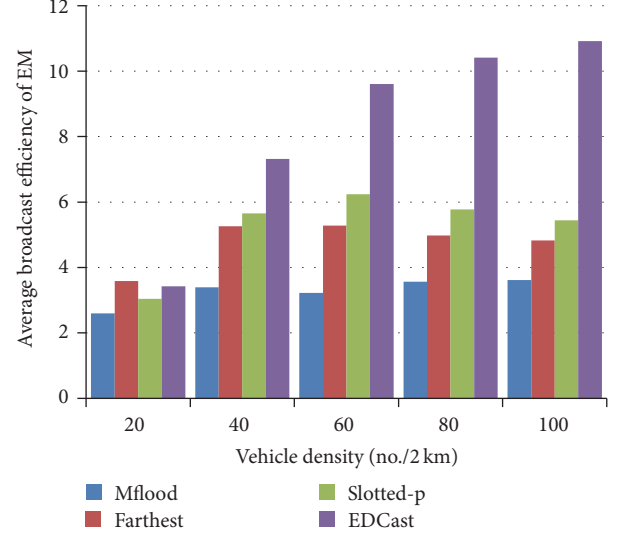
FIGURE 9: Broadcast efficiency of emergency packets.
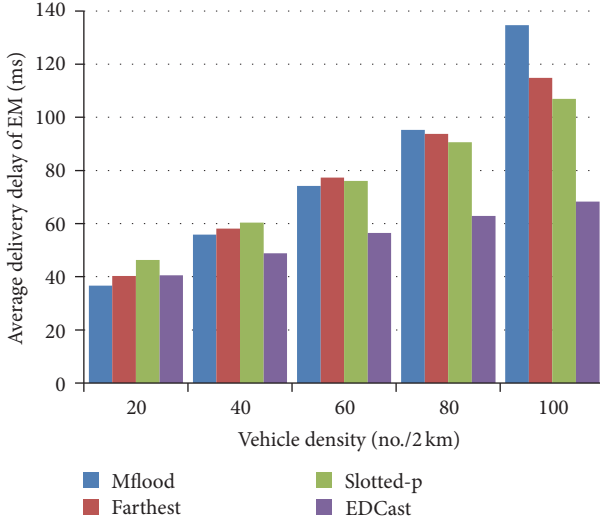
FIGURE 8: Delivery delay of the first three hops of emergency packets.
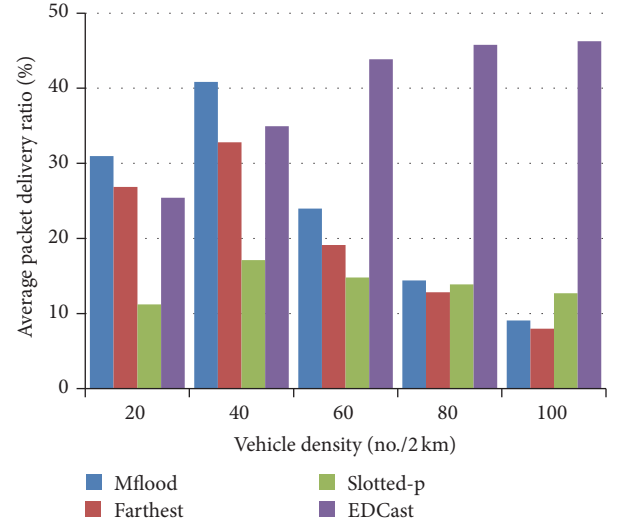
FIGURE 10: Packet delivery ratio of emergency packets.

dissemination in various traffic conditions and alleviates broadcast storm as well.

Furthermore, we study the broadcast performance of emergency messages separately. Figures 8–10 show the average delivery delay of the first three hops, the average emergency packet delivery ratio, and the average broadcast efficiency of emergency packet, respectively. Similar to Figure 5, we can see from Figure 8 that the average delivery delay of the first three hops of EDCast is much less than that of other three protocols. We also can see from Figure 8 that the broadcast efficiency of emergency packets of EDCast is much higher than other protocols. What is more, both the improvement of the first three-hop delivery delay and emergency broadcast efficiency are larger than that of one-hop delivery delay in Figure 5 and broadcast efficiency for all packets in Figure 7.

We plot the e-PDR of all protocols in Figure 10, which represents the reliability of broadcast protocols. Unlike other

protocols reaching the maximal e-PDR in the case of 40 vehicles and decreasing with the increase of vehicles, EDCast reaches its peak when there are 60 vehicles in the network and keeps steady afterwards. Besides, in the sparse network, for example, 20–40 vehicles, e-PDR of mflood and farthest are litter higher than EDCast, while in the scenario of dense network, for example, cases with more than 60 vehicles, EDCast performs better ($3x$-$4x$ times) than other protocols. That is because EDCast allows only certain vehicles in the farthest segment to forward packets to suppress redundancy. So the probability for packets to cover all the vehicles in RoI is less especially in the sparse region, leading to lower e-PDR of EDCast. In the crowed traffic condition of VANET, there are so many vehicles to forward packets that collision occurs frequently, leading to significant degradation of e-PDR in mflood and farthest. On the contrary, by adaptively adjusting forwarding probability and size of CW, EDCast

gets increasing e-PDR. It is demonstrated that it is reliable and efficient for EDCast to disseminate packets based on the metric of emergency-degree.

It is worth mentioning that though slotted-p produces smaller redundancy and lower deliver delay than farthest by forwarding packets in certain probability, it performs much lower e-PDR. On the contrary, EDCast reaches high broadcast efficiency and less delivery delay while introducing small redundancy at the same time. This phenomenon shows that using fixed probability and the same size of CW for broadcasting all packets is insufficient. Therefore, it is necessary to differentiate packets' probability and the minimal CW adaptively as EDCast does.

In summary, we illustrate the feasibility and effectiveness of broadcasting packets based on the information quantity for safety applications through simulation under a typical highway scenario. Compared to three other representative broadcast schemes, EDCast yields a higher broadcast efficiency and a less end-to-end delay with a less redundancy. Besides, it reaches a higher packet delivery ratio of emergency messages while consuming a lower delivery delay of the first three hops. In conclusion, EDCast can provide vehicles recent and precise safety information on road and traffic reliably.

7. Conclusions

In this paper, we study the influence of information quantity on broadcast performance. We propose the concept of emergency-degree to evaluate packets' emergency level on temporal, spatial, and type information quantity. EDCast, an adaptive emergency-degree-based broadcast protocol, is presented to disseminate safety messages, providing vehicles with recent and precise safety information about road and traffic. We compare the performance of EDCast with those of three other representative broadcast strategies by simulation under a typical highway scenario. The results show that EDCast obtains a substantial improvement on latency, broadcast efficiency, redundancy, and emergency packet delivery ratio. Our work proves that it is effective and feasible to design broadcast protocols based on safety information quantity of messages. In the future, we will consider the network load in the design of adaption scheme based on EDCast for improving broadcast performance.

Competing Interests

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

Acknowledgments

References

[1] W. Wang, S. S. Liao, X. Li, and J. S. Ren, "The process of information propagation along a traffic stream through intervehicle communication," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 1, pp. 345–354, 2014.

[2] IEEE 802.11 Working Group, "Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications," ANSI/IEEE Std 802.11, 1999.

[3] "Standard Specification for Telecommunications and Information Exchange Roadside and Vehicle Systems-5 GHz Band Dedicated Short Range Communications (DSRC) Medium Access Control (MAC) and Physical Layer (PHY) Specifications," April 2009.

[4] J. A. Sanguesa, M. Fogue, P. Garrido, F. J. Martinez, J.-C. Cano, and C. T. Calafate, "A survey and comparative study of broadcast warning message dissemination schemes for VANETs," *Mobile Information Systems*, vol. 2016, Article ID 8714142, 18 pages, 2016.

[5] S. Panichpapiboon and W. Pattara-Atikom, "A review of information dissemination protocols for vehicular ad hoc networks," *IEEE Communications Surveys and Tutorials*, vol. 14, no. 3, pp. 784–798, 2012.

[6] L. Briesemeister and G. Hommel, "Role-based multicast in highly mobile but sparsely connected ad hoc networks," in *Proceedings of the 1st Annual Workshop on Mobile and Ad Hoc Networking and Computing (MobiHOC '00)*, pp. 45–50, IEEE, Boston, Mass, USA, 2000.

[7] N. Wisitpongphan, O. K. Tonguz, J. S. Parikh, P. Mudalige, F. Bai, and V. Sadekar, "Broadcast storm mitigation techniques in vehicular ad hoc networks," *IEEE Wireless Communications*, vol. 14, no. 6, pp. 84–94, 2007.

[8] J. Sahoo, E. H.-K. Wu, P. K. Sahu, and M. Gerla, "Binary-partition-assisted MAC-layer broadcast for emergency message dissemination in VANETs," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 3, pp. 757–770, 2011.

[9] F. J. Martinez, M. Fogue, M. Coll, J.-C. Cano, C. Calafate, and P. Manzoni, "Evaluating the impact of a novel warning message dissemination scheme for VANETs using real city maps," in *NETWORKING 2010: 9th International IFIP TC 6 Networking Conference, Chennai, India, May 11–15, 2010. Proceedings*, vol. 6091 of *Lecture Notes in Computer Science*, pp. 265–276, Springer, Berlin, Germany, 2010.

[10] H. Yoo and D. Kim, "ROFF: RObust and fast forwarding in vehicular ad-hoc networks," *IEEE Transactions on Mobile Computing*, vol. 14, no. 7, pp. 1490–1502, 2015.

[11] A. Mohammed, M. Ould-Khaoua, L. M. Mackenzie, and J.-D. Abdulai, "Dynamic probabilistic counter-based broadcasting in mobile Ad Hoc networks," in *Proceedings of the 2nd International Conference on Adaptive Science and Technology (ICAST '09)*, pp. 120–127, Accra, Ghana, December 2009.

[12] A. M. Vegni, A. Mostafa, and D. P. Agrawal, "CAREFOR: collision-aware reliable forwarding technique for vehicular ad hoc networks," in *Proceedings of the International Conference on Computing, Networking and Communications (ICNC '13)*, pp. 773–777, IEEE, San Diego, Calif, USA, January 2013.

[13] X. Ma, J. Zhang, X. Yin, and K. S. Trivedi, "Design and analysis of a robust broadcast scheme for VANET safety-related services," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 1, pp. 46–61, 2012.

[14] K. A. Hafeez, L. Zhao, Z. Liao, and B. N.-W. Ma, "A new broadcast protocol for vehicular ad hoc networks safety applications,"

in *Proceedings of the 53rd IEEE Global Communications Conference (GLOBECOM '10)*, pp. 1–5, Miami, Fla, USA, December 2010.

[15] W. Wang and T. Luo, "The minimum delay relay optimization based on nakagami distribution for safety message broadcasting in Urban VANET," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '16)*, IEEE Press, Duha, Qatar, 2016.

[16] M. Barradi, A. S. Hafid, and S. Aljahdali, "Highway multihop broadcast protocols for vehicular networks," in *Proceedings of the IEEE International Conference on Communications (ICC '12)*, pp. 5296–5300, IEEE Press, Ottawa, Canada, June 2012.

[17] M. Li, K. Zeng, and W. Lou, "Opportunistic broadcast of event-driven warning messages in Vehicular Ad Hoc Networks with lossy links," *Computer Networks*, vol. 55, no. 10, pp. 2443–2464, 2011.

[18] R. S. Schwartz, A. E. Ohazulike, C. Sommer, H. Scholten, F. Dressler, and P. Havinga, "Fair and adaptive data dissemination for traffic information systems," in *Proceedings of the IEEE Vehicular Networking Conference (VNC '12)*, pp. 1–8, IEEE, Seoul, South Korea, November 2012.

[19] B. Ducourthial, Y. Khaled, and M. Shawky, "Conditional transmissions: performance study of a new communication strategy in VANET," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 6, pp. 3348–3357, 2007.

[20] Q. Xiang, X. Chen, L. Kong, L. Rao, and X. Liu, "Data preference matters: a new perspective of safety data dissemination in vehicular ad hoc networks," in *Proceedings of the 34th IEEE Annual Conference on Computer Communications and Networks (IEEE INFOCOM '15)*, pp. 1149–1157, Kowloon, Hang Kong, May 2015.

[21] J. B. Kenney, "Dedicated short-range communications (DSRC) standards in the United States," *Proceedings of the IEEE*, vol. 99, no. 7, pp. 1162–1182, 2011.

[22] M. Torrent-Moreno, S. Corroy, F. Schmidt-Eisenlohr, and H. Hartenstein, "IEEE 802.11-based one-hop broadcast communications: understanding transmission success and failure under different radio propagation environments," in *Proceedings of the 9th ACM Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems (ACM MSWiM '06)*, pp. 68–77, Malaga, Spain, October 2006.

[23] X. Ma and X. Chen, "Delay and broadcast reception rates of highway safety applications in vehicular Ad Hoc networks," in *Proceedings of the IEEE Workshop on Mobile Networks for Vehicular Environments (INFOCOM '07)*, pp. 85–90, Anchorage, Alaska, May 2007.

[24] Society of Automotive Engineers DSRC Committee, *SAE J2735 Dedicated Short Range Communications (DSRC) Message Set Dictionary*, 2016.

[25] I.-C. Chang, H.-T. Tai, F.-H. Yeh, D.-L. Hsieh, and S.-H. Chang, "A VANET-based A* route planning algorithm for travelling time- and energy-efficient GPS navigation app," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 794521, 14 pages, 2013.

[26] "Simulation of urban mobility (sumo)," http://sumo-sim.org/.

Research Article

Intergroup Joint Scheduling for Mitigating Asymmetric Uplink Interference in Self-Organizing Virtual Cell Networks

Ohyun Jo,[1] Gyung-Ho Hwang,[2] and Dong-Ho Cho[3]

[1]*Electronics and Telecommunications Research Institute (ETRI), Daejeon, Republic of Korea*
[2]*Department of Computer Engineering, Hanbat National University, Daejeon, Republic of Korea*
[3]*School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea*

Correspondence should be addressed to Gyung-Ho Hwang; gabriel@hanbat.ac.kr

We introduce the concept of self-organizing VCN (virtual cell network). Here self-organizing VCN topology for efficient operation will be configured, and the functions of the each element will be defined. Also, the operation scenarios of VCN will be described. Then, we propose an efficient scheduling algorithm that considers the asymmetry of interference between downlink and uplink to mitigate intercell interference with little computing overhead. The basic concept is to construct scheduling groups that consist of several users. Each user in a scheduling group is affiliated with a different cell. Then, the intercell groups are managed efficiently in the proposed VCNs. There is no need for the exchange of a lot of information among base stations to schedule the users over the entire network.

1. Introduction

Interference needs to be managed strictly, because, in recent years, the portion of the cell area that is overlapping with other cells has been increasing as cell sizes have been getting smaller. This increase of overlapped region means that the effect of intercell interference is increased, due to the increased cell edge area. Users located in the edge area of a cell are seriously affected by the increased interference. This is the main reason for the degradation of system performance. Conventional algorithms that have been deployed to tackle this problem are usually restricted by a single-cell optimization [1].

However, the news is not all bad. A new network topology that offers more coordination between antennas has been proposed for next-generation systems. This system environment is considered in recent communication specifications for improving system performance [2]. For instance, the third-generation partnership project (3GPP), which is the representative association for system standardization, introduces coordinated multipoint transmission/reception (CoMP) for joint processing among multiple base stations [3]. Each base station is connected to a coordinator, which controls cooperation among the base stations by collecting and passing of information. Intra-e-nodeB of the 3GPP is a further example of coordinated networks [4]. In an attempt to improve system performance further, more coordinated network topology has been proposed for 5th-generation systems [5, 6]. This system environment has featured prominently in recent communication specifications for improving system performance. For instance, each base station is connected to a coordinator, which controls cooperation. Some information for cooperation is exchanged or collected by the coordinator.

Motivated by the research described above, we herein propose an innovative network architecture and a group-scheduling algorithm for efficient resource management. Traditional cellular system, where each base station independently manages all processing within the coverage, has many obstacles in view of drastically improving the system performance due to interference. Particularly, when cell size is reduced, cell edge area increases and handoff occurs more frequently. Even though partial cooperation between base stations somehow helps this limitation, it also causes much overhead due to information exchange. Motivated by the above-mentioned recent research, we herein propose a coordinated VCN topology in which intercell interference is mitigated. The particular means of coordination that we
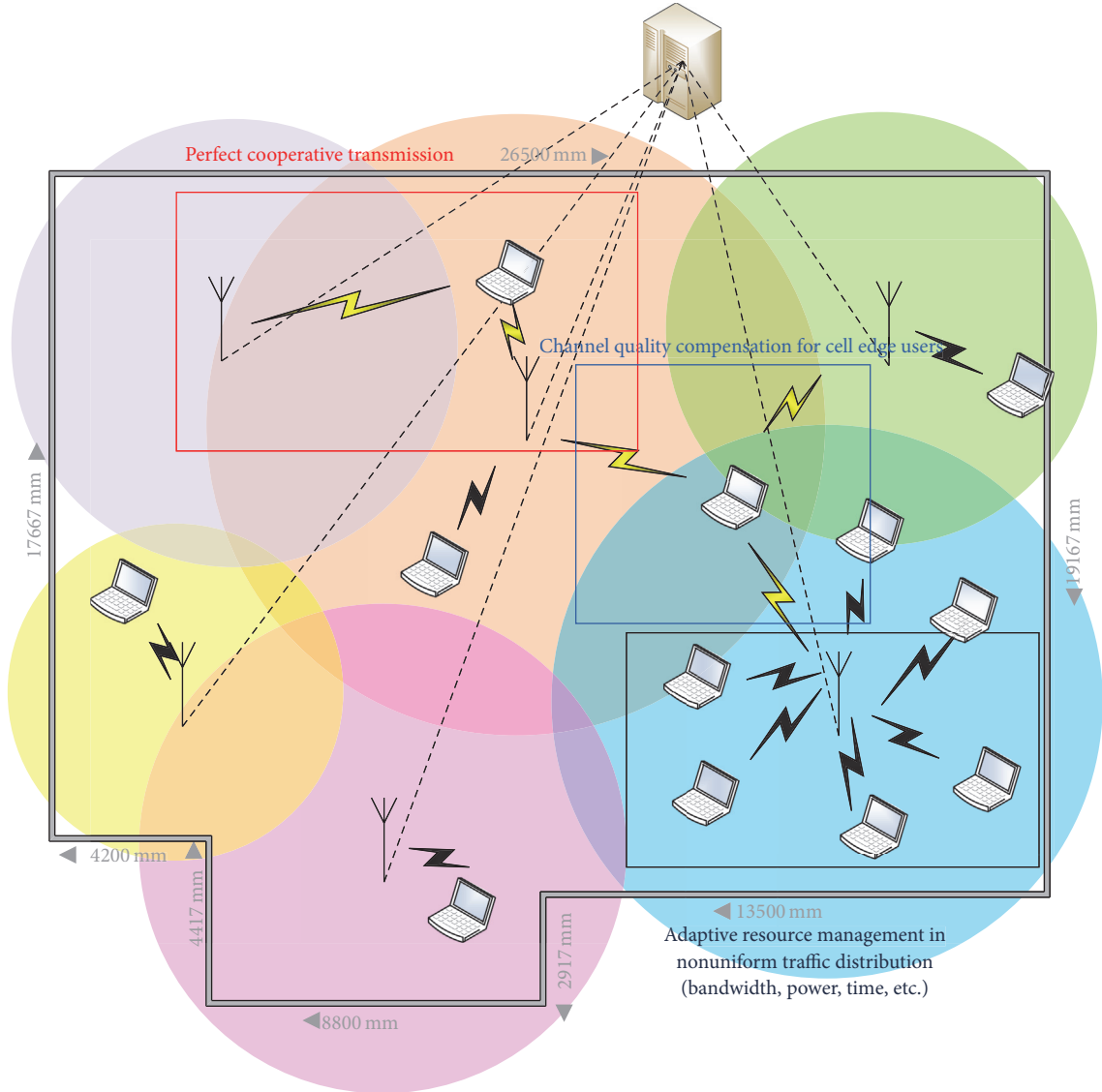
FIGURE 1: System configuration for VCN.

propose enable us to implement the joint scheduling of multiple base stations in the network efficiently, with low computational complexity. The basic idea of VCN is totally self-organizing network architecture to avoid signal overhead among base stations and make feasible full cooperation. Each cell which is represented as RAU (Radio Access Unit) in VCN only has minimal mandatory functions for working and is actually managed by one master unit so-called CU (Central Unit) in the proposed network topology. Moreover, multiple cells can be dynamically merged into one virtual cell which is also managed by CU. So, each cell corresponding to an RAU is not a real cell. This is why we call the new network topology the virtual cell network. The new network topology manages a new resource domain.

2. VCN System Configuration

The basic idea of VCN is totally centralized architecture to avoid signal overhead among base stations and make full

cooperation feasible. Each cell which is represented as RAU in VCN only has minimal mandatory functions for working and is actually managed by one master unit so-called CU in the proposed network topology.

New network architecture for VCN is given in Figure 1. VCN consists of two main components which are CU and RAU. CU has multiple high-level BS controllers and it is the core component in VCN. RAU has a low-level BS controller. And central controller in CU manages all BS controllers in the network. In order for an RAU to build up an independent single-cell, the RAU should have a connection to the high-level BS controller and the central controller. In the case that multiple RAUs are merged into one virtual cell, multiple RAUs are connected to one high-level BS controller and the central controller. The function set of high-level BS controller and low-level controller can be defined in two ways. The first case is that high-level BS controller has function set of RRC, MAC, and PHY, and low-level BS controller has function set
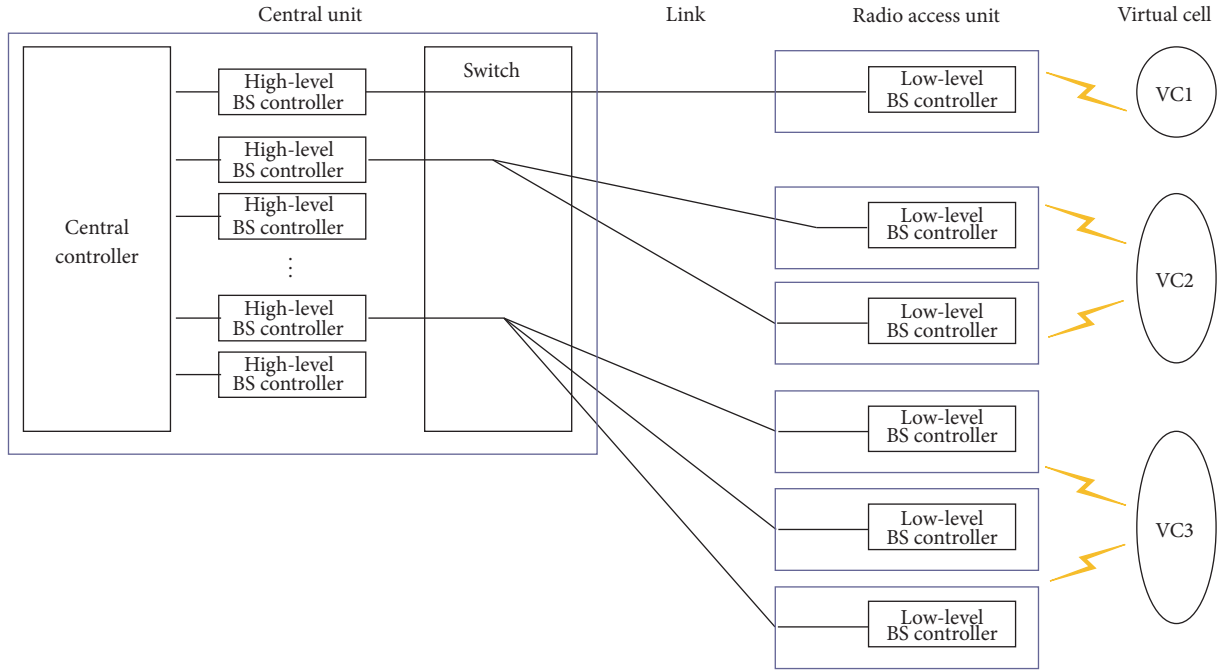
Figure 2: Basic conceptual diagram for VCN.

of RF and antenna. In this case, baseband signal is transmitted in the link between CU and RAU. The second case is the high-level BS controller that has function set of RRC, MA, PHY, and RF, and low-level BS controller has only function set of antenna. In this case, RF signal is transmitted in the link between CU and RAU.

Figure 2 shows the basic conceptual diagram for VCN. CU contains a switch module which carries out mapping function between high-level controller in CU and low-level controller in RAU. The connection is essential to make a virtual cell. To build up a cell, the functionalities of both controllers should be combined. If only one RAU is mapped to a high-level controller, the virtual cell contains only one antenna. In this case, the virtual cell can be considered as an independent normal cell which is just managed by CU. On the other hand, if multiple RAUs are connected to a high-level controller in CU, the virtual cell has multiple antennal. Therefore, the virtual cell can dynamically change the configuration according to the system environment. This is the main characteristic.

Figure 3 shows the functionality of CU. The backhaul interface is a submodule for the communication toward the outer networks, such as wired/wireless network. It is also useful for the cooperation between VCNs. Virtual cell database manages the overall information for the base station and terminals which are working in virtual cell mode. It contains channel information and traffic information and so on. Virtual cell manager carries out functions for managing virtual cell. It performs scheduling, resource allocation, and so on. User mobility manager manages user mobility in a virtual cell. In a virtual cell, new handover scenarios happen, because the cell structure is different from conventional communication systems. Users can move from VCN to

VCN (inter-VCN handover) and also can move to another antennal coverage area in a same virtual cell area (intra-VCN handover). Therefore, the complicated procedure should be carefully managed by this submodule. Intercell interference coordinator operates for the interference management in a virtual cell. The interference model is more complex in a virtual cell. Interference can be generated by neighbor antenna, neighbor VCN, or outer cellular network (in the case of overlaid network structure). In a virtual cell, careful interference management is required for avoiding performance degradation. Therefore, this submodule performs the important roles. Traffic QoS manager contains traffic information of users in a virtual cell. In a virtual cell, all kinds of traffic adaptive algorithm can be implemented based on sharing of traffic information. Finally, BS interface provides the connection between central controller and high-level BS controller in CU.

More advanced and efficient system configuration can be considered for improving the system performance. In this case, RAUs also should be more improved than the basic configuration to perform smarter operations. In this configuration, hierarchical resource management is possible. That makes it possible to increase the system efficiency. The system configuration is shown in Figure 4.

In this configuration, some smarter RAUs contain enhanced low-level BS controller. The RAU which has enhanced low-level BS controller can behave just like conventional pico cell or femto cell. The information of the cell is independently managed by the RAD by itself without any aid of CU. If an individual antenna builds up an independent cell, the enhanced low-level BS controller is activated. In this case, the RAU is not connected to high-level BS controller. Instead, it is directly connected to the central controller for enabling

Figure 3: CU (Central Unit) functionality.

Figure 4: Conceptual diagram for advanced VCN.

network management. When the RAU builds up a virtual cell with other RAU, the enhanced low-level BS controller is not activated and the low-level BS controller is connected to the high-level BS controller in CU. For the feasibility of this new configuration, RAU needs to have a switch module which can configure the mode of RAU. Also, the RAUs with enhanced low-level BS controller can provide backward compatibility with the conventional pico cell base stations or femto cell base stations.

RAUs which have enhanced low-level BS controller can build up a virtual cell by itself. In this case, system configuration is shown in Figure 5.

In this configuration, the functions of high-level BS controller are replaced with enhanced low-level BS controller. Also, the RAU is directly connected to central controller not to the high-level BS controller. In the case that any kind of cooperation and management is not required, the Master RAU manages the virtual cell totally independently without any aid of Central Unit.

When the network size is very huge, sometimes only one CU is not capable to manage all the network. In this case, each CU forms a virtual cell network independently, and the several virtual cell networks can cooperate through an improved system configuration as shown in Figure 6.

Figure 7 shows the system configuration in the case that some RAUs independently build up a virtual cell without CU. Here, the head RAU which replaces the functions of CU is called Master RAU.

For enabling the virtual cell system configuration, mobile terminals also should be improved. Mobile terminals in VCN can request to build up or break up a virtual cell in some cases, considering channel quality, QoS, RAU status, cell shape, and so on. The RAU needs to request some information from mobile terminals to decide on the network topology. When the network topology is changed, the mobile terminals should be notified with this change from RAU for adaptation. For this, mobile terminals can receive the information and configure the functionality. Therefore, some submodules for performing those operations are required at mobile terminals as shown in Figure 8. Also, signaling procedure is described in Figure 9.

3. Operation Scenario

When the new network topology is applied in real communication system, abundant advantages can be obtained as shown in Figure 1. For example, unintended resource waste can be reduced in nonuniform traffic distribution environments.
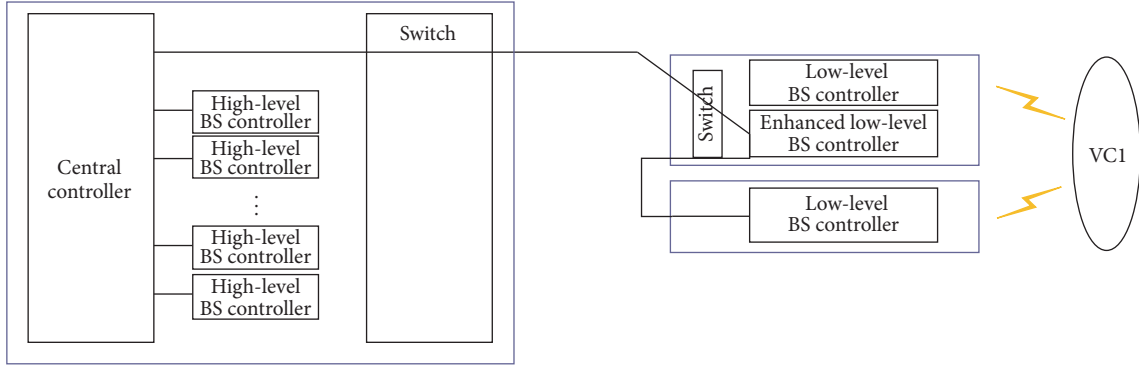
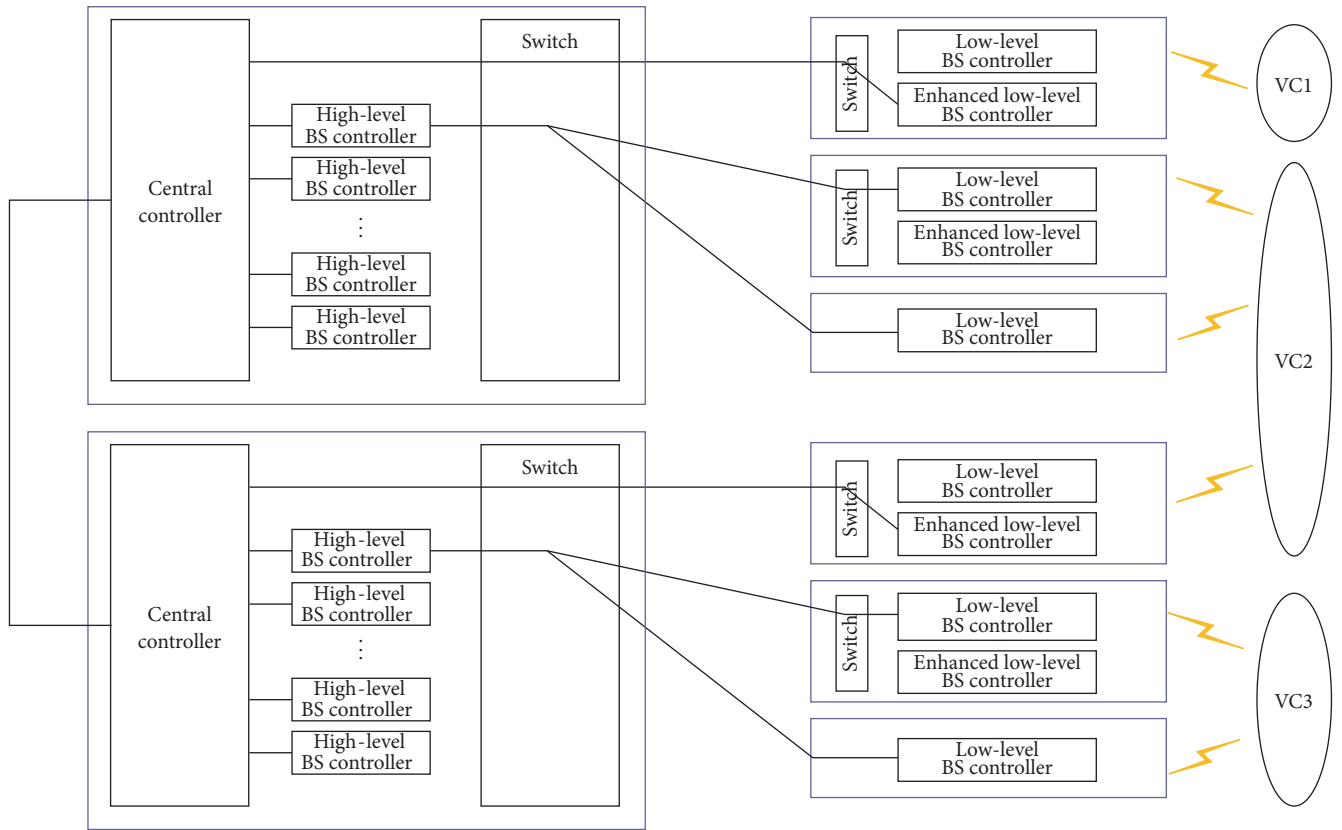FIGURE 5: Conceptual diagram for self-organizing VCN.

FIGURE 6: CU-concatenated VCN.

Also, full cooperation can increase the spectral efficiency, especially in the edge of the cell coverage. Scheduling and resource management also can be more efficient based on the information of traffic distribution, channel quality, and user mobility in all network.

Based on various network status, for example, traffic distribution, user priority, mobility, and so forth, VCN can dynamically change the form of a virtual cell. This is the basic and typical operation of VCN. Pretty much benefit is attained from this characteristic. If there is no user in a neighbor cell, the neighbor cell can be totally merged. In this case, the two

antennas are managed by a common high-level BS controller; then it acts like a distributed MIMO system. In other words, it forms a virtual cell. If there is an idle resource in a neighbor cell, the neighbor cell can be partially merged. In this case, the resource of the partially merged cell is managed hierarchically. In the VCN frame structure, some part can be managed by low-level BS controller independently, and the other part can be separately managed by high-level BS controller in CU for cooperation. Finally, if there is no idle resource in a neighbor cell, each cell operates separately. One user associates with only one antenna in this case.
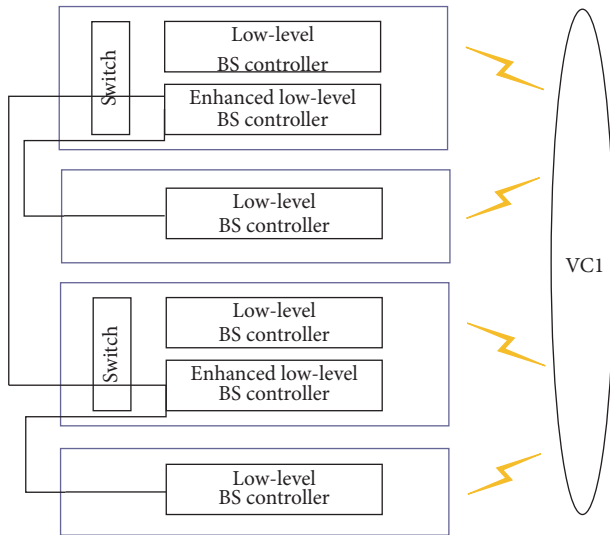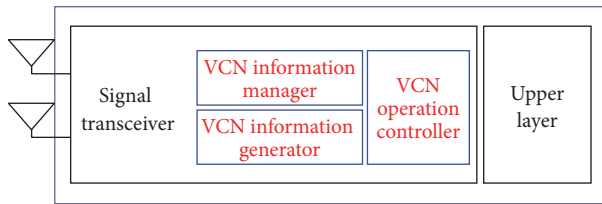
Figure 7: RAU-concatenated VCN.

Figure 8: Conceptual diagram of mobile terminal.
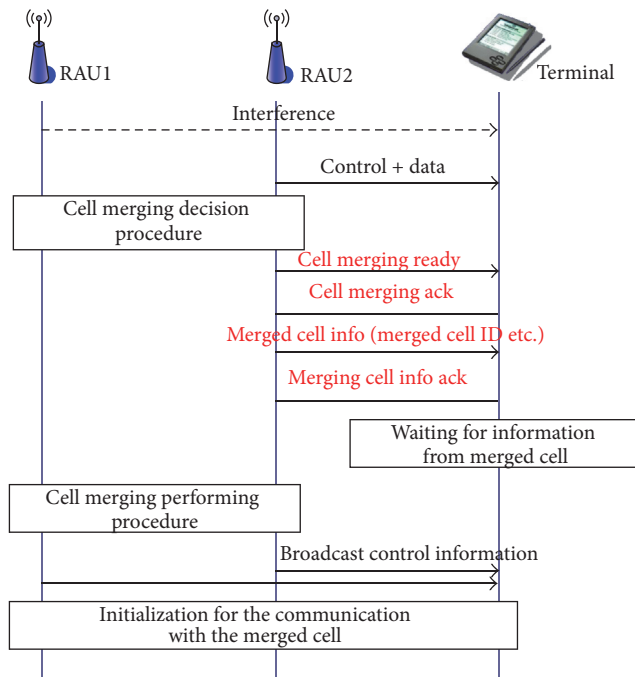
Figure 9: Signaling procedure for building up a virtual cell.
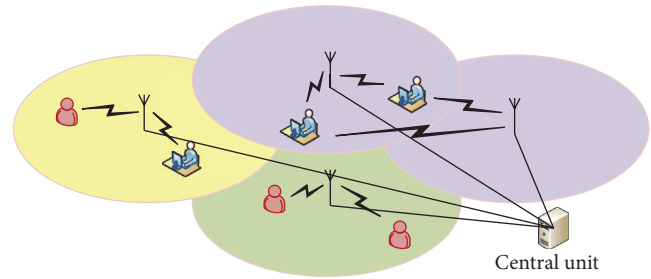
Figure 10: Complete cell merge scenario in VCN.
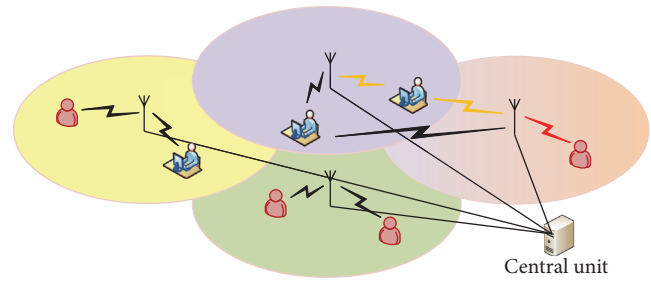
Figure 11: Partial cell merge scenario in VCN.
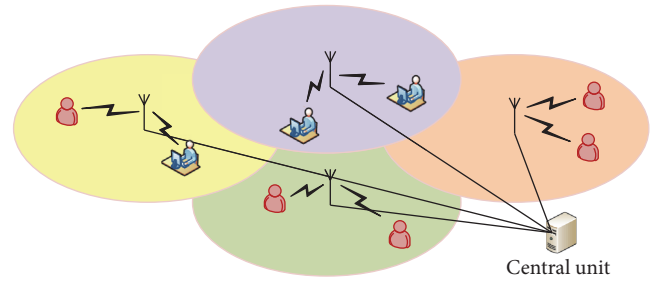
Figure 12: Cell separation scenario in VCN.

However, each cell can be managed by a common high-level BS controller for interference mitigation. These operations are shown in Figures 10, 11, and 12.

4. Asymmetry Interference Problem

There exists an asymmetric characteristic in interference between uplink and downlink. Mobile users in communication systems usually report the Channel Quality Index (CQI) based on the pilot signal from base stations. The pilot signal is a predefined signal pattern that enables mobile users to recognize the base stations. The base stations periodically transmit the pilot signal by using a predefined resource. Through this operation, mobile users can recognize the precise state of the channel and measure the Signal to Interference and Noise Ratio (SINR) in downlink. The situation for uplink is not so helpful. In this case, it is not easy for the BSs to determine the precise state of channels and measure the SINR, because they cannot recognize all the signal patterns that identify users of
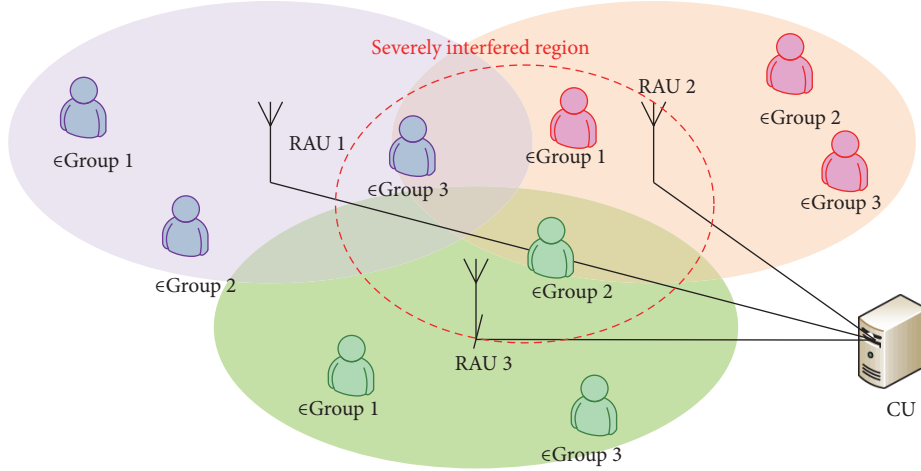
FIGURE 13: System model and the basic idea of the proposed intergroup joint scheduling algorithm.

neighbor cells. Moreover, mobile users may change location, unlike base stations, which are deployed in a fixed position. Consequently, uplink interference model is hard to predict. Therefore, it is much more difficult to manage interference in the uplink than in the downlink case. In most communication systems, the uplink channel quality is assumed to be the same as the downlink channel quality. Consequently, uplink scheduling is performed on the basis of the information about the channel quality that is obtained from CQI reports sent by mobile terminals. However, the CQI reports do not provide information that is correct for uplink. In some cases, even though the downlink SINR is very good, the uplink channel quality may, depending on the position of the mobile users, be pretty poor because of the asymmetry in interference. However, using the benefits of coordinated network topology, we can derive more accurate information about the uplink channel, thereby improving system performance.

5. Intergroup Joint Scheduling

5.1. Basic Idea. In VCNs, RAUs are able to cooperate through a CU, as shown in Figure 13. Here, we propose another approach for uplink scheduling using an intelligent network topology. The proposed approach, which we call intergroup joint scheduling, builds up uplink scheduling groups composed of multiple users who may be located in different cells. The members of a particular scheduling group are allocated to the same resource region by the CU. When the members of a scheduling group are allocated to the same resources, they are assumed to be hardly correlated with each other for the purposes of mitigating intercell interference. Figure 13 illustrates the basic idea of the intergroup scheduling.

5.2. Conditional SINR-Based Grouping. For the proposed uplink intergroup scheduling algorithm, we first define an exclusive cluster. The users in an exclusive cluster have low SINR in the long term in their channel. Usually, some of the users who have low SINR are located in the area that overlaps with other cells. In this area, the received signal strength from

other neighbor cells is quite high. However, that means the users in the edge of the cell can also interfere with the uplink signal transmission of the neighbor cell. Most of the members of an exclusive cluster are located in the edge of the cell. Therefore, these severe interferers in the cluster should not share resources. Then, we build up a scheduling group for a user in the exclusive cluster based on conditional SINR.

Step 1. The CU collects information from users in VCNs to make an exclusive cluster. The information collected by the CU contains the signal strength of users from neighbor RAU cells. The CU makes an exclusive cluster from those users who satisfy the following condition:

$$P_{rx} < \alpha \cdot \sum P_{\text{inter}}, \tag{1}$$

where P_{rx} is the received power from the corresponding RAU of a user, $\sum P_{\text{inter}}$ is the sum of the received power of neighbor RAUs, and α is a scaling factor related to the bound of the exclusive cluster area. If s users satisfy the above condition, P_{ec} is defined as a set composed of s users who are in the exclusive cluster and P_{nec} is defined as a set composed of $N - s$ users who are not in the exclusive cluster. Here, u represents a user in P_{ec}, ψ represents a user in P_{nec}, and N is the total number of users in the network.

$$P_{\text{ec}} = \{u_1, \ldots, u_s\}$$
$$P_{\text{nec}} = \{\psi_1, \ldots, \psi_{N-s}\}. \tag{2}$$

Step 2. In the second step, we find a dominant interferer of each u in P_{ec}. Here, we use the conditional SINR [7].

The conditional SINRs of subchannel n for user m in base station $\pi(m)$, that is, $\gamma_{m,n|\Psi=\{\}}^{\pi(m)}, \gamma_{m,n|\Psi=\{\psi_1\}}^{\pi(m)}, \gamma_{m,n|\Psi=\{\psi_2\}}^{\pi(m)}, \ldots, \gamma_{m,n|\Psi=\{\psi_{N-s}\}}^{\pi(m)}$, which correspond to restrictions of none, $\psi_1, \psi_2, \ldots, \psi_{N-s}$, respectively, are calculated from the SINR expression. $\pi(m)$ denotes the corresponding RAU of user m

$$\gamma_{m,n}^{\pi(m)} = \frac{P_c H_{m,n}^{\pi(m)}}{P_c \sum_{k \neq m} H_{k,n}^{\pi(m)} I_{k,n} + P_N}, \tag{3}$$

where P_c is the transmit power supplied on each subchannel derived from equal power distribution, P_N is the thermal noise power over the subchannel bandwidth, and $H_{m,n}^{\pi(m)}$ and $H_{k(k\neq m),n}^{\pi(m)}$ are the link gains to the registered RAU of user m and the interferer k, respectively. The indicators $I_{k,n}$ take the value 1 or 0 depending on whether or not the interferer k is scheduled in subchannel n. If user k is restricted in subchannel n, $I_{k,n}$ will be definitely 0.

Using the definition of conditional SINR, we calculate the conditional SINRs in the case that each user in P_{nec} is restricted. Then, to build up a scheduling group, we determine whether or not each conditional SINR satisfies the following condition.

$$c_{m,n|\Psi=\{\}}^{\pi(m)} > c_{m,n|\Psi=\{\psi_i\}}^{\pi(m)} + c_{\psi_i,n|\Psi=\{m\}}^{\pi(\psi_i)} \quad \text{for } i = 1,\dots,N-s. \quad (4)$$

Each term of the above equation is defined by a Shannon's equation as follows. Here, B is the bandwidth of subchannel n.

$$
\begin{aligned}
c_{m,n|\Psi=\{\}}^{\pi(m)} &= B \cdot \log_2\left(1 + \gamma_{m,n|\Psi=\{\}}^{\pi(m)}\right) \\
c_{m,n|\Psi=\{\psi_i\}}^{\pi(m)} &= B \cdot \log_2\left(1 + \gamma_{m,n|\Psi=\{\psi_i\}}^{\pi(m)}\right) \\
c_{\psi_i,n|\Psi=\{m\}}^{\pi(\psi_i)} &= B \cdot \log_2\left(1 + \gamma_{\psi_i,n|\Psi=\{m\}}^{\pi(\psi_i)}\right).
\end{aligned}
\quad (5)
$$

Using this condition as a basis, we define and identify preferred candidates of each member in P_{ec} for building a scheduling group. The preferred candidates are determined for every member in P_{ec} as follows. Then, S_m, which is a set of preferred candidates of user u_m in P_{ec}, can be expressed as

$$S_m = \left\{\psi_k c_{m,n|\Psi=\{\}}^{\pi(m)} > c_{m,n|\Psi=\{\psi_k\}}^{\pi(m)} + c_{\psi_k,n|\Psi=\{m\}}^{\pi(\psi_i)}\right\} \quad (6)$$
$$\text{for } m = 1,\dots,s.$$

Step 3. Now, we build the scheduling groups. The number of scheduling groups is equal to the cardinality of P_{ec}. If no other user satisfies the above condition for user u_m, the corresponding scheduling group for user u_m is composed of user m alone. Then, user m occupies a subchannel on its own in the network. To build a scheduling group for u_m, we elect an appropriate member out of S_m for every neighbor RAU. Therefore, at most one user that is being served by a neighbor RAU belongs to the scheduling group for u_m.

Let g_m be the scheduling group for u_m in P_{ec}. Then, g_m is defined by following algorithm.

Algorithm 1 (scheduling group).

(1) Pick one user u_m from P_{ec}.

(2) Determine the preferred candidate set S_m for u_m.

(3) Select a RAU Φ from the neighbor RAU of u_m

(4) Perform $g_m \ni \psi_j$, if $\psi_j = \arg\{\max_j c_{m,n|\Psi=\{\psi_j\}}^{\pi(m)} + c_{\psi_j,n|\Psi=\{m\}}^{\pi(\psi_j)}$, $\psi_i \in S_m$, and $\pi(\psi_j) = \Phi\}$ exists.

(5) Update $S_m = S_m - \{\psi_j\}$.

(6) Repeat (3)~(5) for all neighbor base stations of u_m.

(7) Repeat (1)~(6) for every user in P_{ec}.

After building the scheduling groups, we calculate the capacity of each group for each subchannel.

$$c_{g_i,n} = \sum_{k=1}^{n(g_i)} c_{k,n}^{\pi(k)}. \quad (7)$$

Here, $n(g_i)$ is the cardinality of set g_i. Then, we allocate the subchannels for each scheduling group according to the following policy when N_{ch} is the number of subchannels and $n(P_{\text{ec}})$ is the cardinality of P_{ec}:

$$
\begin{aligned}
\max \quad & \sum_{n=1}^{N_{\text{ch}}} \sum_{i=1}^{n(P_{\text{ec}})} I_{g_i,n} \cdot c_{g_i,n} \\
\text{s.t.} \quad & \sum_{i=1}^{n(P_{\text{ec}})} I_{g_i,n} = 1 \quad \text{for each } n.
\end{aligned}
\quad (8)
$$

Note that the above optimal scheduling problem for overall network has exactly the same computational complexity as that of the simple scheduling problem for single-cell optimization [8]. This is because the unit of scheduling is expanded from a user to a scheduling group. Therefore, the complexity will not be increased compared to single-cell scheduling problem. And the scheduling algorithm can be implemented much more easily in practical systems.

6. Performance Evaluation

We assume that users are distributed randomly over the network area. Let $N_{i,\text{in}}$ be the number of users in the inner side of RAU i in the exclusive cluster. Let $N_{i,\text{out}}$ be the number of users in the outer side of RAU i in the virtual cell who do not cause severe interference to each other because they are not located in the exclusive cluster. Further, there exist N_r RAUs in the virtual cell. Then, the total number of users is defined as $N = \sum_{i=1}^{N_r} N_{i,\text{in}} + N_{i,\text{out}}$.

If the network resources are utilized fully, every subchannel will be occupied by a certain user. Then, in this case, every subchannel is used by N_r users in the entire network because at least one user occupies the subchannel in each cell. Therefore, the probability P_{inter} that a user in P_{ec} receives strong interference in a subchannel due to a severe interferer is calculated as follows:

$$P_{\text{inter}} = 1 - \frac{\prod_{i=1}^{N_r} \binom{N_r}{N_{i,\text{out}}}}{\prod_{i=1}^{N_r} \binom{N_r}{N_{i,\text{in}}+N_{i,\text{out}}}}. \quad (9)$$

Figure 14 shows the variation of P_{inter} according to $\sum_i^{N_r} N_{i,\text{in}}$. Here, $\sum_i^{N_r}(N_{i,\text{in}} + N_{i,\text{out}})$ is fixed as 15. Then, we vary $\sum_i^{N_r} N_{i,\text{in}}$ from 5 to 10. In the case that $\sum_i^{N_r} N_{i,\text{in}}$ is 5, the probability that the some of inner-area users receive strong interference is greater than 0.7. Moreover, in the case that $\sum_i^{N_r} N_{i,\text{in}}$ is 10, the probability approaches 1. However, in the proposed algorithm, this problem can be avoided.
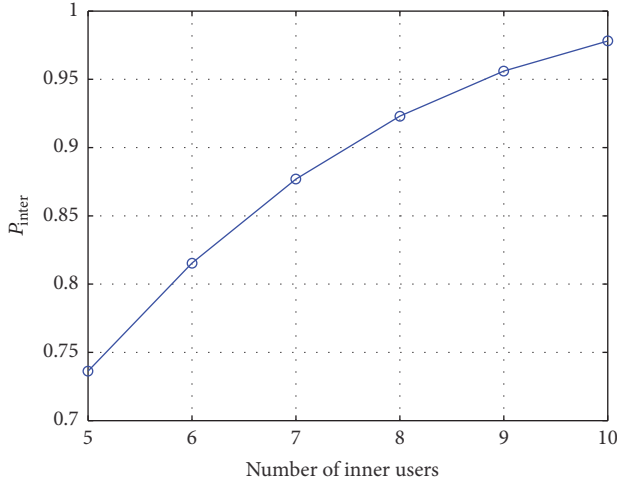
FIGURE 14: Probability that a user in an exclusive cluster receives strong interference due to a severe interferer versus number of inner-area users.
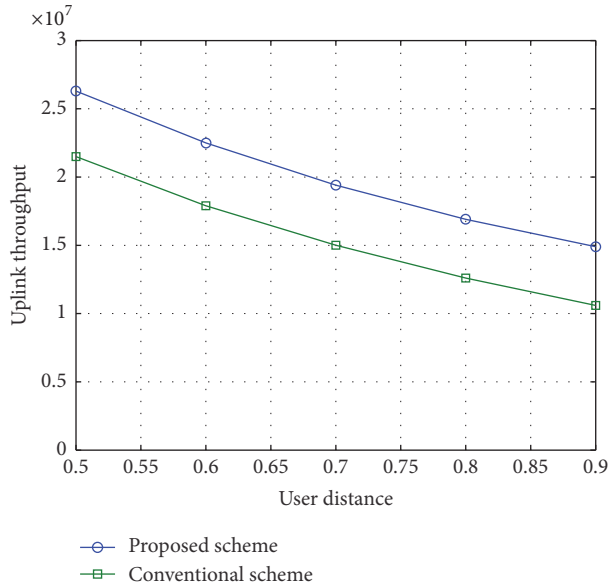

FIGURE 15: Uplink throughput versus user distance.

location to the coverage of the RAU is plotted. The proposed scheme outperforms the conventional scheme in terms of throughput. In the conventional scheme, asymmetric inter-cell interference cannot be taken into account because the uplink channel quality of neighbor cell users is not used for uplink scheduling. Thus, uplink traffic is scheduled by using only the information of the downlink channel that is measured from the pilot signal. In this conventional network configuration [1, 2, 8], the performance of conventional scheme is measured by using single-cell optimization of uplink scheduling based on downlink channel information.

The uplink throughput falls as the distance of the users from the RAU increases because the path loss increases as the distance increases. In contrast, when the proposed algorithm is used, the capacity gain in uplink increases up to 50%, relative to the conventional algorithm. This is because the effect of intercell interference is reduced by applying the proposed algorithm. Moreover, the asymmetric property of uplink intercell interference is intensified as the distance of the user is increased. This is because the uplink interference to neighbor cells is increased and the signal strength to the serving cell is decreased as the distance of the user is increased. Thus, the downlink channel information is more incorrect for uplink scheduling of the users in outer regions. Also, the amount of performance enhancement is relatively increased in case that the users are located in outer regions as the distance of the user is increased.

7. Conclusion

We have proposed an efficient intergroup scheduling algorithm for self-organizing virtual cell networks. To overcome the problem of the asymmetry in interference between downlink and uplink, we introduced scheduling groups, in which the same resource is shared by multiple users. Then, network-wise scheduling was performed to mitigate uplink interference. The numerical and simulation results show that the proposed algorithm can reduce the probability that near interferers in the network share the same resource, thereby improving uplink performance.

Competing Interests

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

References

[1] J. Huang, V. Subramanian, R. Agrawal, and R. Berry, "Joint scheduling and resource allocation in uplink OFDM systems for broadband wireless access networks," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 2, pp. 226–234, 2009.

[2] A. Kumar, Y. Liu, and A. Wason, "LTE-advanced: the roadmap to 4G mobile wireless networks," *Global Journal of Computer Science and Technology*, vol. 10, no. 4, pp. 50–53, 2010.

[3] 3GPP, "Aspects of joint processing for downlink CoMP, 3GPP TSG RAN WG1," 3GPP R1-090942, 2009.

[4] 3GPP R1-083050, Inter-eNB and Intra-eNB Schemes for CoMP in LTE-Advanced, 3GPP TSG RAN WG1, August 2008.

To analyze the performance of the proposed algorithm, we validated the proposed algorithm in a coordinated network by computer simulations using MATLAB. Then, we observed the uplink throughput in the following scenario.

We used the following parameters. The frequency was 2 GHz, the available bandwidth was 10 MHz, and the transmission power was 100 mW, which is usual in indoor communication systems. An Okumura-Hata channel model for indoor environments was used. The number of coordinated RAUs was assumed to be 3 and five users were located randomly in the area served by each base station. The distance between RAUs was 20 m, and the number of subchannels was 18.

Figure 15 shows the uplink throughput according to the location of the users. On the x-axis, the ratio of the user

[5] V. Jungnickel, K. Manolakis, W. Zirwas et al., "The role of small cells, coordinated multipoint, and massive MIMO in 5G," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 44–51, 2014.

[6] P. K. Agyapong, M. Iwamura, D. Staehle, W. Kiess, and A. Benjebbour, "Design considerations for a 5G network architecture," *IEEE Communications Magazine*, vol. 52, no. 11, pp. 65–75, 2014.

[7] M. Rahman and H. Yanikomeroglu, "Enhancing cell-edge performance: a downlink dynamic interference avoidance scheme with inter-cell coordination," *IEEE Transactions on Wireless Communications*, vol. 9, no. 4, pp. 1414–1425, 2010.

[8] B. Rengarajan, A. L. Stolyar, and H. Viswanathan, "Self-organizing dynamic fractional frequency reuse on the uplink of OFDMA systems," in *Proceedings of the 44th Annual Conference on Information Sciences and Systems (CISS '10)*, pp. 1–6, Princeton, NJ, USA, March 2010.

Research Article

Utility Maximization for Load Optimization in Cellular/WLAN Interworking Network Based on Generalized Benders Decomposition

Fanqin Zhou, Lei Feng, Peng Yu, Wenjing Li, and Luoming Meng

State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China

Correspondence should be addressed to Wenjing Li; wjli@bupt.edu.cn

Load steering is widely accepted as a key SON function in cellular/WLAN interworking network. To investigate load optimizing from a perspective of system utilization maximization more than just offloading to improve APs' usage, a utility maximization (UTMAX) optimization model and an ASRAO algorithm based on generalized Benders Decomposition are proposed in this paper. UTMAX is to maximize the sum of logarithmic utility functions of user data rate by jointly optimizing user association and resource allocation. To maintain the flexibility of resource allocation, a parameter β is added to the utility function, where smaller β means more resources can be allocated to edge users. As a result, it reflects a tradeoff between improvements in user throughput fairness and system total throughput. UTMAX turns out to be a mixed integer nonlinear programming, which is intractable intuitively. So ASRAO is proposed to solve it optimally and effectively, and an optional phase for expediting ASRAO is proposed by using relaxation and approximation techniques, which reduces nearly 10% iterations and time needed by normal ASRAO from simulation results. The results also show UTMAX's good effects on improving WLAN usage and edge user throughput.

1. Introduction

With the popularization of intelligent mobile terminals and enrichment of Internet services, it arouses surging demands for mobile Internet access in mobile users. So higher requirements on access capacity and data rate are put forward to commercially operated mobile networks. As spectrum band authorized to a mobile network is so scarce, academic and industrial research groups are continuously working on more spectrum-efficient radio communication technologies. Besides endeavors in this direction, taking advantages of available extra frequency bands seems much more economical and practical, which leads to a common deed among network operators of deploying WLAN in public areas, because WiFi works on the open ISM frequency band and wireless traffic can be offloaded from cellular network (CN) to WLAN.

A typical traffic offloading case in cellular/WLAN coexisted scenario is illustrated in Figure 1. In Figure 1(a), users (or UEs without distinction hereafter) are served dominantly by cellular base stations (BSs), while APs are underutilized. As service areas of CN and public WLAN overlap, to perform offloading, we can simply adjust user association and optimize resource allocation accordingly. The optimized scenario is expected to be as in Figure 1(b), where some users handover from BSs to nearby APs.

Initially, public access points (APs) in WLAN were deployed mainly in traffic hot-spots where many mobile users gathered, resulting in high usage of APs. However, with public WLAN gradually being a scale large enough to cover almost all urban areas, the problem is increasingly highlighted that most public WLAN APs are underutilized. The strict power limit of ISM devices is a root cause, while a lack of load distribution optimization approaches in present cellular/WLAN coexisting networks also plays a role.

So load steering between cellular and WLAN was an important research case in SEMAFOUR, a project in EU Framework Program 7 (EU-FP7) for developing multi-RAT/multilayer self-optimizing network (SON) function and integrated SON management system [1]. Its resulting approaches
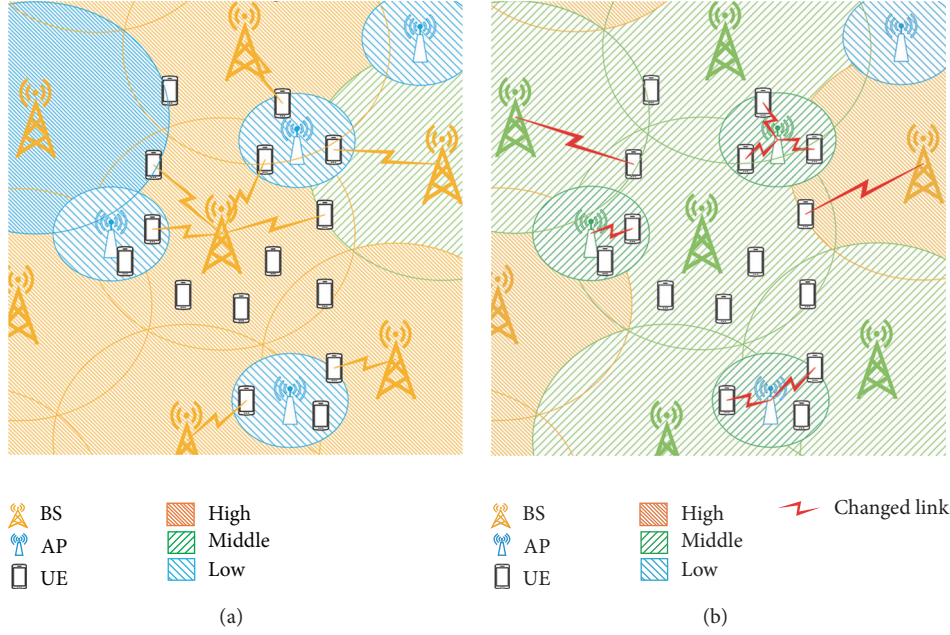
FIGURE 1: Offloading in cellular/WLAN interworking network.

pay more attention to applicability in practice, while a theoretical performance bound is not well investigated. A typical approach proposed is to make AP serve users as many as possible, which is sure to improve AP's usage to the most but will not produce the highest system utilization or overall users' satisfaction. Therefore, offloading is deemed as a basic purpose, but load optimizing is a further goal in cellular/WLAN interworking networks.

Moreover, providing better satisfactory data rate for users with the same infrastructure means more economical utilizing of investment, which is of more significance to commercially operated networks. To this end, optimizing user association and resource allocation to achieve ultimate overall users' satisfaction in cellular/WLAN interworking network are to be investigated in this paper.

Many prior works related to load optimizing in cellular/WLAN interworking networks have been surveyed in [2, 3]. In this paper, we roughly group the related approaches into three categories, according to whether limited, fair, or sufficient information exchange between CN and WLAN is required.

(1) UE-decision based approaches are widely used where information exchange is difficult, and statistical methods are used to estimate proper handover trigger criteria for individual UE on aspects, such as the average amount of data that could be transferred through CN or WLAN [4], the holding and continuing of an existing traffic data stream [5], the hysteresis for handover between CN and WLAN [6], the load status in a BS or AP [7], and the partial packet success [8]. These approaches focus on designing practical rules for individual UE to make proper handover decision without assists from networks.

(2) Network-assisted UE-decision based approaches become possible with more information attainable about a target network through information exchange, such as the achievable data rate in a target eNB or AP [9, 10], UE's velocity, and received signal strength of target AP [11]. These works use different methods, such as TOPSIS [10] and fuzzy logical system [11] to process multidimensional information obtained from network to help UE make better handover choice.

(3) Centralized optimization approaches are applicable when sufficient information is achievable. They usually focus on providing theoretical performance bounds for coexisting networks on various optimization goals, such as balancing resource utilization [12], minimizing power consumption [13, 14], maximizing quality of experience [15], and maximizing multilink aggregation [16]. These approaches may not be intended for application in actual networks, but they provide theoretical bounds for target optimization goals, which is significant for the design of practical network optimizing methods.

The existence of various types of approaches is partially due to the different requirements in actual practices of network optimizing and theoretical analyses in academic researches and largely due to the development of network technologies. In the early stage of cellular/WLAN coexisting network, the lack of information exchange between noninterworking CN and WLAN caused great difficulty in controlling user access or handover in a centralized way. Therefore, works at that time focused on UE-decision based approaches.

To improve the performance of coexisting networks, information exchange between networks is a prerequisite.

For this purpose, 3GPP proposed solutions for interworking between 3GPP network and WLAN in [17], and a media independent handover (MIH) standard was developed by IEEE in [18]. With more information attainable, especially that about load status in a target BS or AP (we use AN to denote a BS or AP hereafter), individual UE can even estimate its achievable data rate in a target BS or AP, so better handover decision can be made. On these bases, network-assisted UE-decision based became popular.

As multiple users may contend for access chances and resources, UE-decision based schemes neither can guarantee a resulting decision always to be the optimal nor can tell where a performance bound goes from a whole network perspective. Therefore, centralized optimization approaches are widely investigated to derive such a theoretical performance bound. These approaches are usually of high computational complexity; however, deriving such a performance bound will give guidance and direction for more practical approaches with better tradeoff between optimality and complexity.

Therefore, this paper is to derive a centralized logarithmic utility maximization model to improve overall users' satisfaction by jointly optimizing user association and resource allocation. The key contributions are summarized as follows.

(i) A utility maximization (UTMAX) optimization model is formulated to depict the users' satisfaction degree maximization by jointly planning association selection and resource allocation (ASRA). Instead of directly using a logarithmic function of data rate to model user's satisfaction of a data rate, which produces identical resource allocation in a same BS and thus lack flexibly in resource allocation, we add a control parameter β to it. With smaller β value, the users with poor channel conditions are to be assigned more resources and vice versa. Therefore, a tradeoff between user throughput fairness and system throughput gain can be achieved by selecting proper β value.

(ii) The derived UTMAX has both binary and continuous variables for user association and resource allocation, respectively, which is a mixed integer nonlinear programming (MINLP) and seems intractable intuitively. To solve it optimally and effectively, we devise an ASRAO algorithm on the basis of Bender Decomposition (BD) framework [19, 20]. We also propose optional procedures for ASRAO to expedite the computation of UTMAX using relaxation and approximation techniques [21], which dramatically reduces the number of iterations needed before convergence, especially when a UTMAX instance is in large scale. In addition, the convergence of our algorithms is validated both theoretically and experimentally in this paper.

(iii) Our work contributes to SON. The UTMAX and its corresponding ASRAO can serve directly in the scheme-planning module, which forms a self-optimizing closed loop together with monitor, evaluation, and execution modules, to generate the optimal ASRA scheme through a centralized way. It also provides a performance upper bound for load optimization on purposes of a system preferred tradeoff between overall throughput and end user satisfaction. So It can guide the development of novel suboptimal but less computationally complex algorithms by evaluating the performance gaps between them.

The rest of the paper are arranged as follows. In Section 2 we formulate the user data rate and propose the UTMAX optimization model. Section 3 analyzes a special property of UTMAX, and, to exploit this property, we propose the ASRAO algorithm based on generalized Benders Decomposition (GBD) to solve it. Optional procedures for ASRAO to expedite the computing of UTMAX are also given in this section. In Section 4, we evaluate the performance of (A-)ASRAO and effect of UTMAX through extensive simulations. The paper will be concluded in Section 5.

2. System Model and Problem Formulation

A cellular/WLAN interworking network is composed of J BSs and K APs, serving I UEs. We denote by $\mathcal{J}$, $\mathcal{K}$, and $\mathcal{I}$ the sets of all BSs, APs, and UEs, respectively. UE is served uniquely by a BS or AP, from which it gets largest received power. The association between UE i and AN l is indicated by binary variable x_{il}, which equals 1 if UE i is served by AN or 0 otherwise as in (1). A matrix $\mathbf{X} = (x_{il})_{i \in \mathcal{I}, l \in \mathcal{J} \cup \mathcal{K}}$ represents associations between all UEs and ANs. UE is uniquely served by one AN, namely, a BS or AP, which forms the constraint in (2).

$$x_{il} = \begin{cases} 1, & \text{UE } i \text{ is served by AN } l, \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

$$\sum_{l \in \mathcal{J} \cup \mathcal{K}} x_{il} = 1. \tag{2}$$

2.1. Channel and Resource Model of CN. The access rate of a UE in CN is determined by two factors, the channel condition and the allocated resource. $\Lambda_{\mathcal{J}} = (\lambda_{ij})$, $i \in \mathcal{I}$, $j \in \mathcal{J}$ represents the resource allocation scheme in CN, in which λ_{ij} denotes the ratio of resource allocated to UE i from BS j. To assure a basic quality of service (QoS), a minimal ratio of resource, λ_{ε}, should be allocated to UE i from BS j if $x_{ij} = 1$, and $\lambda_{ij} = 0$ if $x_{ij} = 0$, which can be summarized in (3). The channel condition between UE i and BS j can be indicated by SINR_{ij}, as in (4), in which Pr_{ij} is the received signal power of UE i from BS j, and α is a coefficient of cochannel interference. According to Shannon Formula, the data rate of UE i served by BS j, r_{ij} can be expressed in (5), wherein $R_{ij} = W_j \log(1 + \text{SINR}_{ij})$.

$$\lambda_{\varepsilon} x_{ij} \leq \lambda_{ij} \leq x_{ij}, \tag{3}$$

$$\text{SINR}_{ij} = \frac{\text{Pr}_{ij}}{\alpha \cdot \sum_{i' \in B/\{i\}} \text{Pr}_{i'j} + \sigma^2}, \tag{4}$$

$$r_{ij} = \lambda_{ij} W_j \log_2 \left(1 + \text{SINR}_{ij} \right) = \lambda_{ij} W_j \gamma_{ij} = \lambda_{ij} R_{ij}. \tag{5}$$

2.2. Channel and Resource Model of WLAN. Due to the sparse deployment of public WLAN APs and the limited transmit power, for simplification, the interferences between WLAN APs will not be taken into consideration. Reference [9] gives the way to evaluate the service rate of UE $i \in \mathcal{I}$ accessing public WLAN AP $k \in \mathcal{K}$, as in

$$r_{ik} = C_{\text{WLAN}} \cdot \frac{\text{MCS}_{\text{WLAN}}(\text{Pr}_{ik})}{\max\{\text{MCS}_{\text{WLAN}}(\cdot)\}} \quad i \in \mathcal{I}, \ k \in \mathcal{K}, \quad (6)$$

wherein C_{WLAN} denotes the average data rate evaluated by AP when the modulation and coding scheme (MCS) with the highest spectrum efficiency is used and Pr_{ik} is the received power strength of UE i from WLAN AP k. $\text{MCS}_{\text{WLAN}}(\cdot)$ is corresponding to each possible Pr_{ik} in MCS table. Taking resource allocation into consideration, we have $C_{\text{WLAN}} = C_{\text{WLAN\_FULL}}\lambda_{ik}$, where $C_{\text{WLAN\_FULL}}$ is the available data rate if the observing UE occupies resources exclusively in AP. As interference to AP is not taken into consideration and only a limited number of UEs are served in each AP, we assume $C_{\text{WLAN\_FULL}}/\max\{\text{MCS}_{\text{WLAN}}(\cdot)\} = 1$. Therefore, $r_{ik} = \lambda_{ik}R_{ik}$.

AP uses Round Robin (RR) for packet data scheduling, which from long time average perspective can be viewed as UEs in AP equally sharing channel resources (7). Network Operators commonly limit the data rate of UEs in WLAN APs to avoid resources being occupied dominantly by user requesting big volume of data with good channel condition. For simplification, in this paper we assume the ratio of resources occupied by UE in AP has an upper bound $\lambda_m \leq 1$ as in (8). In addition, UEs contend to access AP via shared wireless channel, and too many UEs would severely deteriorate the throughput performance, so the number of UEs served simultaneously by AP has an upper bound N_m as in (10), which amounts to setting a lower bound to resources allocated to UE. Hence, we have the following constraints on resource allocation in AP.

$$\lambda_{ik} = \lambda_k^{\text{avr}} x_{ik}, \quad i \in \mathcal{I}_k, \ k \in \mathcal{K}, \quad (7)$$

$$\lambda_{ik} \leq \lambda_m x_{ik}, \quad (8)$$

$$\sum_{i \in \mathcal{I}} \lambda_{ik} \leq 1, \quad (9)$$

$$\sum_{i \in \mathcal{I}} x_{ik} \leq N_m. \quad (10)$$

As UE is served uniquely by a BS or AP, the overall data rate of UE i can be formulated as $r_i = \sum_i \lambda_{il}R_{il}$. It should be noted that R_{il} is not a variable to be optimized in our model that means more accurate approaches for full band rate estimation can be used, which is not a key point in this paper and we will pay no more attention to it. After formulating the data rate of UE in CN and WLAN, we take a utility function perspective to investigate a utility maximization problem for the data rate r_i of each UE i in the system, to get the optimal association and resource allocation.

2.3. Utility Maximization Problem. For utility maximization, utility function $U_i(\cdot)$ is often chosen to be continuously differentiable, monotonically increasing, and strictly concave.

UE i gets utility $U_i(r_i)$ when its receiving rate is r_i. The overall system utility can then be represented as $\sum_i U_i(r_i)$. Maximize it with some practical constraints, we can get the general utility maximization problem. $U_i(\cdot)$ should be chosen carefully. If $U_i(\cdot)$ is a linear function, $\max \sum_i U_i(r_i)$ turns out to be a throughput maximization and it produces a trivial solution, where each BS allocates resources only to its user with best channel quality. It is not a satisfactory solution for a multiple user system. So we seek a utility function that would naturally encourage fairness resource allocation among users. To this end, a logarithmic function is widely used. It is concave and has diminishing returns. This property approximates the fact that a well-served user has low priority in getting more resources, so it encourages load balancing and improves overall users' satisfaction.

However, logarithmic utility function, like in [15], always results in equal resource allocation among users in a same BS. Because $\sum_i \log(R_i \lambda_i) = \sum_i \log(R_i) + \sum_i \log(\lambda_i)$, the utility of individual BS has nothing to do with R_i, and $\sum_i \log(\lambda_i) = \log(\prod_i \lambda_i) \leq n \log(\sum_i \lambda_i/n)$ always hold. It reaches maximum only when resources are all shared equally, $\lambda_{i_1} = \lambda_{i_2}$, $\forall i_1, i_2 \in \mathcal{I}_l$. This lacks flexibility in resource allocation, so we slightly adjust this logarithm utility function and add a BS related parameter β_l to r_i, to make resource allocation scheme represent certain system preference, like caring more about fairness or demanding higher system throughput. If $b_l = 0$, it means an identical resource allocation. If $b_i > 0$, users with better channel condition get more resources and vice versa. The logarithmic utility function then becomes $\log(\sum_l R_{il}\lambda_{il} + \beta_l x_{il})$. As x_{il} is binary variable and $\sum_l x_{il} = 1$, $\log(\sum_l R_{il}\lambda_{il} + \beta_l x_{il})$ can be transformed into $\sum_l x_{il}\log(\sum_l R_{il}\lambda_{il} + \beta_l)$. And our utility maximization problem is formed as P1 in P1, where for simplicity we assume in CN $\lambda_m = 1$, and in WLAN $\lambda_\varepsilon = 0$.

$$\text{P1: } \max_{(\Lambda, \mathbf{X})} \quad \Phi(\Lambda, \mathbf{X}),$$

$$\text{s.t.} \quad \Phi(\Lambda, \mathbf{X})$$

$$= \left(\sum_i \sum_l \left(x_{il} \log \left(\sum_l R_{il}\lambda_{il} + \beta_l \right) \right) \right),$$

$$\sum_l x_{il} = 1,$$

$$\sum_i \lambda_{il} \leq 1,$$

$$\lambda_\varepsilon x_{il} \leq \lambda_{il} \leq \lambda_m x_{il}, \quad (11)$$

$$\lambda_{ik} = \lambda_k^{\text{avr}} x_{ik},$$

$$\sum_{i \in \mathcal{I}} x_{ik} \leq N_m,$$

$$x_{il} \in \{0, 1\},$$

$$i \in \mathcal{I},$$

$$k \in \mathcal{K},$$

$$l \in \mathcal{J} \cup K.$$

P1 has both binary and continuous variables and nonlinear part in goal function. These properties make it an MINLP, which is intractable intuitively. However, this problem has a special structure, utilizing which we propose ASRAO algorithm based on GBD to solve it effectively. The problem analyses and ASRAO algorithm will be presented in next section.

3. ASRAO Algorithm

3.1. Problem Analyses. Assume association indicator $\mathbf{X}$ is fixed, then each UE i is served by a specific AN l, and each AN l has a fixed UE collection $\mathscr{I}_1 = \{i \mid x_{il} = 1\}$. UE i contributes $\log(R_{il}\lambda_{il} + \beta_l)$ to utility of AN l. Its value is determined by λ_{il}, which is only related to AN l. So resource allocation in each AN can be solved independently with given user association $\dot{\mathbf{X}}$. Therefore, $\Phi(\mathbf{\Lambda}, \dot{\mathbf{X}})$ can be identically transformed to

$$\Phi\left(\mathbf{\Lambda}, \dot{\mathbf{X}}\right) = \sum_{\substack{i\in\mathscr{I}, \\ l\equiv\arg\{x_{il}=1\}}} \log\left(R_{il}\lambda_{il} + \beta_l\right)$$

$$= \sum_{l\in\mathscr{J}\cup K}\left(\sum_{i\in I_l}\log\left(R_{il}\lambda_{il} + \beta_l\right)\right) \quad (12)$$

$$= \sum_{l\in\mathscr{J}\cup K}\Phi_l\left(\boldsymbol{\lambda}_l, \dot{\mathbf{x}}_l\right),$$

where $\Phi_l(\boldsymbol{\lambda}_l, \dot{\mathbf{x}}_l) = \sum_{i\in I_l}\log(R_{il}\lambda_{il} + \beta_l)$ is the utility of AN l. So we get a subproblem of P1 on condition of $\dot{\mathbf{X}}$.

$$\max_{(\mathbf{\Lambda})}\quad \sum_{i\in I_l}\log\left(R_{il}\lambda_{il} + b_l\right)$$

$$\text{s.t.}\quad \sum_i\lambda_{il} \leq 1,$$

$$\lambda_\varepsilon\dot{x}_{il} \leq \lambda_{il} \leq \lambda_m\dot{x}_{il}, \quad (13)$$

$$\lambda_{ik} = \lambda_k^{\text{avr}}\dot{x}_{ik}.$$

This is a classical problem in information theory and a water filling (WF) principle is specially developed to solve it, which is derived from Karush-Kuhn-Tucker (KKT) optimal conditions. It provides direction solutions to problems with the form like (13). To enable its applicability in P1, we extend WF to a multi-AN case and denote its results as MWF, short for multi-AN water filling. With $\mathbf{X}$ fixed as $\dot{\mathbf{X}}$ in P1, we have a MWF applicable problem as in

$$\max_{(\mathbf{\Lambda})}\quad \Phi\left(\mathbf{\Lambda}, \dot{\mathbf{X}}\right),$$

$$\text{s.t.}\quad \Phi\left(\mathbf{\Lambda}, \dot{\mathbf{X}}\right)$$

$$= \left(\sum_i\sum_l\left(\dot{x}_{il}\log\left(\sum_l R_{il}\lambda_{il} + \beta_l\right)\right)\right), \quad (14)$$

$$\lambda_\varepsilon\dot{x}_{il} \leq \lambda_{il} \leq \lambda_m\dot{x}_{il},$$

$$\sum_i\lambda_{il} \leq 1,$$

$$\lambda_{ik} = \lambda_k^{\text{avr}}x_{ik}.$$

According to generalized duality theory, the dual function of $\Phi(\mathbf{\Lambda}, \dot{\mathbf{X}})$ is as in

$$\mathscr{L}\left(\mathbf{\Lambda}, \boldsymbol{\nu}, \boldsymbol{\omega}, \boldsymbol{\gamma}, \dot{\mathbf{X}}\right)$$

$$= \left(\sum_i\sum_l\left(\dot{x}_{il}\log\left(\sum_l R_{il}\lambda_{il} + \beta_l\right)\right)\right)$$

$$- \sum_l\nu_l\left(\sum_i\lambda_{il} - 1\right) - \sum_i\sum_l\omega_{il}\left(\lambda_{il} - \dot{x}_{il}\right) \quad (15)$$

$$+ \sum_i\sum_l\gamma_{il}\left(\lambda_{il} - \varepsilon\dot{x}_{il}\right) - \sum_i\sum_k\mu_{ik}\left(\lambda_{ik} - \dot{x}_{ik}\lambda_k^{\text{avr}}\right),$$

where $\boldsymbol{\nu}$, $\boldsymbol{\omega}$, $\boldsymbol{\gamma}$ are nonnegative relaxed variables and $\boldsymbol{\mu}$ is nonzero. The local optimal is reached, when the following KKT conditions are satisfied.

$$\frac{\partial\Phi\left(\mathbf{\Lambda}\right)}{\partial\lambda_{ij}} = \frac{R_{ij}}{\sum_j R_{ij}\lambda_{ij} + \sum_l\dot{x}_{ij}b_j} - \nu_j - \omega_{ij}$$
$$+ \gamma_{ij} = 0,$$

$$\frac{\partial\Phi\left(\mathbf{\Lambda}\right)}{\partial\lambda_{ik}} = \frac{R_{ik}}{\sum_k R_{ik}\lambda_{ik} + \sum_l\dot{x}_{ik}b_k} - \nu_k - \omega_{ik}$$
$$+ \gamma_{ik} - \mu_{ik} = 0,$$

$$\frac{\partial\Phi\left(\mathbf{\Lambda}\right)}{\partial\lambda_k^{\text{avr}}} = \mu_{ik}\dot{x}_{ik} = 0, \quad (16)$$

$$\omega_{il}\left(\lambda_{il} - \dot{x}_{il}\right) = 0,$$

$$\nu_l\left(\sum_i\lambda_{il} - 1\right) = 0,$$

$$\gamma_{il}\left(\lambda_{il} - \varepsilon\dot{x}_{il}\right) = 0,$$

$$\mu_{ik}\left(\lambda_{ik} - \dot{x}_{ik}\lambda_k^{\text{avr}}\right) = 0,$$

$$\nu_l, \omega_{il}, \gamma_{il} \in [0, +\infty), \quad \mu_{ik} \neq 0.$$

The solutions to (16) are as follows. MWF results for CN is as in

$$\nu_j^* = \left(\sum_i\dot{\overline{x}}_{ij}\right)\left(\overline{\rho} + \sum_i\frac{b_j\dot{\overline{x}}_{ij}}{R_{ij}}\right)^{-1},$$

$$\lambda_{ij}^* = \max\left\{\dot{x}_{ij}\left(\frac{1}{\nu_j^*} - \frac{\beta_j}{R_{ij}}\right), 0\right\} + \lambda_\varepsilon\dot{x}_{ij},$$

$$\gamma_{ij} = \begin{cases} -\left(\dfrac{R_{ij}}{R_{ij'}\lambda_{ij'} + \beta_{j'}} - \nu_j^*\right), & \dfrac{R_{ij}}{R_{ij'}\lambda_{ij'} + \beta_{j'}} \leq \nu_j^*, \\ 0, & \text{else,} \end{cases} \quad (17)$$

$$\omega_{ij} = \left(\frac{R_{ij}}{R_{ij'}\lambda_{ij'} + \beta_{j'}} - \nu_j^*\right) + \gamma_{ij}.$$

The following (18) is MWF results for WLAN.

$$\lambda_k^{\mathrm{avr}} = \begin{cases} \min\left\{\lambda_m, \dfrac{1}{\sum_i \dot{x}_{ik}}\right\}, & \sum_i \dot{x}_{ik} > 0, \\ 0 & \sum_i \dot{x}_{ik} = 0, \end{cases}$$

$$\lambda_{ik}^* = \lambda_k^{\mathrm{avr}} \dot{x}_{ik},$$

$$\nu_k^* = \begin{cases} \min\left\{\dfrac{R_{ik}}{\beta_k + \sum_i \dot{x}_{ik}}\right\}, & \lambda_m \sum_i \dot{x}_{ik} \ge 1, \\ 0, & \lambda_m \sum_i \dot{x}_{ik} < 1, \end{cases}$$

$$\mu_{ik} = \begin{cases} -\left(\dfrac{R_{ik}}{R_{ik'}\lambda_{ik'} + \beta_{k'}} - \nu_k^*\right), & \lambda_m \sum_i \dot{x}_{ik} > 1, \\ 1, & \lambda_m \sum_i \dot{x}_{ik} \le 1, \end{cases} \quad (18)$$

$$\gamma_{ik}$$
$$= \begin{cases} -\left(\dfrac{R_{ik}}{R_{ik'}\lambda_{ik'} + \beta_{k'}} - \nu_k^*\right), & \dfrac{R_{ik}}{R_{ik'}\lambda_{ik'} + \beta_{k'}} < \nu_k^*, \\ 0, & \text{else}, \end{cases}$$

$$\omega_{ik} = \left(\dfrac{R_{ik}}{R_{ik'}\lambda_{ik'} + \beta_{k'}} - \nu_k^*\right) + \gamma_{ik} + \mu_{ik}.$$

Using above MWF results, we can directly calculate optimal $\mathbf{\Lambda}$ with a given $\dot{\mathbf{X}}$. To utilize MWF in solving problem P1, an optimal $\mathbf{X}$ should be retrieved. However, as $\mathbf{X}$ is a matrix of binary variable entries and $\mathbf{\Lambda}$ is closely related with $\mathbf{X}$, it is difficult to find such an optimal. In this paper, instead of directly solving an optimal $\mathbf{X}$, we devise ASRAO algorithm to solve it iteratively and effectively. ASRAO is originated from BD, which provides a framework for addressing mixed integer programming problems by decomposing it into two smaller subproblems and solving them iteratively. The details are in the sequel.

3.2. Generalized Benders Decomposition.
BD is originally proposed to solve mixed integer linear programming (MILP). Instead of solving all variables and constraints simultaneously, it decomposes an MILP into two subproblems, a master problem (MP) and a slave problem (SP), and the original MILP can be solved by solving MP and SP iteratively. In BD frameworks, MP is an MILP problem which consists of all integer variables in the original problem, while SP is a Linear Programming problem with all the continuous variables from the original problem. In each iteration, BD utilizes an extreme point or extreme ray derived from the dual of SP to generate an optimality cut or feasibility cut to trim the feasible domain of MP, and MP will eventually reach the same optima as the original MILP.

GBD extends the BD approach to a more general class of problems by adopting nonlinear duality theory, and some nonlinear problems are brought into range. When solving MINLP, GBD follows the framework of BD, but SP is derived as NLP. To generate the feasibility cuts and optimality cuts,

GBD utilized KKT conditions to calculate an extreme point or extreme ray of the Lagrange dual functions of NLP SP. A GBD applicable problem has a general form as in

$$\max_{(\mathbf{x},\mathbf{y})} \quad F(\mathbf{x}, \mathbf{y}),$$
$$\text{s.t.} \quad G(\mathbf{x}, \mathbf{y}) \ge 0,$$
$$\qquad H(\mathbf{x}, \mathbf{y}) = 0, \qquad (19)$$
$$\qquad \mathbf{x} \in \{0, 1\}_{m\times 1},$$
$$\qquad \mathbf{y} \in D \subseteq \mathbb{R}_+^{n\times 1}.$$

It should meet following situations: (a) for fixed $\mathbf{x}$, (19) separates into a number of independent optimization problems; (b) for fixed $\mathbf{x}$, (19) assumes a well-known special structure that efficient solution procedures are available; (c) (19) may be not concave program in $\mathbf{x}$ and $\mathbf{y}$ jointly, but fixing $\mathbf{x}$ renders it so in $\mathbf{y}$. According to the analyses, UTMAX obviously satisfies these conditions.

The initial MP of (19) in GBD procedures is as follows:

$$\max_{(\mathbf{x},\eta)} \quad \eta,$$
$$\text{s.t.} \quad G(\mathbf{x}) \ge 0,$$
$$\qquad H(\mathbf{x}) = 0, \qquad (20)$$
$$\qquad \eta \in \mathbb{R},$$
$$\qquad \mathbf{x} \in E \subseteq \{0, 1\}_{m\times 1},$$

where $G(\mathbf{x}) \ge 0$ and $H(\mathbf{x}) = 0$ are constraints from (19) that only related with $\mathbf{x}$. Solving a solution $(\mathbf{x}^{(t)}, \eta)$ from MP, SP can be derived by simply fixing $\mathbf{x}$ in (19) as $\mathbf{x}^{(t)}$,

$$\max_{(\mathbf{y})} \quad F\left(\mathbf{x}^{(t)}, \mathbf{y}\right),$$
$$\text{s.t.} \quad G\left(\mathbf{x}^{(t)}, \mathbf{y}\right) \ge 0,$$
$$\qquad H\left(\mathbf{x}^{(t)}, \mathbf{y}\right) = 0, \qquad (21)$$
$$\qquad \mathbf{y} \in D \subseteq \mathbb{R}_+^{n\times 1},$$

where $\mathbf{x}^{(t)}$ is the solution to MP in tth iteration. The Lagrange function to problem (14) is $\mathscr{L}(\mathbf{y}, \nu, \upsilon) = F(\mathbf{x}^{(t)}, \mathbf{y}) + \nu G(\mathbf{x}^{(t)}, \mathbf{y}) + \upsilon H(\mathbf{x}^{(t)}, \mathbf{y})$. If optimal solution $\mathbf{y}^*$ to SP exists, it can be derived by solving KKT point $(\mathbf{y}^{(u)}, \nu^{(u)}, \mu^{(u)})$, and due to the strictly convex property of SP, $\mathscr{L}(\mathbf{y}^{(u)}, \nu^{(u)}, \mu^{(u)}) = \max_{(\mathbf{y})} F(\mathbf{x}^{(t)}, \mathbf{y})$. After getting $(\mathbf{y}^{(u)}, \nu^{(u)}, \mu^{(u)})$, an optimality cut $\Gamma(\mathbf{x}, \mathbf{y}^{(u)}, \nu^{(u)}, \upsilon^{(u)}) \ge \eta$ can be generated and added to MP, which actually constructs an upper bound for MP. If the optimal solution to SP does not exist, a feasibility cut should be added to cut off the infeasible $\mathbf{x}^{(v)} = \mathbf{x}^{(t)}$. Specifically, the feasible cut is $\sum_{i\in Y_v} x_i - \sum_{i\in N_v} x_i \le |Y_v| - 1$, derived from $\mathbf{x} \ne \mathbf{x}^{(v)}$, wherein Y_v is the set of indexes indicating the nonzero elements in $\mathbf{x}^{(v)}$ and N_v is that indicating the zero ones.

Therefore, the MP in tth iteration is as

$$
\max_{(\mathbf{x},\eta)} \quad \eta,
$$

$$
\text{s.t.} \quad \Gamma\left(\mathbf{x}, \mathbf{y}^{(u)}, \nu^{(u)}, v^{(u)}\right) \geq \eta, \quad \forall u = 1, 2, \ldots, t_1,
$$

$$
\sum_{(i,l) \in Y_\nu} - \sum_{(i,l) \in N_\nu} \leq |Y_\nu| - 1, \quad \forall \nu = 1, 2, \ldots, t_2, \tag{22}
$$

$$
G(x) \geq 0,
$$

$$
H(x) = 0,
$$

$$
\mathbf{x} \in (0, 1)_{m \times 1},
$$

where $t_1 + t_2 = t$. Based on above problem decomposition framework, we propose the following ASRAO algorithm to solve P1.

3.3. ASRAO Algorithm.
As is depicted in BD framework, problem P1 can be decomposed into MP and SP as in (26) and (24), separately. The initial MP to be solved is

$$
\max_{(\mathbf{X})} \quad \eta,
$$

$$
\text{s.t.} \quad \sum_{l \in \mathcal{J} \cup \mathcal{K}} x_{il} = 1,
$$

$$
\sum_{i \in \mathcal{J}} x_{ik} \leq N_m, \tag{23}
$$

$$
x_{il} \in \{0, 1\},
$$

$$
i \in \mathcal{I},
$$

$$
l \in \mathcal{J} \cup \mathcal{K}.
$$

In the tth iteration, denote the solution to MP as $(\mathbf{X}^{(t)}, \eta^{(t)})$, and the SP has the form

$$
\max_{(\mathbf{\Lambda})} \quad \Phi\left(\mathbf{\Lambda}, \mathbf{X}^{(t)}\right),
$$

$$
\text{s.t.} \quad \left(\mathbf{\Lambda}, \mathbf{X}^{(t)}\right)
$$

$$
= \left(\sum_i \sum_l \left(x_{il}^{(t)} \log\left(\sum_l R_{il}\lambda_{il} + \beta_l\right)\right)\right), \tag{24}
$$

$$
\lambda_\varepsilon x_{il}^{(t)} \leq \lambda_{il} \leq \lambda_m x_{il}^{(t)},
$$

$$
\lambda_{ik} = \lambda_{ik}^{\text{avr}} x_{ik},
$$

$$
\sum_{i \in \mathcal{J}} \lambda_{ij} \leq 1, \quad \lambda_{ij} \in \mathbb{R}_+.
$$

If optimal solution exists in problem (24), the KKT point is $(\mathbf{\Lambda}^{(u)}, \boldsymbol{\nu}^{(u)}, \boldsymbol{\omega}^{(u)}, \boldsymbol{\gamma}^{(u)}, \boldsymbol{\mu}^{(u)}) = (\mathbf{\Lambda}_\mathcal{J}^{(t)}, \boldsymbol{\nu}^{(t)}, \boldsymbol{\omega}^{(t)}, \boldsymbol{\gamma}^{(t)}, \boldsymbol{\mu}^{(t)})$, and

an optimal cut $\Gamma(\mathbf{X}, \mathbf{\Lambda}^{(u)}, \boldsymbol{\nu}^{(u)}, \boldsymbol{\omega}^{(u)}, \boldsymbol{\gamma}^{(u)}, \boldsymbol{\mu}^{(u)}) \geq \eta$ should be added to MP, wherein

$$
\Gamma\left(\mathbf{X}, \mathbf{\Lambda}^{(u)}, \boldsymbol{\nu}^{(u)}, \boldsymbol{\omega}^{(u)}, \boldsymbol{\gamma}^{(u)}, \boldsymbol{\mu}^{(u)}\right)
$$

$$
= \left(\sum_i \sum_l \left(x_{il}\log\left(\sum_l R_{il}\lambda_{il}^{(u)} + \beta_l\right)\right)\right)
$$

$$
- \sum_l \nu_l^{(u)}\left(\sum_i \lambda_{il}^{(u)} - 1\right)
$$

$$
- \sum_i \sum_l \omega_{il}^{(u)}\left(\lambda_{il}^{(u)} - \lambda_m x_{il}\right) \tag{25}
$$

$$
+ \sum_i \sum_l \gamma_{il}^{(u)}\left(\lambda_{il}^{(u)} - \lambda_\varepsilon x_{il}\right)
$$

$$
- \sum_i \sum_k^{(u)} \mu_{ik}^{(u)}\left(\lambda_{ik}^{(u)} - x_{ik}\lambda_k^{\text{avr}(u)}\right).
$$

Due to strong duality of SP in (24), we have $\Phi(\mathbf{\Lambda}^{(u)}, \mathbf{X}^{(t)}) = \Gamma(\mathbf{X}^{(t)}, \mathbf{\Lambda}^{(u)}, \boldsymbol{\nu}^{(u)}, \boldsymbol{\omega}^{(u)}, \boldsymbol{\gamma}^{(u)}, \boldsymbol{\mu}^{(u)})$. If an optimal solution does not exist, a feasibility cut $\sum_{(i,l) \in Y_\nu} x_{il} - \sum_{(i,l) \in N_\nu} x_{il} \leq |Y_\nu| - 1$ is added to MP to cut off the infeasible $\mathbf{X}^{(\nu)} = \mathbf{X}^{(t)}$, where $Y_\nu = \{(i,l) \mid x_{il} = 1, i \in \mathcal{I}, l \in \mathcal{J} \cup \mathcal{K}\}$ and $N_\nu = \{(i,l) \mid x_{il} = 0, i \in \mathcal{I}, l \in \mathcal{J} \cup \mathcal{K}\}$. And the derived MP in tth iteration is as (26) and $t_1 + t_2 = t$.

$$
\max_{(\mathbf{X}_\mathcal{J}, \mathbf{X}_\mathcal{K}, \eta)} \quad \eta,
$$

$$
\text{s.t.} \quad \Gamma\left(\mathbf{X}, \mathbf{\Lambda}^{(u)}, \boldsymbol{\nu}^{(u)}, \boldsymbol{\omega}^{(u)}, \boldsymbol{\gamma}^{(u)}, \boldsymbol{\mu}^{(u)}\right) \geq \eta,
$$

$$
\forall u = 1, \ldots, t_1,
$$

$$
\sum_{(i,l) \in Y_\nu} x_{il} - \sum_{(i,l) \in N_\nu} x_{il} \leq |Y_\nu| - 1,
$$

$$
\forall \nu = 1, \ldots, t_2, \tag{26}
$$

$$
\sum_{l \in \mathcal{I} \cup \mathcal{J}} x_{il} = 1,
$$

$$
\sum_{i \in \mathcal{J}} x_{ik} \leq N_m,
$$

$$
x_{il} \in \{0, 1\},
$$

$$
i \in \mathcal{I},
$$

$$
l \in \mathcal{J} \cup \mathcal{J}.
$$

Denote the lower bound and upper bound of problem P1 as $LB^{(t)}$ and $UB^{(t)}$, respectively. They can be derived from problem (24) and (26), which will be proved later in Lemma 1. The iteration procedure terminates when the gap between $UB^{(t)}$ and $LB^{(t)}$ becomes zero.

The pseudocodes of the above procedures are as shown in Algorithm 1.

(1) **Initialize**: $UB^{(0)} = +\infty$, $LB^{(0)} = -\infty$, Let $t = 0$, $u = 0$, $v = 0$.
(2) **repeat**
(3)   Let $t = t + 1$ and solve (26) to obtain current optimal solution $(\mathbf{X}^{(t)}, \boldsymbol{\eta}^{(t)})$, $UB^{(t)} = \eta^{(t)}$
(4)   **if** (24) with $\mathbf{X}^{(t)}$ is bounded: $u = u + 1$, and solve it with MWF to get a KKT point $(\boldsymbol{\Lambda}^{(u)}, \boldsymbol{\nu}^{(u)}, \boldsymbol{\omega}^{(u)}, \boldsymbol{\gamma}^{(u)}, \boldsymbol{\mu}^{(u)}) =$
          $(\boldsymbol{\Lambda}^{(t)}, \boldsymbol{\nu}^{(t)}, \boldsymbol{\omega}^{(t)}, \boldsymbol{\gamma}^{(t)}, \boldsymbol{\mu}^{(t)})$. The lower bound is set to $LB^{(t)} = \max\{\max_{1 \le s \le t}\{\Phi(\boldsymbol{\Lambda}^{(s)}, \mathbf{X}^{(s)})\}, LB^{(0)}\}$ and
          add $\Gamma(\mathbf{X}, \boldsymbol{\Lambda}^{(u)}, \boldsymbol{\nu}^{(u)}, \boldsymbol{\omega}^{(u)}, \boldsymbol{\gamma}^{(u)}, \boldsymbol{\mu}^{(u)}) \ge \eta$ to (26)
(5)   **elseif** (24) with $\mathbf{X}^{(t)}$ is infeasible: $v = v + 1$, $\mathbf{X}^{(v)} = \mathbf{X}^{(t)}$ $UB^{(t)} = UB^{(t-1)}$, and add $\sum_{(i,l) \in Y_v} x_{il} - \sum_{(i,l) \in N_v} x_{il} \le |Y_v| - 1$ to (24).
(6)   **else** as (24) is unbounded, P1 is unbounded, ASRAO stops and return an indication of infeasible problem.
(7) **until** $UB^{(t)} - LB^{(t)} \le 0$
(8) Return $(\mathbf{X}^*, \boldsymbol{\Lambda}^*) = (\mathbf{X}^{(t)}, \boldsymbol{\Lambda}^{(t)})$ as the optimal solution to problem P1.

ALGORITHM 1: The ASRAO algorithm.

Lemma 1. *In tth iteration, denote the values of goal functions in problem (24) and (26) as $L^{(t)}$ and $U^{(t)}$, and then $LB^{(t)} = \max_{0 \le s \le t}\{L^{(t)}\} = \max_{0 \le s \le t}\{\Phi(\boldsymbol{\Lambda}^{(s)}, \mathbf{X}^{(s)})\}$ and $UB^{(t)} = L^{(t)} = \eta^{(t)}$.*

Proof. First, we prove $UB^{(t)} = U^{(t)}$. Due to the strictly convex property, P1 has the strong duality, by which we have P1 which is equivalent to

$$\max_{(\mathbf{X})} \left\{ \min_{\boldsymbol{\nu}, \boldsymbol{\omega}, \boldsymbol{\gamma}, \boldsymbol{\mu}} \left\{ \sup_{(\boldsymbol{\Lambda})} \left\{ \Gamma\left(\mathbf{X}_{\mathcal{F}}, \mathbf{X}_{\mathcal{K}}, \boldsymbol{\Lambda}_{\mathcal{F}}, \boldsymbol{\nu}\right) \right\} \right\} \right\},$$

$$\text{s.t.} \quad \sum_{l \in \mathcal{F} \cup \mathcal{J}} x_{il} = 1,$$

$$\sum_{i \in \mathcal{F}} x_{ik} \le N_m, \quad (27)$$

$$x_{il} \in \{0, 1\},$$

$$i \in \mathcal{F},$$

$$l \in \mathcal{F} \cup \mathcal{J}.$$

Therefore, problem P1 and (27) have the same optimal solution and objective value of goal function. Denote the optimal solution of P1 as $(\mathbf{X}^*, \boldsymbol{\Lambda}^*)$ and the corresponding goal function value as $\Phi(\boldsymbol{\Lambda}^*, \mathbf{X}^*) = (\sum_i \sum_l (x_{il}^* \log(\sum_l R_{il} \lambda_{il}^* + \beta_l))) = M^*$. In tth iteration, the optimal solution to problem (26) is $(\mathbf{X}^{(t)}, \eta^{(t)})$. Since (26) is a relaxation of (27), it follows that $\Phi(\boldsymbol{\Lambda}^*, \mathbf{X}^*) \le U^{(t)} = \eta^{(t)}$. Therefore, $U^{(t)}$ is a upper bound of the objective function in P1.

Next, we prove $LB^{(t)} = \max_{0 \le s \le t}\{L^{(s)}\}$ and $L^{(s)} = \Phi(\boldsymbol{\Lambda}^{(s)}, \mathbf{X}^{(s)})$ is an upper bound of P1. Whether $L^{(t)}$ is bounded is decided by (24). If $L^{(s)} = -\infty$ ($\forall 0 \le s \le t$), $LB^{(t)} = \max_{1 \le s \le t}\{\Phi(\boldsymbol{\Lambda}^{(s)}, \mathbf{X}^{(s)})\} = -\infty$, which is obviously the lower bound of P1, so we focus on the case in which problem (24) is bounded. If (24) is bounded, $LB^{(t)} > -\infty$. We denoted $\underline{s} = \arg\max_{1 \le s \le t}\{\Phi(\boldsymbol{\Lambda}^{(s)}, \mathbf{X}^{(s)})\}$, and thereby $LB^{(t)} = L^{(\underline{s})} = \Phi(\boldsymbol{\Lambda}^{(\underline{s})}, \mathbf{X}^{(\underline{s})})$. If $LB^{(t)}$ is not the lower bound of P1, $LB^{(t)} > M^*$. Due to the strong duality of problem (24), $L^{(\underline{s})} = \Phi(\boldsymbol{\Lambda}^{(\underline{s})}, \mathbf{X}^{(\underline{s})}) > M^* = \Phi(\boldsymbol{\Lambda}^*, \mathbf{X}^*)$. This means $(\boldsymbol{\Lambda}^{(\underline{s})}, \mathbf{X}^{(\underline{s})})$ generates a larger objective function value of P1 than that $(\boldsymbol{\Lambda}^*, \mathbf{X}^*)$ does, which is in contradiction to the fact

that $(\boldsymbol{\Lambda}^*, \mathbf{X}^*)$ is the optimal solution of P1. Therefore, $LB^{(t)} = \max_{0 \ge s \ge t}\{L^{(s)}\}$ is a lower bound of P1. □

Theorem 2. *The ASRAO algorithm converges to a global optimal solution to P1 with finite number of iterations.*

Proof. As is indicated in [19, 20], there always exists a pointed convex polyhedral cone $\mathcal{C}$ that makes problem (24) and (26) equivalent to

$$\max_{(\mathbf{X}, \eta)} \quad \eta,$$

$$\text{s.t.} \quad (\mathbf{X}, \eta) \in \mathcal{C},$$

$$\sum_{l \in \mathcal{F} \cup \mathcal{K}} x_{il} = 1,$$

$$\sum_{i \in \mathcal{F}} x_{ik} \le N_m, \quad (28)$$

$$x_{il} \in \{0, 1\},$$

$$\forall i \in \mathcal{F},$$

$$\forall k \in \mathcal{K},$$

$$\forall l \in \mathcal{J} \cup \mathcal{K},$$

wherein $\mathcal{C}$ can be expressed as a convex hull of finitely many half-lines. In each iteration of ASRAO the set of solution $(\mathbf{X}, \eta)$ to problem (26) is shrunk by introducing a new extreme half-line being either an optimality cut $\Gamma(\mathbf{X}, \boldsymbol{\Lambda}^{(u)}, \boldsymbol{\nu}^{(u)}, \boldsymbol{\omega}^{(u)}, \boldsymbol{\gamma}^{(u)}, \boldsymbol{\mu}^{(u)}) \ge \eta$ or a feasibility cut $\sum_{(i,l) \in Y_v} x_{il} - \sum_{(i,l) \in N_v} x_{il} \le |Y_v| - 1$. Because $\mathcal{C}$ is a convex hull of finally many extreme half-lines and the new introduced extreme half-line is different from the preceding ones, the complete set of constraints determining the $\mathcal{C}$ can be obtained within a finite number of iterations. This means the optimal UE-AN associations $\mathbf{X}^*$ and solution to problem (24) can be obtained within finite iterations by solving

$$\Phi\left(\boldsymbol{\Lambda}^{(u)}, \mathbf{X}^{(t)}\right) = \Gamma\left(\mathbf{X}^{(t)}, \boldsymbol{\Lambda}^{(u)}, \boldsymbol{\nu}^{(u)}, \boldsymbol{\omega}^{(u)}, \boldsymbol{\gamma}^{(u)}, \boldsymbol{\mu}^{(u)}\right). \quad (29)$$

Therefore, the optimal UE-AN associations $\mathbf{X}^*$ and solution to problem P1 can also be obtained within a finite number of

iterations. As is analyzed above, the sequence $\{UB^{(t)}\}$ is nonincreasing, and after the set $\mathscr{C}$ is determined by constraints in (29), $UB^{(k)} = M^*$, where M^* denotes the optimal objective function value. The sequence $\{LB^{(t)}\}$ is nondecreasing and satisfies $LB^{(t)} = M^*$ after $\mathbf{X}^*$ is found through solving (29). Therefore, we can claim that $UB^{(t)} - LB^{(t)} = 0$ will guarantee that the ASRAO converges to the optimal solution to problem P1 within a finite number of iterations. $\qquad\square$

3.4. Accelerated ASRAO Algorithm.

Problem (26) is an MILP, which is generally calculated by branch and bound approach. Thus, the computation of problem (24) dominates the computation complexity of the whole BD process. In order to reduce the cost of solving problem (26), in this paper we propose to accelerate ASRAO algorithm by relaxing integer constraints in (26) into continuous constraints in intermediate iterations. Specifically, in tth iteration, the MP can be relaxed into a LP problem, as is in

$$\begin{aligned}
\max_{(\mathbf{X}_{\mathscr{J}}, \mathbf{X}_{\mathscr{K}}, \eta)} \quad & \eta, \\
\text{s.t.} \quad & \Gamma\left(\mathbf{X}, \Lambda^{(u)}, \boldsymbol{\nu}^{(u)}, \boldsymbol{\omega}^{(u)}, \boldsymbol{\gamma}^{(u)}, \boldsymbol{\mu}^{(u)}\right) \geq \eta, \\
& \qquad\qquad\qquad\qquad \forall u = 1, \ldots, t_1, \\
& \sum_{(i,l) \in Y_v} x_{il} - \sum_{(i,l) \in N_v} x_{il} \leq |Y_v| - 1, \\
& \qquad\qquad\qquad\qquad \forall v = 1, \ldots, t_2, \\
& \sum_{l \in \mathscr{I} \cup \mathscr{J}} x_{il} = 1, \\
& \sum_{i \in \mathscr{I}} x_{ik} \leq N_m, \\
& x_{il} \in [0, 1], \\
& i \in \mathscr{I}, \\
& k \in \mathscr{K}, \\
& l \in \mathscr{I} \cup \mathscr{J},
\end{aligned}$$

where x_{il} is a relaxed continuous variable ranged in $[0, 1]$. As problem (30) is a LP problem, it can be solved utilizing standard algorithms, simplex method for instance. Denote the solution derived from problem (30) as $(\widehat{\mathbf{X}}, \widehat{\eta})$, in which the entries in $\widehat{\mathbf{X}}$ are possibly not integers. Theorem 3 addresses the fact that relaxation of problem (26) will not exclude the optimal solution to P1.

Theorem 3. *Denote $(\widehat{\mathbf{X}}, \widehat{\eta})$ as an arbitrary feasible solution to the relaxed problem. The optimality cuts and feasibility cuts generated with $\widehat{\mathbf{X}}$ will not exclude the optimal solution $(\Lambda^*, \mathbf{X}^*)$ from problem P1.*

Proof. If problem (24) is bounded, denote $(\widehat{\Lambda}, \widehat{\nu})$ as the optimal solution to problem (24) with $\widehat{\mathbf{X}}$ and an optimality cut $\Gamma(\mathbf{X}, \widehat{\Lambda}_{\mathscr{J}}, \widehat{\nu}, \widehat{\omega}, \widehat{\gamma}, \widehat{\mu}) \geq \eta$ is generated. Otherwise, a feasibility

cut $\sum_{(i,l) \in Y_v} x_{il} - \sum_{(i,l) \in N_v} x_{il} \leq |Y_v| - 1$ is generated. To prove Theorem 3, we show that, in either cases, optimal solution $(\Lambda^*, \mathbf{X}^*)$ to problem P1 will not violate the newly introduce constraints. Hence, it will not be excluded from the feasible set by the generated cuts.

Denote the optimal solution to problem (24) as $(\Lambda^*, \boldsymbol{\nu}^*, \boldsymbol{\omega}^*, \boldsymbol{\gamma}^*, \boldsymbol{\mu}^*)$ with $\mathbf{X}^*$; the corresponding optimal value of the objective function is $M^* = \eta^* = \Phi(\Lambda^*, \mathbf{X}^*)$. In the case where problem (24) is bounded with $\mathbf{X}$ being $\widehat{\mathbf{X}}$, suppose $(\mathbf{X}^*, \eta^*)$ violates $\Gamma(\mathbf{X}, \widehat{\Lambda}, \widehat{\nu}, \widehat{\omega}, \widehat{\gamma}, \widehat{\mu}) \geq \eta$, that is $\Gamma(\mathbf{X}^*, \widehat{\Lambda}, \widehat{\nu}, \widehat{\omega}, \widehat{\gamma}, \widehat{\mu}) < \eta^*$, due to the completeness of the constraints in (26), it means a smaller $\eta' = \Gamma(\mathbf{X}^*, \widehat{\Lambda}, \widehat{\nu}, \widehat{\omega}, \widehat{\gamma}, \widehat{\mu})$ is found, which violates the fact that $M^* = \eta^*$ is the optimal solution to problem P1. Hence, the optimality cut will not be violated. In the case where the problem (24) is unbounded, suppose $\mathbf{X}^*$ violates the feasibility cut $\sum_{(i,l) \in Y_v} x_{il} - \sum_{(i,l) \in N_v} x_{il} \leq |Y_v| - 1$, which means $\mathbf{X}^*$ makes problem (24) unbound, and therefore, $\mathbf{X}^*$ will not be the optimal solution to P2. This conflicts the fact that $(\mathbf{X}^*, \eta^*)$ is the optimal solution to P1. So the feasibility cut will not be violated.

Theorem 3 means we can safely replace (26) by (30) in our ASRAO algorithm without concerning about losing the optimal solution of P1. However, it should be noted that $\widehat{\mathbf{X}}$ in the optimal solution to problem (30) can be used to calculate the lower bound of P1, but when problem (24) is solved with $\widehat{\mathbf{X}}$, the upper bound derived from the objective function value of problem (24) may not be a valid upper bound for P1, since entries in $\widehat{\mathbf{X}}$ are not integers and thereby infeasible to constraints in P1. One way to cope with the problem is to round $\widehat{\mathbf{X}}$ to the nearest matrix $\widetilde{\mathbf{X}}$ with binary entries. However, as the summation of elements in each row of $\mathbf{X}$ equals 1, a direct approximation would produce an all-zero matrix. Hence, in view of the special property of $\mathbf{X}$ that the entries in the matrix are binary variables and only one element in each row equals 1, we use a technique inspired by feasible pump method proposed in [21] to get $\widetilde{\mathbf{X}}$.

To select a 1 element in x_i, the ith row of $\widehat{\mathbf{X}}$, we divide a range $[0, 1]$ into $J + K$ slots, and each element in x_i successively occupies a slot, whose length equals to the value of element x_{il}. Then a randomizer uniformly generates a pointer valued within $[0, 1]$. This element with the slot where the pointer locates is selected, and all other elements are 0 elements. To control randomness, a parameter T is used to decide only $1/T$ rows are allowed to perform randomly selection and other rows just select their largest elements as 1 element.

Thereby, from $\widehat{\mathbf{X}}$ we get an approximate matrix $\widetilde{\mathbf{X}}$ which is possibly feasible to problem (26) and P1, but with very low probability $\widetilde{\mathbf{X}}_{\mathscr{K}}$ fails the constraint $\sum_{i \in \mathscr{I}} x_{ik} \leq N_m$, and therefore, $\widetilde{\mathbf{X}}_{\mathscr{K}}$ should be checked. And to avoid the fact that $\widetilde{\mathbf{X}}$ stays at a fixed point, novelty of the solution should be checked. If no feasible solution or solution stays unchanged, we revert back to the normal ASRAO algorithm, and the current solution calculated from problem (30) serves as an initial $\widetilde{\mathbf{X}}$ to the following normal ASRAO process.

The pseudocodes of accelerated ASRAO algorithm are as shown in Algorithm 2. $\qquad\square$

(1) **Initialize**: $UB^{(0)} = +\infty$, $LB^{(0)} = -\infty$, Let $t = 0$, $u = 0$, $v = 0$.
   PHASE-I
(2) **repeat**
(3)   Let $t = t + 1$ and solve (30) to obtain current optimal solution $(\widehat{\mathbf{X}}^{(t)}, \widehat{\eta}^{(t)})$, $UB^{(t)} = \widehat{\eta}^{(t)}$
(4)   **if** $\widehat{\mathbf{X}}^{(t)} = \widehat{\mathbf{X}}^{(t-1)}$ or feasible $\widetilde{\mathbf{X}}^{(t)}$ can not be found: $t = t - 1$, **break**
(5)   **else**: get approximative matrix $\widetilde{\mathbf{X}}^{(t)}$, and $\mathbf{X}^{(t)} = \widetilde{\mathbf{X}}^{(t)}$
(6)   **if** (24) with $\mathbf{X}^{(t)}$ is bounded: $u = u + 1$, and solve it with $\mathbf{X}^{(t)}$ to obtain KKT point $(\mathbf{\Lambda}^{(u)}, \boldsymbol{\nu}^{(u)}, \boldsymbol{\omega}^{(u)}, \boldsymbol{\gamma}^{(u)}, \boldsymbol{\mu}^{(u)})$
        $= (\mathbf{\Lambda}^{(t)}, \boldsymbol{\nu}^{(t)}, \boldsymbol{\omega}^{(u)}, \boldsymbol{\gamma}^{(u)}, \boldsymbol{\mu}^{(t)})$. The lower bound is set to $LB^{(t)} = \max\{\max_{1 \le s \le t}\{\Phi(\mathbf{\Lambda}^{(s)}, \mathbf{X}^{(s)})\}, LB^{(0)}\}$.
        Add $\Gamma(\mathbf{X}, \mathbf{\Lambda}^{(u)}, \boldsymbol{\nu}^{(u)}, \boldsymbol{\omega}^{(u)}, \boldsymbol{\gamma}^{(u)}, \boldsymbol{\mu}^{(u)}) \ge \eta$ to (26) and (30)
(7)   **else**: $v = v + 1$, $\mathbf{X}^{(v)} = \mathbf{X}^{(t)}$, $LB^{(t)} = LB^{(t-1)}$, and add $\sum_{(i,l) \in Y_v} x_{il} - \sum_{(i,l) \in N_v} x_{il} \le |Y_v| - 1$ to (26) and (30).
(8) **until** $UB^{(t)} - LB^{(t)} \le 0$
   PHASE-II
(9) ASRAO step (2) to (7).
(10) Return $(\mathbf{X}^*, \mathbf{\Lambda}^*) = (\mathbf{X}^{(t)}, \mathbf{\Lambda}^{(t)})$ as the optimal solution to problem P1.

ALGORITHM 2: The A-ASRAO algorithm.

4. Performance Evaluation

In this section, we first introduce network scenarios and parameter settings used in our simulations. After this, the performance of ASRAO is evaluated in scenarios with different AP and UE settings. The effectiveness of the accelerative technique is also discussed here. Then, we analyze effects the value of β may cause to our UTMAX model by evaluating the optimized network performances on metrics, such as system throughput and access fairness. In the last part of this section, our UTMAX model is evaluated under different AP densities to validate its effectiveness and compatibility.

4.1. Scenario and Parameter Settings. The simulations are carried out in a scenario consists of 2 fixed BSs and 6 or 12 APs. The distance between the two BSs is about 300 m. The number of UEs varies from 10 to 20 by a step size of 2. In order to reduce random errors, each point of the simulation results is averaged over 100 different network topologies among which the positions of APs and UEs differ but their amounts stay unchanged. According to [15], we simplify the propagation loss as $37 + 37.6 \log d$ (km) $+ 21 \log (2.6/2)$ dB in CN, and $34 + 37.6 \log d$ (km) $+ 21 \log (2.4/2)$ dB in WLAN. To model the shadowing effects in CN a log-normal random variable with zero mean and 8 db variation is introduce in CN, while in WLAN shadowing effects are neglected. Moreover, a received power maximization (PRMAX) user association scheme, which is commonly used in present commercial networks, serves as a performance benchmark. Another scheme maximally uses WLAN by getting AP to serve each possible user in its range. We denote it as APMAX. Our UTMAX optimization model with different β values for different system preferences, such as system throughput and user data rate fairness, is compared with PRMAX and APMAX to validate its performance. The main parameters and their values are as listed in Table 1, from [7, 15].

4.2. Algorithm Performance Analyses. In this subsection, we validate the convergence of ASRAO and A-ASRAO in simulations and analyze the effects the value of a randomness

TABLE 1: Simulation parameter.

Items (unit)	Values
Transmitting power of BSs (dBm)	46
Transmitting power of WiFi AP (dBm)	20
Distance between BSs (m)	300
LTE bandwidth (MHz)	10
WiFi bandwidth (MHz)	20
eNB antenna height from eNB (m)	40
AP antenna height from AP (m)	5
LTE operating frequency (MHz)	2600
WiFi operating frequency (MHz)	2400
Number of eNode BSs and APs	$2:6, 2:12$
Thermal noise (dBm/Hz)	-174
Shadowing for LTE (dB)	8
LTE cochannel interference coefficient α	1
Number of UEs	$10:2:20$
Min resource for UE in eNB λ_ε	0.01
Max resource for UE in AP λ_m	0.3
Fairness coefficient β	$-0.05, 0, 1, 5$

control parameter T produces on the convergence of the latter. The simulations are carried out in Matlab under the settings mentioned above, and our algorithms are implemented with YALMIP [22], an optimization toolbox, which is able to solve our UTMAX model directly, but at unsatisfying speed. We denote it by B&B, because an MINLP solver based on branch and bound is used. It works differently in each detailed iteration with (A-)ASRAO, so only the computation time spent by these approaches is compared.

Figure 2 illustrates the convergent processes of ASRAO and A-ASRAO in a UTMAX instance where 20 UEs exist. In essence, ASRAO performs fixed path searching, whose interim solutions are decided by their preceding solutions. ASRAO is sure to converge, but it lacks the ability to try better searching paths. While A-ASRAO adopts a random
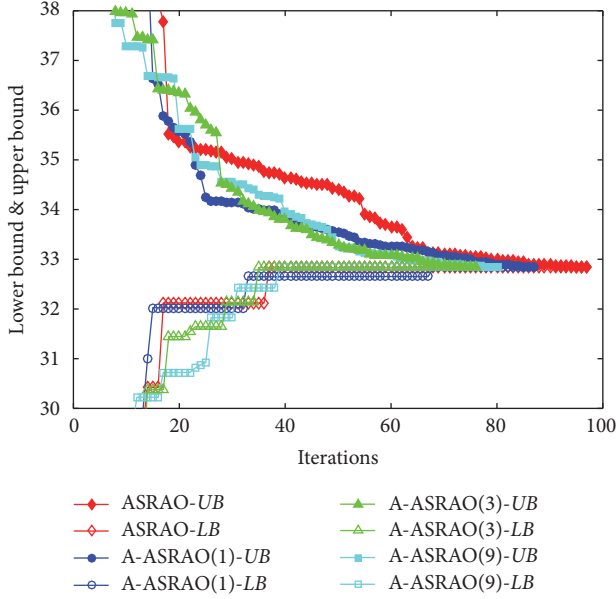
Figure 2: The converge process of ASRAO and A-ASRAO with $T = (1, 3, 9)$.

Figure 3: The number of iterations versus the number of UEs.

searching strategy, with wider searching directions, it shows more steady descending upper bound. From Figure 2, it is obvious that A-ASRAO converges much faster in the later part of the iterations than ASRAO as is depicted by lines with hollow marks. As these upper bounds reach the same optimal objective function value and all lower bounds get to optima much earlier than upper bounds, A-ASRAO will converge much faster than ASRAO.

In addition, A randomness control parameter T is set in A-ASRAO to control its behavior, and A-ASRAO with T is denoted by A-ASRAO(T). With different values of T, A-ASRAO has different converging rates. This is probably caused by random generation of approximation matrix $\widetilde{\mathbf{X}}$. Smaller T means more rows in $\widetilde{\mathbf{X}}$ perform random 1-element selection, while large T will make less rows perform random selection. Extremely small or large values of T will make too many or few rows perform random selection, but neither of them is good for finding ideal initial solution for Phase II in A-ASRAO. So A-ASRAO(3) performs better than A-ASRAO(1) and A-ASRAO(9).

In Figures 3 and 4, respectively, the number of iterations performed and computation time spent by ASRAO and A-ASRAO are plotted against the number of UEs ranging from 10 to 20. Each point is obtained by averaging over 100 different network topologies. It shows that ASRAO needs a little fewer iterations than A-ASRAO when UE number is less than 14. However, when UE number is larger, A-ASRAO reduces the number of iterations by up to 15% compared with ASRAO in average.

The reason for the different convergent properties lies in the scale of a UTMAX instance. A-ASRAO works in two phases and in Phase II A-ASRAO works the same as ASRAO, which implies that A-ASRAO in Phase I works actually to obtain an initial solution to Phase II in A-ASRAO. That
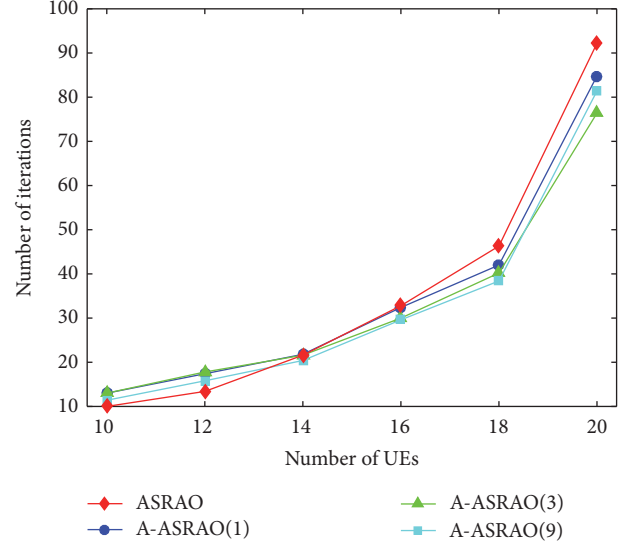
is to say, if a better solution, which produces smaller gap between the upper bound and lower bound, found by the end of Phase I in A-ASRAO than that by ASRAO within the same iterations, A-ASRAO will have much larger probability to converge within fewer iterations than ASRAO. Therefore, the dominant factor relating to the different convergent rates of ASRAO and A-ASRAO is whether a better solution would be found by the end of Phase I in A-ASRAO.

When a UTMAX instance is small-scale problem, its optimal solution can be derived within a few iterations even through exhaustive search, not alone ASRAO. While A-ASRAO tries a more steady but random searching, the finding of optimal solution may thus be postponed. So it hardly performs better than ASRAO. When a UTMAX instance is of large scale, a large solution space means a bolder searching practice may produce better searching path. So A-ASRAO(3) performs better than ASRAO and A-ASRAO(9). Nevertheless, too strong randomness goes against the finding of optimal solution, like in A-ASRAO(1) where all rows in $\widetilde{\mathbf{X}}$ perform random 1-element selection. We speculate that is because a too wild searching will waste many iterations wandering around rather than heading towards a converging direction. That is why A-ASRAO(3) outperforms A-ASRAO(1). Moreover, in Figure 4 it shows that A-ASRAO spends less time computing optimal solution than ASRAO, and the performance gap is even larger than iteration numbers. This benefits from relaxing the MILP master problem into LP in Phase I of A-ASRAO, and LP can be solved much faster than MILP. Figure 4 also shows that ASRAO computes UTMAX much quicker than a common B&B method. Above results prove the effectiveness of (A-)ASRAO in solving UTMAX, and in most cases A-ASRAO(3) performs better than others.

4.3. Effect of β to UTMAX. As is mentioned in Section 2, β indicates UTMAX's preference on throughput fairness among UEs served in a BS by setting effects on resource
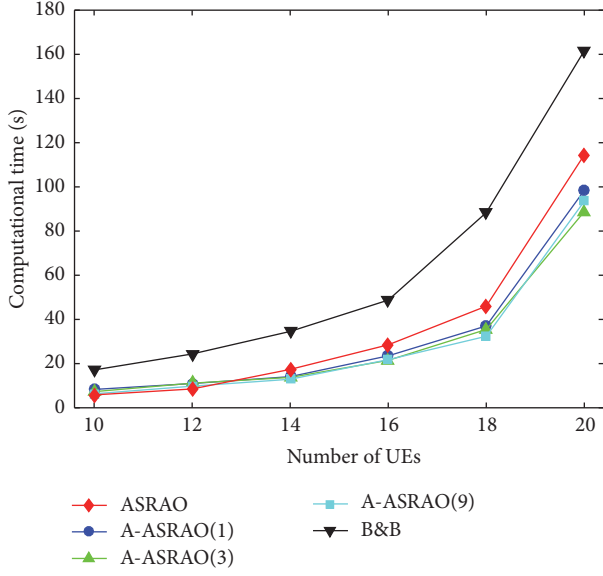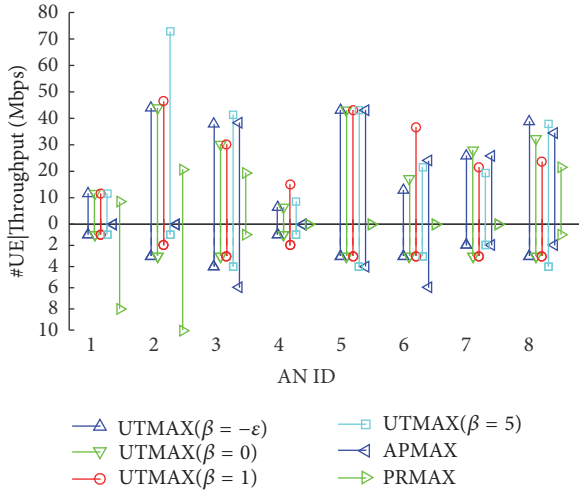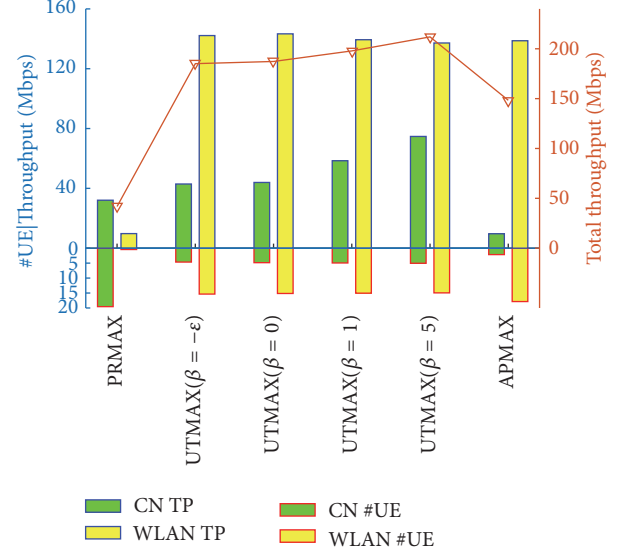
FIGURE 4: Computational time versus the number of UEs.

FIGURE 6: Performance at network granularity of PRMAX, APMAX, and UTMAX(β).

FIGURE 5: Performance at AN granularity of PRMAX, APMAX, and UTMAX(β).

allocation. As UE's attainable data rate is changed, user associations are inevitably affected, which is reflected in Figure 5. And as a consequence, the performance of the entire interworking network is influenced. To analyze these effects, the performance of optimized network is evaluated on throughput produced and the number of UEs served in WiFi tier and cellular tier in Figure 6 and throughput gains of UEs in Figure 7 through APMAX and UTMAX with β in $\{-0.05, 0, 1, 5\}$. PRMAX is a default user association scheme and adopts identical resource allocation in each BS and AP.

Figure 5 plots the throughput and number of UEs served in each BS and AP with different approaches marked by different colors and symbols. It shows that, from a AN perspective, both APMAX and UTMAX do well in improving AP's utilization, and APMAX performs even better, as it tends to maximize AP's usage. The improvements of WLAN's

offloading performance by different approaches are much more obvious in Figure 6.

Figure 6 plot the throughput and number of UEs served in cellular tier and WiFi tier. APMAX gets most UEs served in WLAN, while, as resources are fully occupied, each UE in WLAN will get less resources and the throughput of WLAN will not be improved greatly compared with UTMAX. However, due to lack of careful UE selection from CN to WLAN in APMAX, UE amount decreases greatly in CN and the left UEs are probably with poor wireless channel condition or too far from a BS, which causes the poor CN throughput performance in APMAX. As for UTMAX, with different β value, it performs differently. Larger β value will get BS to allocate more resources to UEs with higher unit band data rate, which probably makes the contribution to the throughput growing in CN. However, as UEs with poor channel condition in CN get much less resources, they turn to WLAN to compensate the throughput loss. That would cause a minor UE amount increase in WLAN, while, as AP's resources are equally shared, new coming UEs with poor channel condition share resources with indigenous WLAN UEs, causing the decrease in WLAN's throughput. Nevertheless, the whole system throughput shows a continuous growth with the increase of β. It seems that larger β performs better on total throughput and throughput fairness between WLAN and CN, whereas it is on the expense of throughput fairness among UEs.

Figure 7 depicts UE throughput gain versus probability $\mathbb{P}(r < \alpha)$ for various approaches versus PRMAX, where $\mathbb{P}(r < \alpha)$ represents the ratio of UE whose throughput is less than α in all UE. The UEs' throughput gains are quite large at 10% ratio point by all approaches, among which UTMAX($-\varepsilon$) brings the largest throughput gain. With the growing of β, UE throughput gain decreases. However, UTMAX always has higher throughput gain over APMAX. With the growth of probability $\mathbb{P}(r < \alpha)$ UE throughput gain decreases,
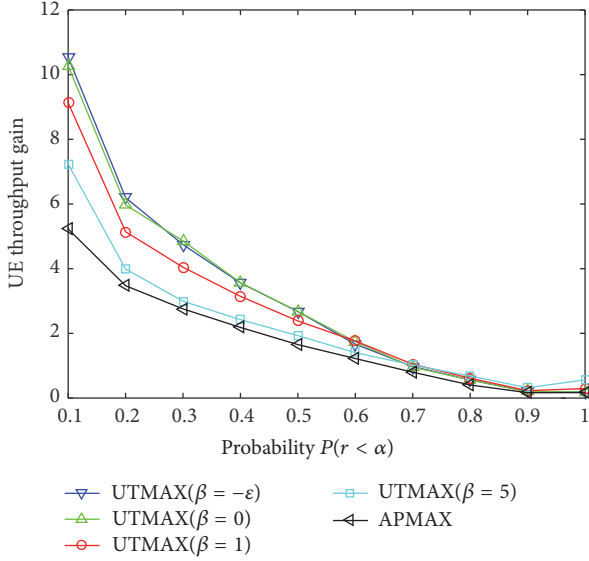
Figure 7: Performance at UE granularity of PRMAX, APMAX, and UTMAX(β).

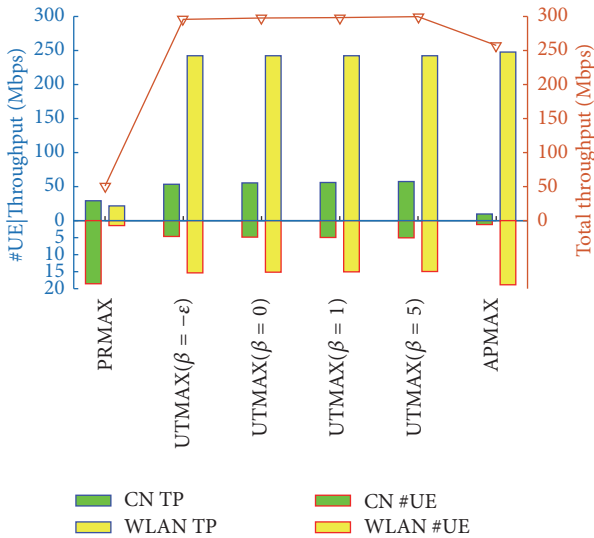Figure 8: Performance at network granularity of PRMAX, APMAX, and UTMAX(β).

Figure 9: Performance at UE granularity of PRMAX, APMAX, and UTMAX(β).

because PRMAX also has relatively larger UE throughput at larger ratio points. At the meanwhile, UTMAX with larger β value has larger UE throughput gain, as larger β makes UTMAX allocate more resources to UEs with better channel conditions.

4.4. Effects of AP Density to UTMAX. The performances of UTMAX with different values of β are evaluated on larger AP density, to validate whether the same conclusion holds with different AP settings. The results are in Figures 8 and 9.

In Figure 8, compared with Figure 6 the most obvious changes are the WLAN tier throughput increase for PRMAX and minimization of performance gap between UTMAXs with each of two different values of β. Due to the increase of
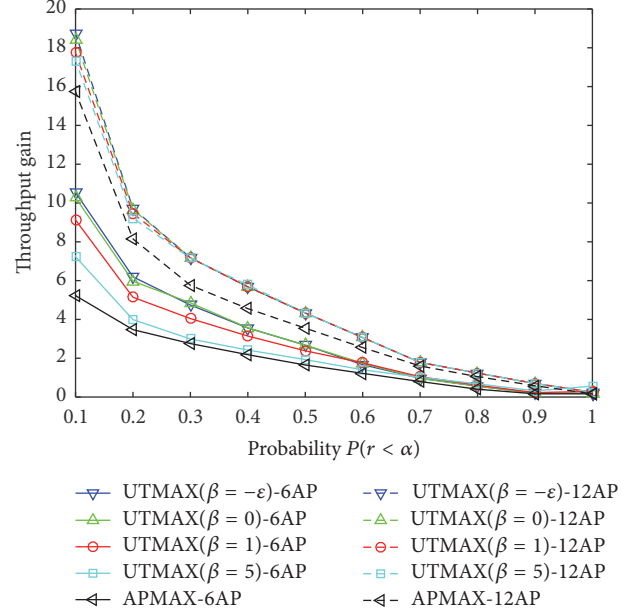
AP density, more UEs access WLAN and the average distance from a UE to an AP is shorten. This is a direct cause to WLAN tier throughput increase for PRMAX and APMAX. As for the latter difference, it is possibly because of much fewer UEs in each BS or AP, and β probably has weaker effect on resource allocation among fewer UEs in a serving node. Figure 9 is plotted by adding curves of UE throughput gain where 12 APs exist. The additional UE throughput gain is largely contributed by newly added APs. A similar performance relation is shown between UTMAX with larger AP density, while the performance gap is narrowed and the curves meet and overlap earlier than that with smaller AP density.

From these results, the effect of β to UTMAX is further checked that a smaller β makes UTMAX perform better on improving throughput of the part of UEs with poor channel conditions, while a larger β does better on improving overall system throughput.

5. Conclusion

To solve a key research case in multi-RAT SON about alleviating the biased load distribution in cellular/WLAN interworking network, we propose to optimize access load from users satisfaction perspective, and a UTMAX optimization model with sum of logarithmic utilities is derived. This is an MINLP and seems hard to solve intuitively. In order to solve it optimally and efficiently, we devise an ASRAO algorithm on the basis of general Benders Decomposition and prove its convergence in finite iterations to optimal. To reduce the computation complexity in ASRAO, especially when dealing with large-scale problem instance, we propose to accelerate the ASRAO algorithm with a relaxation and approximation technique inspired by feasible pump. Simulation results validate the convergence and effectiveness of both ASRAO

and A-ASRAO and prove the effects of UTMAX optimization model on improving throughput fairness among users or the system throughput, and in either case, the utilization of WLAN is improved obviously especially on the number of users and throughput in it.

As shown in simulation results, when dealing with large-scale UTMAX instance, A-ASRAO has distinct advantage over ASRAO on convergent speed, which implies that A-ASRAO is a proper algorithm to solve UTMAX. The resource allocation among users in BS is affected by β, whereas simulation results imply it, and also functions as an indicator for preference on throughput fairness among users or total system throughput, which is a tradeoff or preference from viewpoints of network operators. However, due to computation complexity limitation, it is hard to perform extensive simulations on much larger UTMAX instances. So our future work includes a heuristic approach to produce solution close to that of ASRAO but with much lower computational complexity, and therefore effects of UTMAX can be further checked with larger UE density.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

References

[1] R. Litjens, "D6.6 final report on a unified self-management system for heterogeneous radio access networks," Tech. Rep. INFSO-ICT-316384, SEMAFOUR, 2015.

[2] D. Laselva, "D4.1 SON functions for multi-layer LTE and multi-RAT networks (first results)," Tech. Rep. INFSO-ICT-316384, SEMAFOUR, 2013.

[3] F. Rebecchi, M. Dias de Amorim, V. Conan, A. Passarella, R. Bruno, and M. Conti, "Data offloading techniques in cellular networks: a survey," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 580–603, 2015.

[4] A. Balasubramanian, R. Mahajan, and A. Venkataramani, "Augmenting mobile 3G using WiFi," in *Proceedings of the 8th Annual International Conference on Mobile Systems, Applications and Services (MobiSys '10)*, pp. 209–222, ACM, San Francisco, Calif, USA, June 2010.

[5] M. D. Nisar, V. Pauli, and E. Seidel, "Multi-RAT traffic steering—why, when, and how could it be beneficial?" Whitepaper, 2011.

[6] D. H. Hagos, *The performance of WiFi offload in LTE networks [M.S. thesis]*, Lulea University of Technology, Lulea, Sweden, 2012.

[7] I. Balan, D. Laselva, S. Redana, and A. Lobinger, "RSRP-based LTE-WLAN traffic steering," in *Proceedings of the 81st IEEE Vehicular Technology Conference (VTC '15)*, pp. 1–5, May 2015.

[8] S. Shin, D. Han, H. Cho, and J.-M. Chung, "Improved association and disassociation scheme for enhanced WLAN handover and VHO," *Mobile Information Systems*, vol. 2016, Article ID 4868479, 6 pages, 2016.

[9] 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, January 2011.

[10] K. Piamrat, A. Ksentini, C. Viho, and J.-M. Bonnin, "QoE-aware vertical handover in wireless heterogeneous networks," in *Proceedings of the 7th International Wireless Communications and Mobile Computing Conference (IWCMC '11)*, pp. 95–100, Istanbul, Turkey, July 2011.

[11] A. I. Aziz, S. Rizvi, and N. M. Saad, "Fuzzy logic based vertical handover algorithm between LTE and WLAN," in *Proceedings of the International Conference on Intelligent and Advanced Systems (ICIAS '10)*, pp. 1–4, IEEE, Kuala Lumpur, Malaysia, June 2010.

[12] K. Premkumar and A. Kumar, "Optimum association of mobile wireless devices with a WLAN-3G access network," in *Proceedings of the IEEE International Conference on Communications (ICC '06)*, vol. 5, pp. 2002–2008, Istanbul, Turkey, July 2006.

[13] J. Chen, L. P. Qian, and Y. J. Zhang, "On optimization of joint base station association and power control via Benders' decomposition," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '09)*, pp. 1–6, IEEE, Honolulu, Hawaii, USA, December 2009.

[14] S. Kim, S. Choi, and B. G. Lee, "A joint algorithm for base station operation and user association in heterogeneous networks," *IEEE Communications Letters*, vol. 17, no. 8, pp. 1552–1555, 2013.

[15] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2706–2716, 2013.

[16] S. Singh, S. Yeh, N. Himayat, and S. Talwar, "Optimal traffic aggregation in multi-RAT heterogeneous wireless networks," in *Proceedings of the IEEE International Conference on Communications Workshops (ICC '16)*, pp. 626–631, Kuala Lumpur, Malaysia, May 2016.

[17] T. GPP, "23.234 v6. 2.0, 3gpp system to wireless local area network (wlan) interworking," System description (Release 6), 2004.

[18] "IEEE Standard for Local and metropolitan area networks—Media Independent Handover Services," IEEE Std 802.21-2008, 2009.

[19] J. F. Benders, "Partitioning procedures for solving mixed-variables programming problems," *Numerische Mathematik*, vol. 4, pp. 238–252, 1962.

[20] A. M. Geoffrion, "Generalized Benders decomposition," *Journal of Optimization Theory and Applications*, vol. 10, no. 4, pp. 237–260, 1972.

[21] M. Fischetti, F. Glover, and A. Lodi, "The feasibility pump," *Mathematical Programming*, vol. 104, no. 1, pp. 91–104, 2005.

[22] J. Löfberg, "YALMIP: a toolbox for modeling and optimization in MATLAB," in *Proceedings of the IEEE International Symposium on Computer Aided Control System Design*, pp. 284–289, September 2004.

Research Article

A TDoA Localization Scheme for Underwater Sensor Networks with Use of Multilinear Chirp Signals

En Cheng, Xizhou Lin, Shengli Chen, and Fei Yuan

Key Laboratory of Underwater Acoustic Communication and Marine Information Technology (Xiamen University),
Ministry of Education, Xiamen 361005, China

Correspondence should be addressed to Fei Yuan; yuanfei@xmu.edu.cn

Due to the multipath, Doppler, and other effects, the node location signals have high probability of access collision in the underwater acoustic sensor networks (UW-ASNs), and therefore, it causes the signal lost and the access block; therefore, it constrains the networks performance. In this paper, we take the multilinear chirp (MLC) signals as the location signal to improve the anticollision ability. In order to increase the detection efficiency of MLC, we propose a fast efficient detection method called mixing change rate-fractional Fourier transform (MCR-FrFT). This method transforms the combined rates of MLC into symmetry triangle rates and then separates the multiuser signals based on the transformed rates by using FrFT. Theoretical derivation and simulation results show that the proposed method can detect the locations signals, estimate the time difference of arrival (TDoA), reduce the multiple access interference, and improve the location performance.

1. Introduction

During the last couple of years, we could observe a growing interest in underwater acoustic sensor networks (UW-ASNs). One important reason is that the UW-ASNs improve ocean exploration ability and fulfill the needs of a multitude of underwater applications, including oceanographic data collection, warning systems for natural disasters, ecological applications, military underwater surveillance, assisted navigation, and industrial applications. However, the location of the sensors needs to be determined because sensed data can only be interpreted meaningfully when referenced to the location of the sensor. Due to the absorption of electromagnetic wave, the well-known Global Position System (GPS) receivers, which may be used in terrestrial systems to accurately estimate the geographical locations of sensor nodes, could not work properly underwater [1]. Therefore, localization for UW-ASNs has been one of the major research topics since UW-ASNs started to draw the attention of the networking community in the early 2000s [2].

Sensors' self-localization is the basic and key requirement for the self-organization of UW-ASNs. It can be achieved by leveraging the low speed of sound in water to accurately determine the internode distance [3]. Localization of underwater equipment has also been an essential part of the traditional oceanographic systems where it has been established by one of two techniques: short base line (SBL) [4] or long base line (LBL) [5]. Time difference of arrival (TDoA) method is typically used in LBL systems. In TDoA-based localization, the difference in the measured time of arrivals of signals, received from a pair of reference nodes, is translated to the difference in range estimates with those reference nodes and gives rise to a hyperbola for the unknown position of the node (target node). A unique estimation of the target node position can be obtained by intersecting three such hyperbolas. However, this technique requires that reference nodes transmit at near concurrent time because of the water currents (the motion of the target node).

Traditional LBL systems are designed for small deployments consisting of a few reference nodes (here called beacons), whose acoustic transmissions can be received by all submersibles within the deployment. LBL provides sufficient localization accuracy for such deployments. This is because signals from few beacons can be scheduled to occur within a

short time window. Since all the target nodes are within the communication range of the beacons, these signals also arrive at near concurrent times at each target node. Therefore, the TDoA is suitable. However, as the spatial extent and the size of the system are scaled up, the number of location signals increases too. Therefore, to avoid location signals' collision is also urgent, within the short time window.

2. Related Works

Acoustic-based localization was studied because of its application in underwater sensor networks localization. In [6], a TDoA-based silent positioning algorithm termed as UPS for UW-ASNs was proposed. In [7], the localization problem in sparse 3D underwater sensor networks was studied based on TDoA. In [8], a TDoA for multiple acoustic sources in reverberant environments was proposed, and the ambiguities in the TDoA estimation caused by multipath propagation and multiple sources were resolved by exploiting two TDoA constraints: the raster condition and the zero cyclic sum condition. However, these references do not consider the collisions of location signals in the target sensor.

In [9], a different system where transmissions from multiple reference nodes have to be sufficiently lagged over a time epoch to avoid collisions was presented, but it needs a lot of time to localize the target sensor. In [10], the authors introduced a method for designing localization signal based on Code Division Multiple Access (CDMA), and the modulation mode is ASK; but as described in [11] about the physical layer of underwater channel, ASK modulation is not suitable for underwater channel because of high attenuation. What is more, the location signals are transmitted that are sufficiently lagged, which is the same as [9]. In [8], to avoid collisions and meet the signal concurrency requirement of traditional LBL systems simultaneously, a TDoA location scheme for the orthogonal frequency division multiplexing (OFDM) was presented; this outperforms the location schemes with traditional TDoA estimations. Unfortunately, the OFDM system is very sensitive to Doppler, which is one of the main characters of underwater channel. In [11, 12], a time-varying multichirp rate modulation for multiple access systems was used to avoid collision in terrestrial communication. Chirp signal is widely used in underwater acoustic communication [13], because it is robust to channel noise and resistant to multipath fading [14] and has low Doppler sensitivity [13].

Some different methods are used to estimate the parameters of chirp signal. In [12], the method of time-delay frequency mixing (rate reduction) to convert the chirp signal with different chirp rate into MFSK signal is proposed, but it is not suited for combined chirp signal because of the change of chirp rate. In [11, 12], the detection technique of matched filter receiver (correlation method) was mentioned, but it is difficult and also needs amplitude modulation which is not suitable for underwater channel. The fractional Fourier transform (FrFT) presents best localization performance in a certain FrFT domain, which is useful for the detection and estimation of multicomponent linear frequency modulation (LFM) signals [15] and some improved algorithms based on FrFT are also proposed, such as EEMD-FrFT [16] and
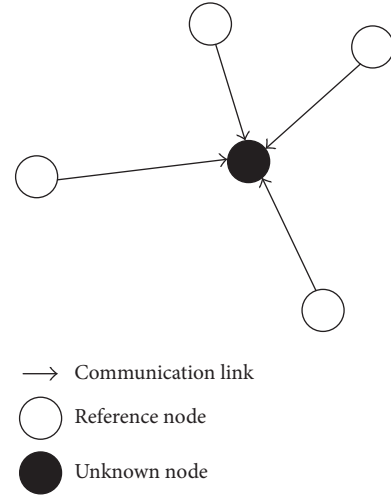

FIGURE 1: Location system architecture using combined chirp signals.

STFT [17]; they overcome some disadvantages such as high computation cost for combined chirp signals. In this paper, a method called mixing change rate-FrFT (MCR-FrFT) is proposed to deal with the drawback.

In this paper, a TDoA location scheme using the multilinear chirp modulation signals (called MLC signals) is proposed for UW-ASNs. The rest of the paper is organized as follows: in Section 3, the system model and location signals are described. In Section 4, we analyze the system with the underwater environment in TDoA estimation. In Section 5, we present our simulation results on our passive localization scheme for the multilinear chirp signals. Section 6 concludes this paper.

3. System Model

The system architecture of the scheme is depicted in Figure 1. The white nodes are reference nodes whose locations are known, while the black node is target node which needs to be located. Arrows indicate the communication links. The system is designed to work with only one-way reference node transmissions. The reference nodes and target node are fully synchronized with each other and the target node is within the communication range of at least four reference nodes (3D network) or three reference nodes (2D network). Since the communication range and distance between reference nodes are all hundred meters' scale, the location signals get high possibility to arrive at the same time. Therefore, avoiding collision is a key problem to node localization. We introduce the chosen location signal and the processing of the signal in the following subsections.

3.1. Location Signal-Multilinear Chirp (MLC) Signal. Due to the match of MLC and underwater channel, the MLC is chosen as the location signal, and its time-frequency characteristic of location signals-MLC is shown as Figure 2, where T is the duration time of location signal. Nodes are
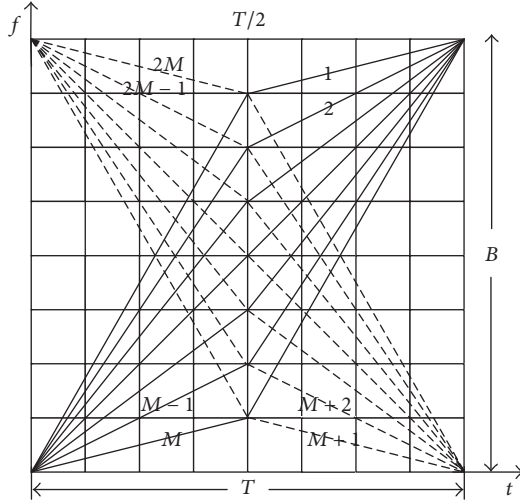
FIGURE 2: The time-frequency characteristic of location signals.

denoted by $1, 2, \ldots, M, M+1, \ldots, 2M-1, 2M$, where the first M nodes are with positive combined slopes and the second M nodes are with negative combined slopes. Each signal is composed of two segments with two different slopes. The signal for mth node is expressed as

$$S_m(t) = S_{mf}(t) + S_{mb}(t). \tag{1}$$

Subscripts f and b indicate the first half and the second half of duration time T, respectively. And

$$S_{mf}(t) = A \exp\left(j\pi\mu_{mf}t^2\right) \cdot \exp\left(j2\pi f_c t\right), \tag{2}$$

when $0 \le t < T/2$, and

$$S_{mb}(t) = A \exp\left(\pi\mu_{mb}\left(t - \frac{T}{2}\right)^2\right)$$
$$\cdot \exp\left(j2\pi\left(f_c + \frac{\mu_{mf}}{2}T\right)\left(t - \frac{T}{2}\right)\right), \tag{3}$$

when $T/2 \le t < T$,

where f_c is the carrier frequency, m is the node number, $m = 1, 2, \ldots, M, \ldots, 2M$, A is the amplitude of signal, and $2M$ is the total number of nodes in UW-ASNs. μ_{mf} is mth node's slope within the first half of signal duration and μ_{mb} is the slope of mth node within the second half of signal duration. The slopes of general combined chirp signal can be expressed as

$$\mu_{mf} = 2\frac{(M+1-m)B}{(M+1)T},$$

$$\mu_{mb} = 2\frac{mB}{(M+1)T}, \tag{4}$$

$$m \in [1, M],$$

$$\mu_{mf} = 2\frac{(m - 2M - 1)B}{(M+1)T},$$

$$\mu_{mb} = 2\frac{(m - M)B}{(M+1)T}, \tag{5}$$

$$m \in (M, 2M].$$

As shown in Figure 2, each location signal has the same bandwidth B and duration time T; thus the system produces equal time-bandwidth product $D = BT$ for all nodes. Equal D makes the nodes in the UW-ASNs have the equal position, which is important for mobile nodes.

3.2. Detection Block Diagram and Principle. The localization signal detection problem could be described as deciding which one of the $2M$ possible signals was transmitted, given the received signal during the interval time of $(0, T)$. The method of parameter estimation proposed in the paper is called MCR-FrFT, whose block diagram is shown as Figure 3(a). The received signal is divided into two branches. Each branch has its own local signals $x\text{local}u(t)$ and $x\text{local}d(t)$, respectively, $x\text{local}u(t)$ for the up branch $(m \le M)$ and $x\text{local}d(t)$ for the down branch $(M < m \le 2M)$.

3.2.1. Mixing Change Rate (MCR). The purpose of MCR is to change the chirp rates of MLC signals into a new set of rates. It includes two parts, multiplier and low pass filter. The block diagram of MCR is shown in Figure 3(b):

$$x\text{local}(t)$$

$$= \begin{cases} x\text{local}u(t) = \exp\left(j\pi\frac{B}{T}t^2\right) \cdot \exp\left(j2\pi f_c t\right), \\ x\text{local}d(t) = \exp\left(-j\pi\frac{B}{T}t^2\right) \cdot \exp\left(j2\pi f_c t\right), \end{cases} \tag{6}$$

$$r_0(t) = r_f(t) * h_{\text{LPF}}(t),$$

$$r_f(t) = r(t) \cdot x\text{local}(t),$$

where $h_{\text{LPF}}(t)$ is the impulse response function of low pass filter (LPF). Assuming that mth node's signal is received and $m < M$, we get the received signal as (7) by substituting (2)–(4) into (1). $n(t)$ is the underwater acoustic channel noise. After the MCR, the high frequency terms have been filtered out and the received signal becomes as (8) and $w(t) = (n(t) \cdot x\text{local}(t)) * h_{\text{LPF}}(t)$.

The frequency and time relationship of received signals after multiplier (r_f) are shown in Figure 4. From the figure, mth location signal, composed of two different positive slopes, is changed into two branches of new multilinear chirp signals. The up branch $(m \le M)$ has chirp rates of positive and negative relationship, while the down branch $(M < m \le 2M)$ is composed of another chirp rate. However, the received signal only has the up branch after being filtered in the LPF of MCR:
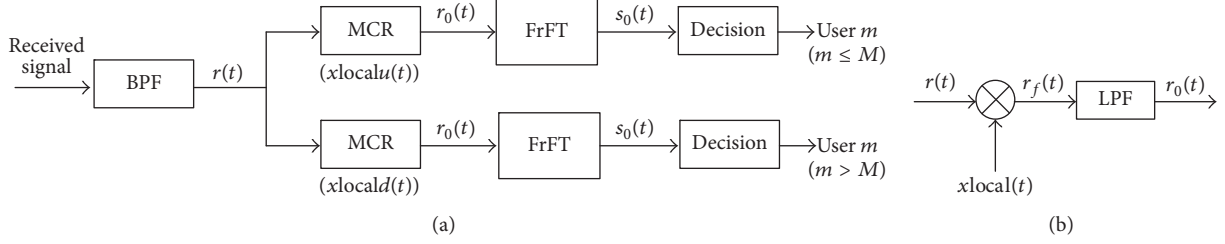
Figure 3: System block diagram. (a) The detection block diagram. (b) The block diagram of MCR.

$$r(t) = S_m(t) + n(t) = A \exp\left(j\pi\mu_{mf}t^2\right) \cdot \exp\left(j2\pi f_c t\right) \cdot \text{rect}\left(\frac{4t-T}{2T}\right) + A \exp\left(j\pi\mu_{mb}\left(t-\frac{T}{2}\right)^2\right) \cdot \exp\left(j2\pi\left(f_c + \frac{\mu_{mf}}{2}T\right)\left(t-\frac{T}{2}\right)\right) \cdot \text{rect}\left(\frac{4t-3T}{2T}\right) \tag{7}$$

$$+ n(t),$$

$$r_0(t) = w(t) + \frac{A^2}{2}$$

$$\cdot \begin{cases} \cos\left(\pi\left(\mu_{mf} - \frac{B}{T}\right)t^2\right) \cdot \text{rect}\left(\frac{4t-T}{2T}\right) + \cos\left[2\pi\left(\frac{M+1-m}{M+1}B - \frac{B}{2}\right)\left(t-\frac{T}{2}\right) + \pi\left(\mu_{mb} - \frac{B}{T}\right)\left(t-\frac{T}{2}\right)^2\right] \cdot \text{rect}\left(\frac{4t-3T}{2T}\right), & m \leq M, \\ \cos\left(\pi\left(\mu_{mf} + \frac{B}{T}\right)t^2\right) \cdot \text{rect}\left(\frac{4t-T}{2T}\right) + \cos\left[2\pi\left(\frac{M+1-m}{M+1}B + \frac{1}{2}B\right) \cdot \left(t-\frac{T}{2}\right) + \pi\left(\mu_{mb} + \frac{B}{T}\right)\left(t-\frac{T}{2}\right)^2\right] \cdot \text{rect}\left(\frac{4t-3T}{2T}\right), & M < m \leq 2M. \end{cases} \tag{8}$$

3.2.2. Fractional Fourier Transform (FrFT).

FrFT is a generalized Fourier transformation form and it can be regarded as the Fourier transform to pth order, where p needs not be an integer; thus it can transform a function to any intermediate domain between time and frequency [18]. For signal $x(t)$, the FrFT has the following form:

$$X_p(u) = \int_{-\infty}^{+\infty} K_p(t,u) x(t) dt, \tag{9}$$

and the transform kernel is

$$K_p(t,u) = \begin{cases} \sqrt{\frac{1-j\cot(\alpha)}{2\pi}} \exp\left[j\left(\frac{1}{2}t^2\cot(\alpha) + \frac{1}{2}u^2\cot(\alpha) - tu\csc(\alpha)\right)\right], & \alpha \neq n\pi, \\ \delta(t-u), & \alpha = 2n\pi, \\ \delta(t+u), & \alpha = (2n \pm 1), \end{cases} \tag{10}$$

where p is the transform order, α is the angle of rotation, and $\alpha = p\pi/2$. When $p = 1$, FrFT will degenerate into Fourier transformation, and when $p = 0$, FrFT is just the original signal.

Simple component LFM signal with noise has the following form:

$$r(t) = A_0 \cos\left(2\pi f_c t + \pi u t^2\right) + n(t), \tag{11}$$

where $0 \leq t \leq T$ and $n(t)$ is white noise. The FrFT of $r(t)$ is

$$X_p(u) = \int_{-\infty}^{+\infty} r(t) K_p(t,u) dt$$

$$= A_0 \sqrt{\frac{1-j\cot\alpha}{2\pi}} e^{(j(1/2)u^2\cot\alpha)}$$

$$\cdot \int_0^T e^{(2j\pi((1/2)t^2(u+\cot\alpha/2\pi)+t(f_c-u\csc\alpha/2\pi)))} dt$$

$$+ \int_0^T n(t) K_p(t,u) dt. \tag{12}$$

In (12), the LFM signal is an impulse function only in the appropriate fractional Fourier domain. The amplitude of energy aggregation of LFM signal will exhibit obvious peak by doing an appropriate p-order FrFT, while the noise, which cannot appear in energy aggregation at any fractional Fourier domain, is distributed on the whole time-frequency plane evenly.

Figure 4 shows that the set of chirp rates is changed from two different slopes, which needs two different p values to complete its parameter estimation, to a new set of chirp rates
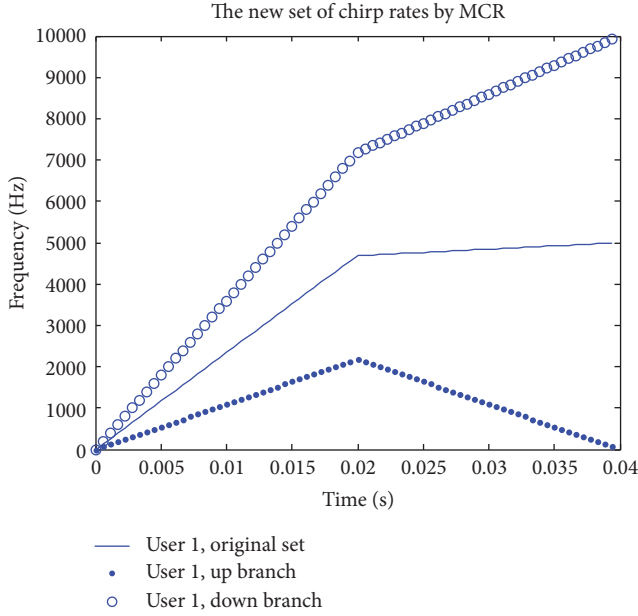
The new set of chirp rates by MCR

FIGURE 4: The time-frequency characteristic after MCR.

- —— User 1, original set
- • User 1, up branch
- ○ User 1, down branch

that the new two slopes are with same absolute value but one positive and one negative by using the MCR. At this condition, only one value needs to complete the parameter estimation, that is, one FrFT for a single branch, while two FrFT are needed for the scheme of FrFT (without MCR). Since FrFT cost more computation than the MCR module, our MCR-FrFT saves computations of system. In practical application, the discrete fractional Fourier transform (DFrFT) is usually used.

4. TDoA Estimation

There are two steps of MCR-FrFT at the receiving end. In the first step, MCR takes effect to convert the chirp rates of location signal into a new set of chirp rates. Then, FrFT estimates the parameters. There are two parameters considered: chirp rate, which represents the reference node, and initial frequency, which can be used for estimation of the processing delay.

4.1. Collision Avoidance. One of the advantages of the paper is to avoid collision. Taking an example with three reference nodes $m, l, n \in [1, M]$,

$$r(t) = S_m(t) + S_l(t) + S_n(t) + n(t). \tag{13}$$

According to (4) and (8), there are different sets of chirp rates included in the MCR results with different m, n, l group:

$$
\begin{aligned}
r_0(t) = {} & \frac{A^2}{2}\left[\cos\left(\pi\frac{(M+1-2m)B}{(M+1)T}t^2\right)\right.\\
& + \cos\left(\pi\frac{(M+1-2l)B}{(M+1)T}t^2\right)\\
& \left.+ \cos\left(\pi\frac{(M+1-2n)B}{(M+1)T}t^2\right)\right]\cdot\text{rect}\left(\frac{4t-T}{2T}\right)\\
& + \frac{A^2}{2}\left[\cos\left(2\pi\left(\frac{M+1-m}{M+1}B-\frac{B}{2}\right)\left(t-\frac{T}{2}\right)\right.\right.\\
& \left.- \pi\frac{(M+1-2m)B}{(M+1)T}\left(t-\frac{T}{2}\right)^2\right)\\
& + \cos\left(2\pi\left(\frac{M+1-l}{M+1}B-\frac{B}{2}\right)\left(t-\frac{T}{2}\right)\right.\\
& \left.- \pi\frac{(M+1-2l)B}{(M+1)T}\left(t-\frac{T}{2}\right)^2\right)\\
& + \cos\left(2\pi\left(\frac{M+1-n}{M+1}B-\frac{B}{2}\right)\left(t-\frac{T}{2}\right)\right.\\
& \left.\left.- \pi\frac{(M+1-2n)B}{(M+1)T}\left(t-\frac{T}{2}\right)^2\right)\right]\cdot\text{rect}\left(\frac{4t-3T}{2T}\right)\\
& + w(t).
\end{aligned}
\tag{14}
$$

From (14), we could see that the performance of MLC signals mainly depends on the cross-coherence between the different location signals even by passing the MCR. Ideally, the signals should be orthogonal with zero cross-coherence to cancel the multiple access interference.

Set $A = \sqrt{2E/T}$, where E is the signal energy in the whole duration, and the cross-correlation of signals ρ takes the following form:

$$\rho_{l,m} = \frac{1}{E}\int_0^T S_l^*(t)\cdot S_m(t)\,dt. \tag{15}$$

Substituting the corresponding values of $S(t)$ in (13) and neglecting the integration over the higher frequencies, we get

$$\rho_{l,m} = \frac{A^4}{8E}\left[\int_0^{T/2}\cos\left(2\pi\frac{(l-m)B}{(M+1)T}\right)t^2dt + \int_{T/2}^{T}\cos\left(2\pi\frac{l-m}{M+1}B\left(t-\frac{T}{2}\right) - 2\pi\frac{l-m}{(M+1)T}B\left(t-\frac{T}{2}\right)^2\right)dt\right]. \tag{16}$$

Take $v = \sqrt{|l-m|B/(M+1)T}\,t$, $u = \sqrt{|l-m|B/(M+1)T}(t-T/2)$; then (16) can be changed to

$$
\begin{aligned}
\rho_{l,m} = {} & \frac{A^2}{4}\sqrt{\frac{(M+1)}{|l-m|BT}}\left[\int_0^{\sqrt{|l-m|BT/4(M+1)}}\cos\left(2\pi v^2\right)dv\right.\\
& \left.+ \int_0^{\sqrt{|l-m|BT/4(M+1)}}\cos\left(2\pi\sqrt{\frac{|l-m|BT}{M+1}}u - 2\pi u^2\right)du\right]\\
= {} & \frac{A^2}{4}\sqrt{\frac{(M+1)}{\Delta n\cdot D}}\int_0^{\sqrt{(\Delta n\cdot D)/4(M+1)}}\cos\left(2\pi t^2\right)dt
\end{aligned}
$$

$M = 30, l\text{-}m = 2$

(a)

$D = 100, l\text{-}m = 2$
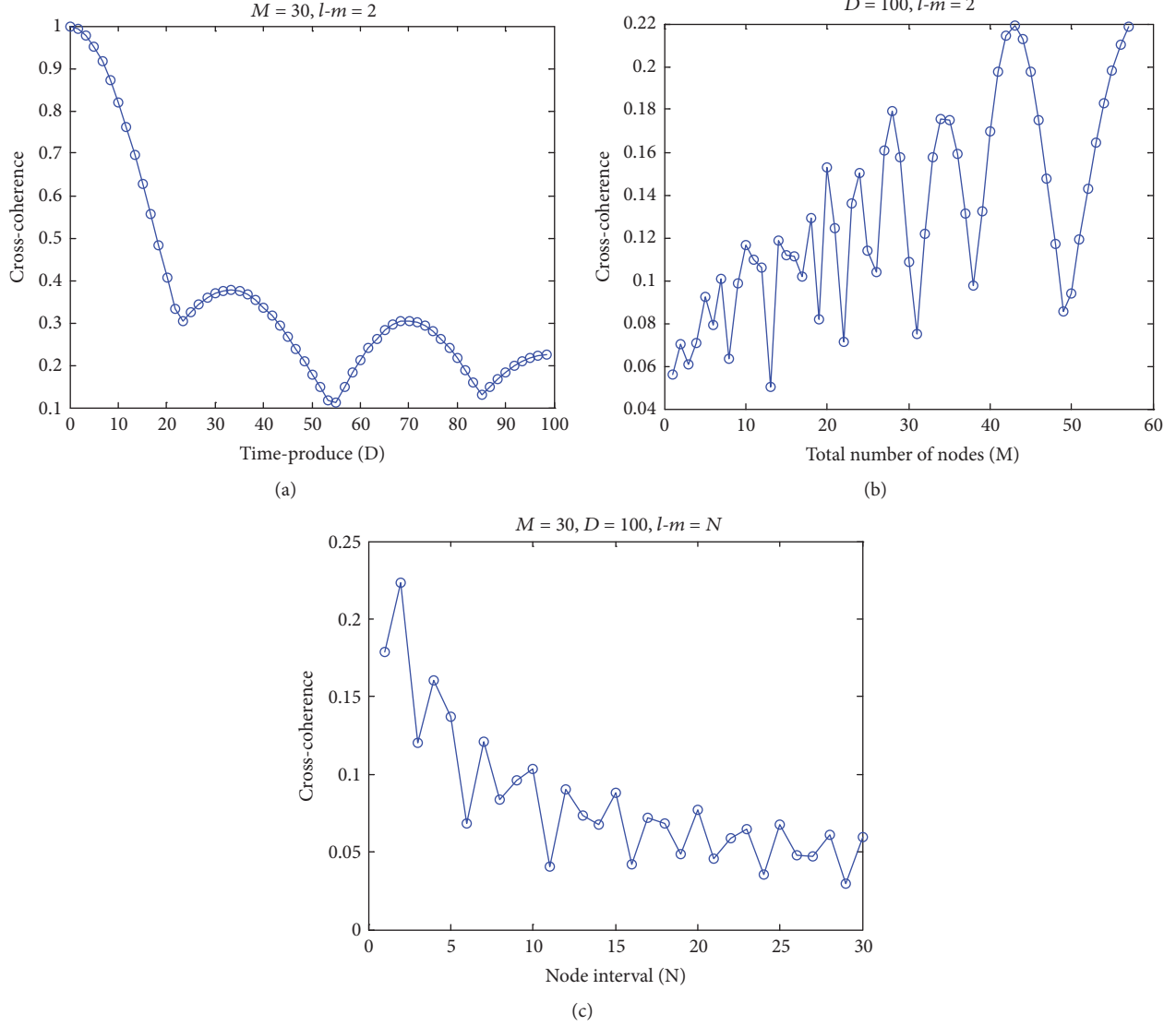
(b)

$M = 30, D = 100, l\text{-}m = N$

(c)

Figure 5: Cross-coherence function ρ as a function of (a) time-bandwidth product with $M = 30$ and $\Delta n = 2$. (b) Total node numbers M with $D = 100$ and $\Delta n = 2$. (c) User difference with $M = 30$ and $D = 100$.

$$+ \cos\left(2\pi\sqrt{\frac{\Delta n \cdot D}{M+1}}t - 2\pi t^2\right) dt,$$

(17)

where $D = BT$ is time-bandwidth product and $l-m = \Delta n$. We could see that ρ has an oscillatory nature as a function of D, M, and Δn, due to the Fresnel function form of (17). Usually, smaller cross-correlation ρ indicates a better performance of the system.

The normalized sample values can be expressed as

$$\rho_{M \times M} = \begin{bmatrix} 1 & \rho_{1,2} & \cdots & \rho_{1,M} \\ \rho_{21} & 1 & \cdots & \rho_{2,M} \\ \vdots & \vdots & & \vdots \\ \rho_{M,1} & \rho_{M,2} & \cdots & 1 \end{bmatrix},$$

(18)

where $\rho_{i,j}$ is the cross-coherence between ith and jth location signal and number 1 on the matrix diagonal represents the location signal itself. Figure 5 shows that the cross-coherence is an oscillatory function of time-bandwidth product D or the total number of MLC signals or the difference between chirp rates Δn, when the other two parameters are set fixed. Figure 5(a) indicates that ρ decreases oscillatorily as D increases when $M = 30$ and $\Delta n = 2$. Figure 5(b) indicates that ρ increases oscillatorily as M increases when $D = 100$ and $\Delta n = 2$. Figure 5(c) indicates that ρ decreases oscillatorily as Δn increases when $D = 100$ and $M = 30$. We could select the assembling parameter according to (17) in practice.

4.2. The Result of Underwater Channel.

The impulse response of an underwater acoustic channel is influenced by the geometry of the channel and its reflection properties, which determine the number of significant propagation paths, their

relative strengths, and delays. Strictly speaking, there are infinitely many signal echoes, but those that have undergone multiple reflections and lost much of the energy can be discarded, leaving only a finite number of significant paths. At this point, the channel impulse response can be expressed as

$$h(t) = A_0 \delta(t - \tau_0) + \sum_{i=1}^{N-1} A_i \delta(t - \tau_i), \qquad (19)$$

where N is the total number of propagation paths, and $i = 0$ corresponding to the direct path, A_i is the amplitude of ith propagation path at receiving side, and τ_i is the propagation delay of ith propagation path.

Assuming that the transmitted signal is $s(t)$, then the received signal $r(t)$ can be expressed as

$$\begin{aligned} r(t) &= A_0 s(t - \tau_0) + \sum_{i=1}^{N-1} A_i s(t - \tau_i) + n(t) \\ &= \sum_{i=0}^{N-1} A_i \cdot s(t - \tau_i) + n(t). \end{aligned} \qquad (20)$$

Taking the Doppler into account, let $s(t) = S_m(t)$; then the received signal can be expressed as

$$\begin{aligned} r(t) &= \sum_{i=0}^{N-1} A_i \cdot S_m(t - \tau_i) \cdot \exp\left(j2\pi\varepsilon_i(t - \tau_i)\right) + n(t) \\ &= \sum_{i=0}^{N-1} A_i \left[\cos\left(2\pi f_0(t - \tau_i) + \pi\mu_{mf}(t - \tau_i)^2\right) \right. \\ &\quad \cdot \operatorname{rect}\left(\frac{4(t - \tau_i) - T}{2T}\right) \\ &\quad + \cos\left(2\pi\left(f_0 + \frac{(M + 1 - m)}{M + 1}B\right)\left(t - \tau_i - \frac{T}{2}\right) \right. \\ &\quad \left. + \pi\mu_{mb}\left(t - \tau_i - \frac{T}{2}\right)^2\right) \cdot \operatorname{rect}\left(\frac{4(t - \tau_i) - 3T}{2T}\right) \right] \\ &\quad \cdot \exp\left(j2\pi\varepsilon_i(t - \tau_i)\right) + n(t). \end{aligned} \qquad (21)$$

Substituting the corresponding values into (21), neglecting the third branch, and using MCR we get

$$r_0(t) = \frac{1}{2}\sum_{i=1}^{N} A_i X(t) \cdot \exp\left(j2\pi\varepsilon_i(t - \tau_i)\right) + w(t), \qquad (22)$$

where

$$\begin{aligned} X(t) &= \left[\cos\left(2\pi\frac{2(M + 1 - m)B\tau_i}{(M + 1)T}t \right.\right. \\ &\quad - \pi\frac{(M + 1 - 2m)B}{(M + 1)T}t^2 + 2\pi f_0\tau \\ &\quad \left. - 2\pi\frac{(M + 1 - m)B}{(M + 1)T}\tau_i^2\right) \right] \cdot \operatorname{rect}\left(\frac{4(t - \tau_i) - T}{2T}\right) \end{aligned}$$

$$\begin{aligned} &+ \left[\cos\left(2\pi\frac{(M + 1 - 2m)BT - 4mB\tau_i}{2(M + 1)T}\left(t - \frac{T}{2}\right) \right.\right. \\ &\quad - \pi\frac{(M + 1 - 2m)B}{(M + 1)T}\left(t - \frac{T}{2}\right)^2 \\ &\quad - 2\pi\frac{(M + 1 - m)}{M + 1}B\tau_i - 2\pi f_0\tau_i \\ &\quad \left.\left. + 2\pi\frac{mB}{(M + 1)T}\tau_i^2\right) \right] \cdot \operatorname{rect}\left(\frac{4(t - \tau_i) - 3T}{2T}\right). \end{aligned} \qquad (23)$$

From the above equations, we could see that the frequencies of combined chirp signal are changed due to the Doppler, but the sets of chirp rates for different nodes are unchanged. Since our location signals are detected according to the different sets of chirp rates, which is only determined by mth order for the fixed system, the MLC signals would not be affected by multipath properties of underwater acoustic channel.

4.3. Positioning Process. The target node receives the location signals and determines the corresponding reference node with time (t_m, t_l, t_n) and position (x_m, y_m), (x_l, y_l), (x_n, y_n), when and where the location signal was transmitted. In this system, the location signals are received at the same time so the difference between t_m and t_l is a constant; in other words, the difference between the corresponding distance r_m and r_l is a constant. Of course, $|r_m - r_n|$ and $|r_l - r_n|$ are constants too. Based on the property of hyperbola, the target node is localized by the intersection of three curves as shown in Figure 6.

In Figure 6, c is the speed of sound in underwater channel in m/s. Hyperbolic curve $H_{l,n}$ is determined by nodes l and n, while $H_{m,n}$ and $H_{l,m}$ are determined by m, n and l, n respectively.

For this system, the scaling parameter and the length of interval are $S = (T/f_s)^{1/2} = N^{1/2}/f_s$ and $\Delta x = (Tf_s)^{1/2} = N^{1/2}$, respectively, where $N = f_s \cdot T$ is the sample number. When $p = p_0$ is the optimal p order, then there are two different energy accumulation peaks (assuming that the coordinates are m_{01} and m_{02}) symmetrical about $N/2$ that are got.

$$\begin{aligned} m_{01} &= \frac{N}{2} + \Delta x \cdot f_i \cdot S \cdot \sin(\alpha_0), \\ m_{02} &= \frac{N}{2} - \Delta x \cdot f_i \cdot S \cdot \sin(\alpha_0), \end{aligned} \qquad (24)$$

where f_i is the frequency of combined chirp signal with two opposite slopes at $T/2$ and $\alpha_0 = p_0(\pi/2)\,(m > N/2)$ or $\alpha_0 = (2 - p_0)(\pi/2)\,(m < N/2)$. Assuming that the time difference between the first arrival location signal and the other location signal at the receiver is τ, the later location signal's "center frequency" is changed from $f_i = f_0 + k(T/2)$ to $f_i' = f_0 + k(T/2 - \tau)$ within the short time window and the energy accumulation peak position is changed from (m_{01}, m_{02}) to (m_{01}', m_{02}').
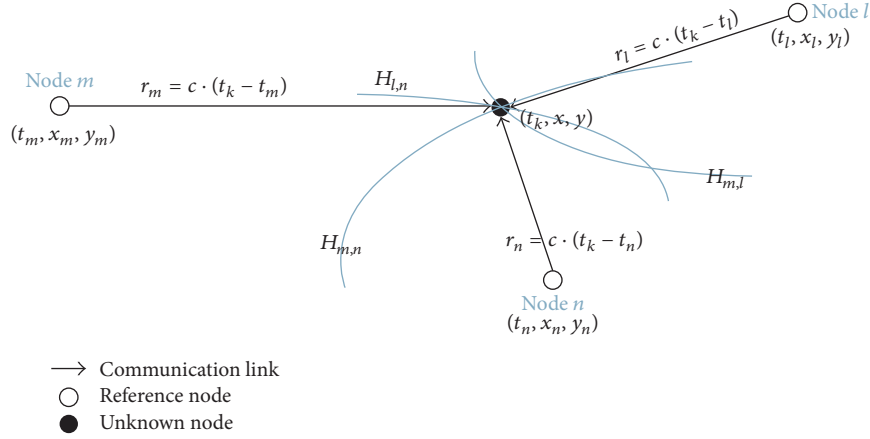
FIGURE 6: The hyperbola-based localization scheme with TDoA.

TABLE 1: The single location signal.

p order	Node					
	Node 1	Node 2	Node 3	Node 4	Node 5	Node 6
$p1 = 1.0299$	**18.02879**	5.429839	6.675838	7.805378	5.134123	5.810376
$p2 = 1.0245$	8.523125	**15.80866**	10.10544	8.86092	6.864997	4.61941
$p3 = 1.0190$	6.683892	8.061306	**23.39217**	8.208405	5.327722	4.381641
$p4 = 1.0136$	5.311406	5.342699	8.484171	**18.18579**	7.588097	7.449769
$p5 = 1.0082$	5.310347	6.943049	10.01721	9.669561	**20.46406**	7.944082
$p6 = 1.0027$	7.35016	7.233013	10.42534	10.35992	8.531059	**25.98142**

According to (24), we get

$$\frac{\left(m_1 - m_1'\right) + \left(m_2 - m_2'\right)}{2} \tag{25}$$
$$= \Delta x \cdot \left(f - f'\right) \cdot S \cdot \sin\left(\alpha\right).$$

So time difference τ can be expressed as

$$\tau = \frac{m_{01} + m_{02} - m_{01}' - m_{02}'}{2 \cdot \Delta x \cdot S \cdot k \cdot \sin\left(\alpha\right)}, \tag{26}$$

where m_{01} and m_{02} are theoretical values, while m_{01}' and m_{02}' are measured values. For each location signal, $\Delta x \cdot S \cdot k \cdot \sin(\alpha)$ is a constant; so, according to the theoretical values and measured values of position about power accumulation peak, the time difference can be got from (26). So the true value of distance difference in TDoA is $|r_m - r_n| + c \cdot \tau$ or $|r_l - r_n| + c \cdot \tau$.

5. Experimental Results

5.1. MATLAB Simulation. The simulation tool MATLAB is used to evaluate the performance of MCR-FrFT. We set typical UW-ASNs parameters set as follows: the total number of nodes, $M = 12$; the bandwidth, $B = 5$ kHz; the signal duration time, $T = 0.04$ s; and the sample frequency, $f_s = 90$ kHz. White Gaussian noise is added and the signal noise ratio (SNR) is set to 0 dB. The number of underwater multipaths is 4, the delay is $\tau = \begin{bmatrix} 0 & 0.01 & 0.013 & 0.017 \end{bmatrix}$, and

TABLE 2: Multilocation signal.

p order	(m, n, l)		
	(1, 4, 5)	(2, 3, 6)	(1, 4, 6)
$p1 = 1.0299$	**30.4154**	22.8408	**33.5639**
$p2 = 1.0245$	17.3811	**35.5565**	24.3017
$p3 = 1.0190$	24.2002	**33.2709**	22.9931
$p4 = 1.0136$	**28.0625**	27.1995	**35.7083**
$p5 = 1.0082$	**30.7174**	23.7714	23.3232
$p6 = 1.0027$	20.0376	**34.6695**	**48.8934**

the amplitude is $A = \begin{bmatrix} 1 & 0.4 & 0.3 & 0.1 \end{bmatrix}$, respectively. Discrete FrFT is used and order p at the receiving end is given by $p = 2/\pi \cdot \text{arc cot}[(M + 1)f_s/((M + 1 - 2m)B)] + 1$ for mth node.

We take single location signal and multilocation signals (three location signals), for example. For simplicity, we only consider the first half of M (here is 6) in the system. The FrFT results are shown in Tables 1 and 2.

Node refers to location signal; node 1 is the first location signal. p is the fractional power. m, l, n in Table 2 are mth, nth, lth location signal. $m = \text{round}(\text{rand}) + 1$ is 1 or 2 and $l = \text{round}(\text{rand}) + 3$ is 3 or 4, while $n = \text{round}(\text{rand}) + 5$ is 5 or 6. The numerical values in the tables are the biggest value of FrFT according to the different order power in the left of table. From Tables 1 and 2, we could see that if we set a suitable detection threshold (decision condition such as the average
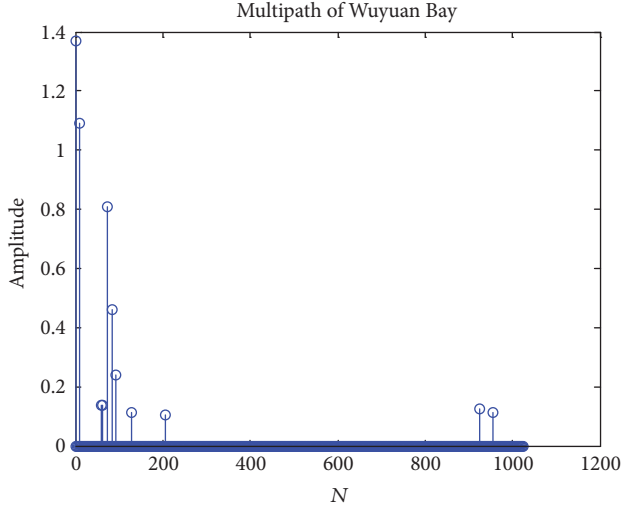
FIGURE 7: The multipath characteristic of Wuyuan Bay.

TABLE 3: The number of location signals is different.

p order	(m, l, n)		
	$(3, 0, 0)$	$(1, 6, 0)$	$(2, 5, 6)$
$p1 = 1.0420$	5.4391	**27.3816**	24.3869
$p2 = 1.0344$	10.7573	10.3540	**34.1121**
$p3 = 1.0268$	**25.2653**	5.8847	15.0960
$p4 = 1.0191$	11.6195	5.1691	10.4770
$p5 = 1.0115$	6.1462	6.5556	**29.5133**
$p6 = 1.0038$	4.3366	**21.7600**	**25.7836**
Threshold	12.594	14.8508	25.228

of all the maximum of p order FrFT), we can distinguish the location signal(s) in the received signal efficiently.

5.2. Simulation with Real Underwater Channel Impulse Response. We get the underwater multipath channel impulse response, shown in Figure 7, from Wuyuan Bay in Xiamen, China. Taking the threshold as the average of all the maximum of p order FrFT response, the result of different number of location signals can be tabulated in Table 3 (0 value of (m, n, l) means that this location node is not selected). It should be noted that we adopt f_s = 64 kHz, thus getting different p values as Table 1.

Tables 1, 2, and 3 indicate that the location signals, MLC signals, have the advance in avoiding collision.

5.3. TDoA Estimation. On the other hand, although all the nodes are synchronized, the time that the different location signals received by target node within the short time window is not considered to be the same. Estimation of the TDoA within the short time window is useful to improve the accuracy of location. The TDoA for different location signals in the short time window can be estimated by our MCR-FrFT. The simulated results are shown in Figure 8.

Assuming that the short time window is the same as signal duration time T. From Figure 8, when the true time-delay

(a)

(b)
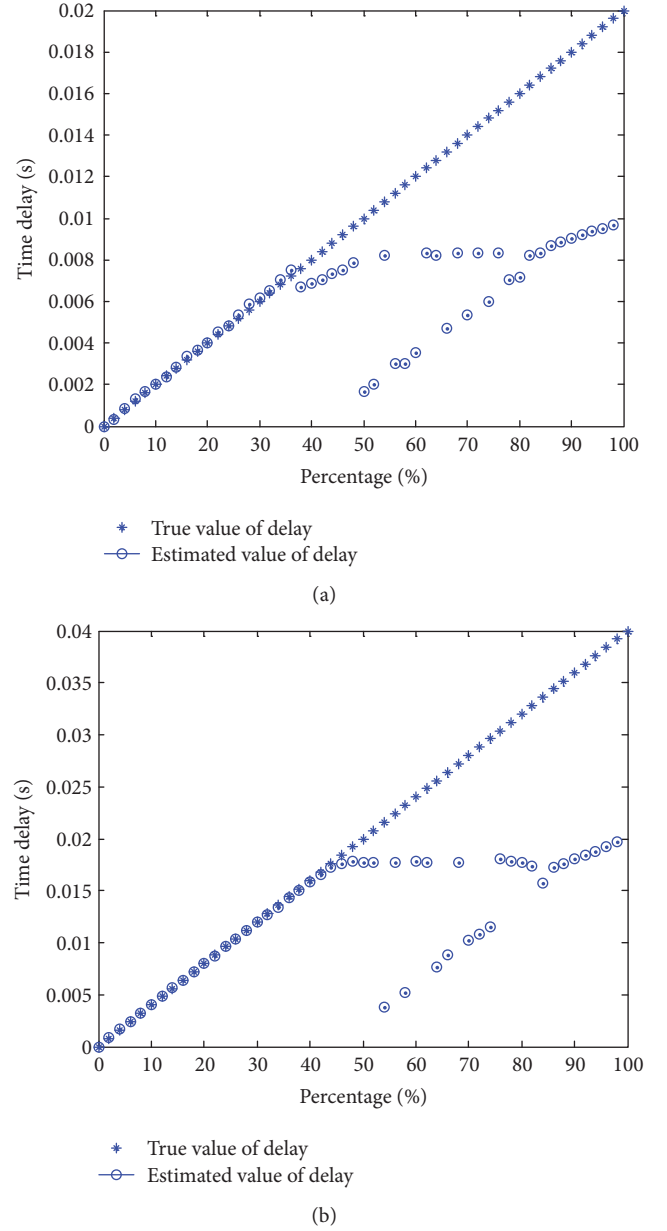
FIGURE 8: Time difference of arrival within the short time window. (a) $T = 0.02$ s; (b) $T = 0.04$ s.

value is larger than $T/2$, the estimated result is unreliable because of bigger error. Comparing Figures 8(a) and 8(b), the bigger T, the better the estimation result of time difference within $T/2$. For example, when $T = 0.02$, the estimation range is about $[0, 0.35 * T]$, but when $T = 0.04$, the estimation ranges change to about $[0, 0.45 * T]$. In summary, Figure 8 shows that our MLC signals could be estimated successfully as long as the time delay is smaller than $T/2$ with certain duration time T.

5.4. Comparisons with OFDM Based Location Scheme. OFDM based TDoA location scheme has been used in wireless metropolitan area networks (WMANs) and it gets high performance [19]. For sensor node location in underwater
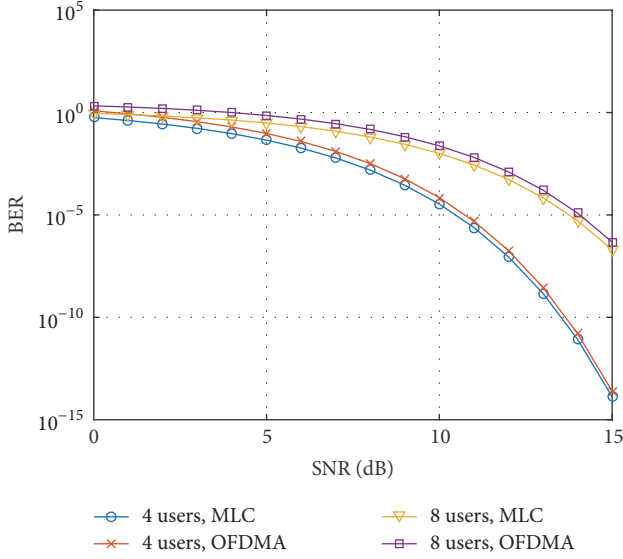
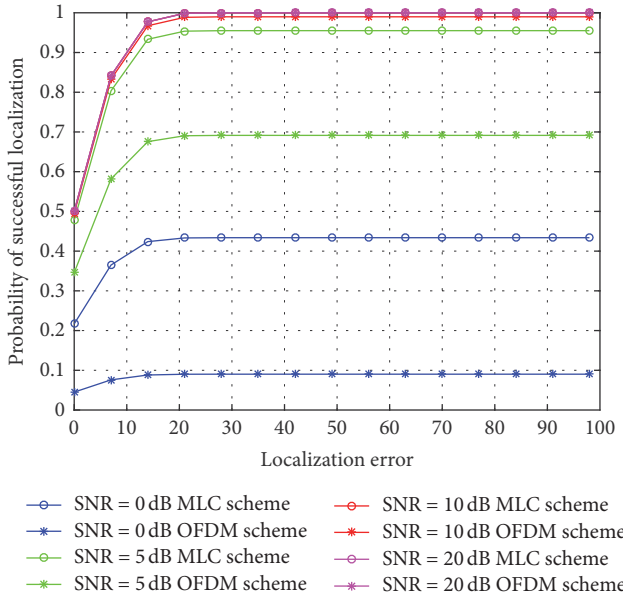FIGURE 9: Comparisons of MLC scheme and OFDM scheme.

FIGURE 10: Probability of successful localization with respect to localization error between MLC scheme and OFDM scheme.

channel environment, the MLC scheme and OFDM scheme are compared.

Firstly, we compare the MLC scheme and OFDM scheme as shown in Figure 9. It should be noted that each user of the networks occupies a subcarrier of the OFDM; therefore, this kind of multiple access could be called OFDMA. The result shows that MLC scheme always outperforms the OFDM scheme with the same total number of location signals in terms of BER with same SNR.

Then, similar to [19], the cumulative distribution function (CDF) for successful localization at SNR = 0 dB, 5 dB, 10 dB, and 20 dB in regard to the localization error (with unit m) is shown in Figure 10. The results reveal that our MLC scheme

outperforms the OFDM scheme at low SNR and is equal to OFDM one at higher SNR. This results in the lower energy requirements of the sensors. It is an excellent characteristic for underwater sensor acoustic networks (UW-ASNs), since the mobile sensors are usually powered by battery.

6. Conclusion

In this paper, we proposed a TDoA location scheme for UW-ASNs. The key idea is that we took the MLC as the location signal to avoid collision and proposed a new detection method MCR-FrFT. Compared to the complicated matching filter, the FrFT make the detection simple and efficient. Due to the time-frequency analysis ability of FrFT, we could estimate the TDoA and thus locate the sensor nodes in UW-ASNs. Both the detection diagram and theoretical derivation were given. Experimental results demonstrated that the MLC signals can avoid collision as location signals in TDoA scheme. Also the corresponding detection method MCR-FrFT could estimate the TDoA between different location signals within a short time window and locate the nodes.

Competing Interests

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

Acknowledgments

References

[1] I. F. Akyildiz, D. Pompili, and T. Melodia, "Underwater acoustic sensor networks: research challenges," *Ad Hoc Networks*, vol. 3, no. 3, pp. 257–279, 2005.

[2] M. Erol-Kantarci, H. T. Mouftah, and S. Oktug, "Localization techniques for underwater acoustic sensor networks," *IEEE Communications Magazine*, vol. 48, no. 12, pp. 152–158, 2010.

[3] I. F. Akyildiz, D. Pompili, and T. Melodia, "State-of-the-art in protocol research for underwater acoustic sensor networks," in *Proceedings of the 1st ACM International Workshop on Underwater Networks (WUW '06)*, pp. 7–16, September 2006.

[4] W.-H. Cheng, "Study mobile underwater positioning system with expendable and multi-functional bathythermographs," *Ocean Engineering*, vol. 32, no. 3-4, pp. 499–512, 2005.

[5] M. V. Jakuba, C. N. Roman, H. Singh et al., "Long-baseline acoustic navigation for under-ice autonomous underwater vehicle operations," *Journal of Field Robotics*, vol. 25, no. 11-12, pp. 861–879, 2008.

[6] X. Cheng, H. Shu, Q. Liang, and D. H.-C. Du, "Silent positioning in underwater acoustic sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 3, pp. 1756–1766, 2008.

[7] A. Y. Teymorian, W. Cheng, L. Ma, X. Cheng, X. Lu, and Z. Lu, "3D underwater sensor network localization," *IEEE Transactions on Mobile Computing*, vol. 8, no. 12, pp. 1610–1621, 2009.

 [8] J. Scheuing and B. Yang, "Disambiguation of TDOA estimation for multiple sources in reverberant environments," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 8, pp. 1479–1489, 2008.

 [9] J. Yi, D. Mirza, C. Schurgers, and R. Kastner, "Joint time synchronization and tracking for mobile underwater systems," in *Proceedings of the 8th ACM International Conference on Underwater Networks and Systems (WUWNet '13)*, p. 38, Kaohsiung, Taiwan, 2013.

[10] W. Lihui, P. Nanjiang, G. Lei, and H. Yang, "CDMA in underwater positioning of application," *Electronic Test*, vol. 8, article 032, 2011.

[11] S. E. El-Khamy, S. E. Shaaban, and E. A. Thabet, "Efficient multiple-access communications using multi-user chirp modulation signals," in *Proceedings of the 4th International Symposium on Spread Spectrum Techniques & Applications (ISSSTA '96)*, pp. 1209–1213, September 1996.

[12] H. Shen, S. Machineni, C. Gupta, and A. Papandreou-Suppappola, "Time-varying multichirp rate modulation for multiple access systems," *IEEE Signal Processing Letters*, vol. 11, no. 5, pp. 497–500, 2004.

[13] C. He, J. Huang, Q. Zhang, and K. Lei, "Reliable mobile underwater wireless communication using wideband chirp signal," in *Proceedings of the WRI International Conference on Communications and Mobile Computing (CMC '09)*, pp. 146–150, Yunnan, China, January 2009.

[14] E. Calvo and M. Stojanovic, "Efficient channel-estimation-based multiuser detection for underwater CDMA systems," *IEEE Journal of Oceanic Engineering*, vol. 33, no. 4, pp. 502–512, 2008.

[15] J. Song, Y. Wang, and Y. Liu, "Iterative interpolation for parameter estimation of LFM signal based on fractional Fourier transform," *Circuits, Systems, and Signal Processing*, vol. 32, no. 3, pp. 1489–1499, 2013.

[16] H. Hao, "Multi component LFM signal detection and parameter estimation based on EEMD–FRFT," *Optik-International Journal for Light and Electron Optics*, vol. 124, no. 23, pp. 6093–6096, 2013.

[17] L. Shen, Q. Yin, M. Lu et al., "Linear FM signal parameter estimation using STFT and FRFT," *Chinese Journal of Electronics*, vol. 22, no. 2, pp. 301–307, 2013.

[18] L. B. Almeida, "The fractional Fourier transform and time-frequency representations," *IEEE Transactions on Signal Processing*, vol. 42, no. 11, pp. 3084–3091, 1994.

[19] H. Ni, G. Ren, and Y. Chang, "A TDOA location scheme in OFDM based WMANs," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 3, pp. 1017–1021, 2008.

Research Article

A Measurement Study of BLE iBeacon and Geometric Adjustment Scheme for Indoor Location-Based Mobile Applications

Jeongyeup Paek,[1] JeongGil Ko,[2] and Hyungsik Shin[3]

[1]*School of Computer Science and Engineering, Chung-Ang University, Seoul, Republic of Korea*
[2]*Department of Software and Computer Engineering, Ajou University, Suwon, Republic of Korea*
[3]*School of Electronic and Electrical Engineering, Hongik University, Seoul, Republic of Korea*

Correspondence should be addressed to Hyungsik Shin; hyungsik.shin@hongik.ac.kr

Bluetooth Low Energy (BLE) and the iBeacons have recently gained large interest for enabling various proximity-based application services. Given the ubiquitously deployed nature of Bluetooth devices including mobile smartphones, using BLE and iBeacon technologies seemed to be a promising future to come. This work started off with the belief that this was true: iBeacons could provide us with the accuracy in proximity and distance estimation to enable and simplify the development of many previously difficult applications. However, our empirical studies with three different iBeacon devices from various vendors and two types of smartphone platforms prove that this is not the case. Signal strength readings vary significantly over different iBeacon vendors, mobile platforms, environmental or deployment factors, and usage scenarios. This variability in signal strength naturally complicates the process of extracting an accurate location/proximity estimation in real environments. Our lessons on the limitations of iBeacon technique lead us to design a simple class attendance checking application by performing a simple form of geometric adjustments to compensate for the natural variations in beacon signal strength readings. We believe that the negative observations made in this work can provide future researchers with a reference on how well of a performance to expect from iBeacon devices as they enter their system design phases.

1. Introduction

Bluetooth is ubiquitous in today's everyday life, used in laptops, mobile devices, keyboards, headsets, and many other consumer electronics for wire replacement. Bluetooth Low Energy (BLE) is an extension of the Bluetooth standard in version 4.0 that enables low-power low-cost short-range wireless communication [1, 2]. It has significantly reduced its power consumption, among other extended features, compared to classical Bluetooth versions, and is now possible to run BLE devices for several months to even a couple years on a coin cell battery. Such aspects make BLE ideal for applications requiring infrequent or periodic transfers of small amount of data; thus, it can be applied in a wide range of medical, industrial, and consumer applications.

iBeacon is a BLE-based proximity sensing framework proposed by Apple [3], which allows a mobile device to detect its proximity to an iBeacon station (and possibly its location) by knowing how close it is to (usually wall-mounted) low-complexity, low-cost BLE transmitters, the iBeacons. Each iBeacon transmits periodically short identification frames that are received by a mobile BLE device to estimate the distance between the mobile device and the iBeacon using received signal strength indicator (RSSI). Based on this detection of proximity, iBeacons provide automatic and location specific triggering of services on the mobile device such as advertising, coupons, or route guidance. Although the specifications of the iBeacon were initially proposed by Apple (iOS7 or later), it is just one way of utilizing BLE's proximity features, and thus it is (and can be made) compatible with other devices (e.g., Android 4.3 or later) that use BLE. More generally, any BLE device can use the same concept to provide similar proximity functionality. Technically, iBeacon

technology is a subset of BLE beacons, but iBeacon and BLE proximity beacons became nearly synonyms today despite Google's recent release of their own open beacon format called Eddystone (Eddystone is a protocol specification by Google that defines a Bluetooth Low Energy message format for proximity beacon messages) [4].

Note here that the original goal of BLE beacons was to provide proximity-based application services, possibly to extend the features to coarse-grained location-based applications. Given that BLE beacons solely focus on RSSI measurements, combined with lessons learned from the long history of research in RF-based localization, it is somewhat expected to be able to achieve only a limited accuracy in indoor localization services. It is well known that the variability of RF signals with respect to the indoor environment complicates the accurate estimation of the mobile devices' location or its relative distance to the BLE beaconing transmitter device. However, most prior work on BLE-based indoor localization merely states the possibility of error due to signal variation but does not quantify them, and their reported results show only the positive side of the picture. For example, the work in [5] reports a positioning accuracy of 0.53 meters, but this work was bounded by a $9 \times 10\,m^2$ area. The work in [6] reports a 1.2-meter accuracy with fingerprinting, but only within a $47.4 \times 15.9\,m^2$ area. Recently work in [7] shows that distance estimation error can increase significantly with longer distance when the measured signal RSSI vary over time but does not show other dimensions of possible error.

In this work, we perform an extensive set of experiments to quantify the impact of various indoor obstacles on the BLE signal variance. Specifically, we show that different iBeacon devices from different vendors along with the paired mobile device platform (e.g., iOS or Android) can give significant impact on the RSSI measurements to complicate the process of designing a generally acceptable distance/location estimation model. Furthermore, practical factors such as the deployment height or environmental factors also introduce a significant amount of impact on the RSSI values to introduce an additional level of complexity in designing highly accurate location estimation systems using iBeacons. Lessons from these empirical studies lead us to design an application that is "practically designable." Specifically, we design and evaluate a class attendance monitoring application system, by combining distance estimations made from locally accessible iBeacon signals and an estimation adjustment scheme designed from the lessons from our empirical signal measurements.

This paper is structured as follows. Section 2 introduces the relationship between RSSI readings and the estimated distances and then motivates our empirical studies on the relationship. Section 3 describes our approaches to the empirical studies and experiment settings including various parameters that we change during our studies. Section 4 is the main section, and it describes our measurements data and illustrates our analysis results based on the data. Section 5 presents an application case study, an automatic attendance monitoring system, which is based on our empirical studies. This system also proposes a novel geometric approach to overcome the limitations of practical iBeacon systems.

Section 6 reviews prior and related work and finally Section 7 concludes the paper.

2. Limitations of Bluetooth Low Energy-Based Localization Systems

As Bluetooth Low Energy (BLE) and Apple's iBeacon protocols begin to see widespread adoption and deployment, indoor proximity and/or location enabled applications are quickly penetrating into our everyday lives and consumer space. iBeacon has been installed in Apple stores in the US to provide product information as well as customer services, and the National Football League (NFL) has used iBeacons in the MetLife Stadium to provide personalized advertisements to football fans during the Super Bowl game [8]. Furthermore, iBeacon can be used to track luggage at the airport [9], provide interactive experiences to visitors in museums [10], plan evacuation paths in emergency guiding systems [11], track patients in emergency rooms [12], guide indoor/outdoor routes [13, 14], and detect the occupancy of rooms [15, 16]. The application space to which iBeacons and BLE proximity applications can be applied is extremely large. These applications all use proximity measurements based on the received signal strength indicator (RSSI) as the sole source of information; thus, these applications exploit information on how close a mobile device is to an iBeacon rather than trying to pinpoint the exact and absolute location of the mobile device.

The original purpose of iBeacon technology is to provide proximity-based application services and coarse-grained indoor location positioning and navigation based solely on proximity. Note that proximity (being near to some object) is related to the location (where you are) but is not necessarily identical. A location (or position) is more than just proximity, and it is an absolute value usually defined by some coordinate system (e.g., latitude and longitude for GPS). Recently, however, there have been several proposals to use BLE for indoor positioning, extending several prior works that use other technologies such as ultrasonic sound, infrared, WiFi, GSM, RFID, IEEE802.15.4, and earlier versions of Bluetooth for indoor localization [17–22] (there are many more related prior work on indoor localization that uses other wireless technologies, but we list only a subset here for brevity and to keep the reference list focused on BLE only since there are too many of them). Indoor positioning and localization are still an active area of research, and BLE is extending the literature using its new wireless features.

There are several approaches (beyond simple proximity-based ones [12]) that can be applied to BLE-based indoor positioning. In all cases, an iBeacon will periodically broadcast an advertisement packet containing a unique ID and a calibrated RSSI value (R^0) corresponding to one-meter distance (d^0). This value allows us to determine the distance between an iBeacon and a device using a model in (1), where γ is a calibration parameter for the path loss exponent. Based on the RSSI or the estimated distances between iBeacons and a mobile device, a weighted average model [5], trilateralization/triangulation model [23], or fingerprinting [7, 24] can

FIGURE 1: Three different types of iBeacons used for the experiments: top row shows the images of the Estimote, GELO, and Wizturn Pebble iBeacons. Bottom row shows the icons of their corresponding smartphone applications.

be applied for the localization process. The estimations can also be combined with Pedestrian Dead Reckoning (PDR) or filtering techniques to improve tracking or to reduce positioning errors [6, 25]. Of course, all of these approaches assume that all installed iBeacons possess their predefined location information including their exact coordinates.

$$\text{RSSI} = R^0 - 10\gamma \log\left(\frac{\widehat{d}}{d^0}\right) \implies$$

$$\widehat{d} = 10^{(R^0 - \text{RSSI})/10\gamma}.$$

(1)

If distances are used instead of raw RSSI values for positioning, RF propagation model such as (1) is used to convert RSSI readings into distances. The main challenge here is the sensitivity of RSSI values to environment changes such as object movements or obstacles, which result in signal propagation or radio map changes. It is obvious that the accuracy and efficiency of this distance estimation (and thus the location estimation) depend heavily on the accuracy of the measured RSSI values and the model used to derive and calculate the distance and also significantly influenced by the surrounding environment. Therefore, in this work, we perform a detailed empirical measurement study on the location/distance estimation performance based on RSSI measurements for different types of iBeacon devices and mobile device platforms. We consider various environmental factors to gather a practical perspective on the performance limitations of iBeacon technologies.

3. Approach and Experiment Setup

We carried out experiments in two different environments: an open soccer field of a university campus without any obstacles and an unblocked corridor of an office building to ensure line-of-sight (LOS) signal propagation. In these environments, we used three different types of iBeacons: Estimote [26], Wizturn Pebble [27], and GELO [28] iBeacons to examine differences in performance among vendors. Furthermore, we used two different mobile devices: iPhone 5 with iOS 7.1.2 and Galaxy Round (SM-G910S) with Android 4.4.2, to verify any existing particularities. Mobile applications used for data collection are the applications provided by the respective iBeacon vendors: "Estimote app" for Estimote iBeacons, "GELO toolkit app" for GELO iBeacons, and "Wizturn beacon manager" app for Wizturn Pebble iBeacons (Figure 1).

The default configuration parameters used in the experiments are as follows. The transmission (TX) power of iBeacon was set to the maximum of 4 dBm unless stated otherwise to verify the maximum distance that can be covered by any iBeacon. We also experiment with minimum TX power (which is −19∼−23 dBm depending on the device) to observe the lower-bounds of connectivity. The advertising interval of the iBeacons was set to 950 milliseconds, which was the factory default for all three devices. Device placement height from ground was set to 1.2 meters for both the iBeacons and the mobile phones unless stated otherwise. We selected this height of 1.2 meters to imitate the height at which a user holds their mobile device in their hands. Finally, the orientation of the iBeacon was set to face the direction of LOS towards the mobile device. Using these configurations, a mobile application continuously detected signals from iBeacons and logged their RSSI measurements while the distance between the iBeacon and the mobile device was increased from 1 meter to the maximum distance and recorded manually until reaching the maximum distance. Here, we note that the maximum distance is defined to be the farthest distance where the RSSI value can be detected by the device. At each location (e.g., relative distance to the iBeacon) we took twenty RSSI measurements for each configuration.

4. Measurement Data Analysis

In this section, we focus on identifying and quantifying the various characteristics related to the performance of
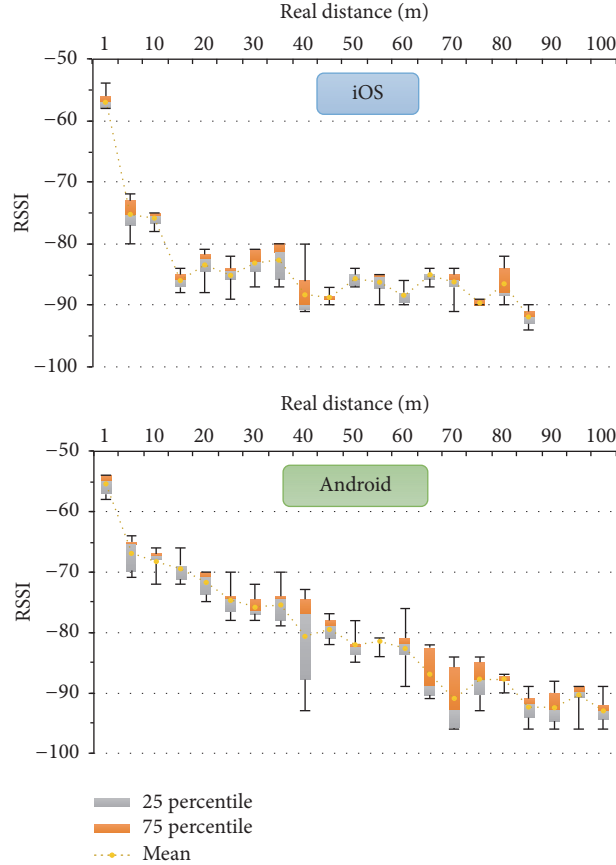
FIGURE 2: Distance versus RSSI plot for iOS and Android comparison. (Estimote, 4 dBm TX power, 1.2 m height, and LOS).

iBeacons for localization applications by carefully analyzing the extensive measurement data that we have collected.

4.1. Difference between iOS and Android Phones. We first start by comparing the differences between iPhone 5 with iOS 7.1.2 and Galaxy Round (SM-G910S) with Android 4.4.2 in receiving iBeacon signals. The experiment was conducted outdoors in an open soccer field using an Estimote beacon with 4 dBm TX power at 1.2-meter height in LOS. Figure 2 presents our results from this experiment, where the dotted line represents the average RSSI reading at each distance, and the boxes represent the 25 and 75 percentile values along with their minimum and maximum error bars.

From this result, we can identify several interesting facts. First, compared to the 100 meters for the Android platform, iOS showed notably shorter maximum distances of 85 meters. Note that the 100 meters observed for the Android is not the actual maximum distance, given that 100 m was the maximum length we could achieve for LOS links in our environment. Therefore, the difference between the maximum distances of the two platforms turned out to be very large. Second, RSSI readings on Android phone decreased more gradually whereas iOS showed a sudden drop in RSSI after ~10 meters. Lastly, RSSI readings on the Android platform had more temporal variation than iOS. While it is difficult to conclude whether the hardware

specification shows any difference or this is an effect of the software platform, it is known that there is a difference between iOS and Android in scanning and sampling iBeacon advertisements [16]. More importantly, the RSSI readings differ between different phones (regardless of it being an impact of the operating system or the hardware), and we should point out that this effect itself will have significant implications for distance estimation and localization, causing difficulties for the application developers to generalize information extracted from the collected RSSI values.

4.2. Effect of Device Placement Height. In this experiment, we examine the impact of the iBeacon installation height from the ground on RSSI. For this, we place an Estimote beacon on the ground (e.g., 0-meter height) and measure the RSSI readings using the Android mobile phone. We compare this value with the case when the same beacon is installed at a 1.2-meter height. Figure 3 shows the result from this experiments. We can first notice that the values are substantially different for the two experiment cases. When the iBeacon is placed on the ground while all other configurations are identical, the maximum distance reduces from 100+ meters to only 12 meters with significant and drastic drop in RSSI. This result implies that the placement of iBeacon is very important when designing and deploying an iBeacon system. Distance estimation and thus the localization accuracy can be
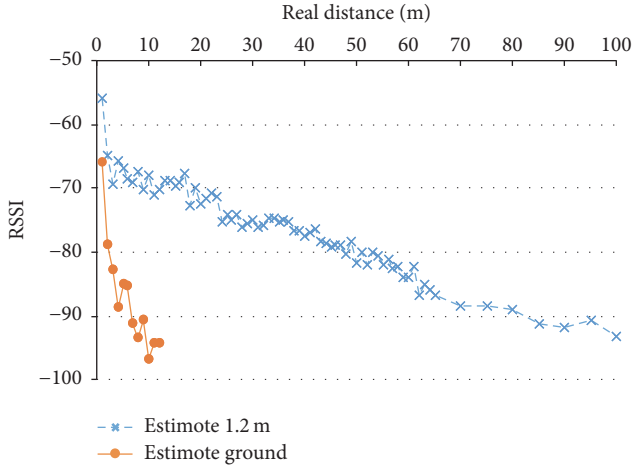
FIGURE 3: Distance versus RSSI plot that shows the impact of device placement height from ground (Estimote, 4 dBm TX power, and Android).
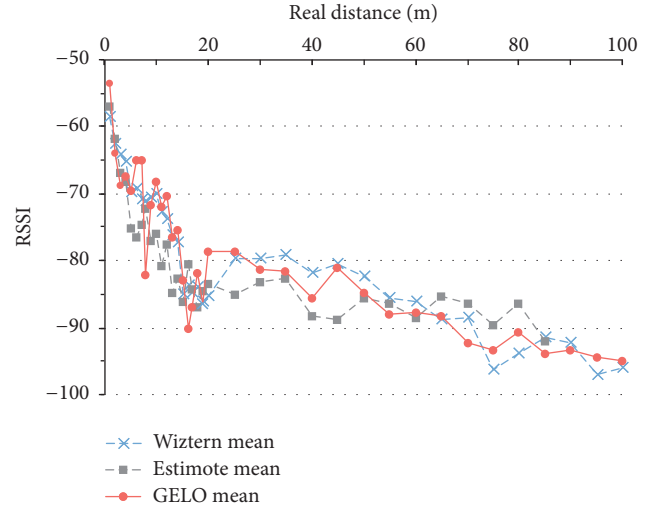

FIGURE 4: Distance versus RSSI plot for three different types of iBeacons: Estimote, GELO, and Wizturn Pebble (4 dBm/0 dBm TX power, 1.2 m height, LOS, and iOS).

significantly impacted by the precise placement of iBeacons, and planning the installation carefully as part of the site survey will be very important. Based on our experimental data, the optimistic advertisements that declare "ease of deployment" and "place nodes anywhere!" may not be valid when targeting precise accuracy on the iBeacon systems.

4.3. Difference between iBeacons from Different Manufacturers. Next, we compare the RSSI reading differences between different iBeacon products for three different manufacturers: Estimote, GELO, and Wizturn Pebble. iPhone5 was used as the mobile device, and both the beacons and the mobile devices were installed at 1.2 meters' height in an outdoor soccer field. One important thing to note is that the maximum transmission power supported by GELO iBeacon was 0 dBm whereas Estimote and Wizturn Pebble allowed configuration up to 4 dBm. We used respective maximum powers for each device to examine the maximum distance coverage. Our initial intuition was that GELO iBeacon will exhibit 4 dBm lower RSSI values compared to other beacons at identical distances, and thus it will result in a shorter maximum transmission distance.

Surprisingly, Figure 4 shows that, unlike our expectation, GELO along with Wizturn beacons shows a maximum distance of 100+ meters, far exceeding Estimote's 85 meters. Furthermore, the RSSI readings were all similar between distinct iBeacon types despite the 4 dBm difference in transmission power. This implies several interesting points. First, the configured TX power difference does not directly translate into received RSSI reading difference, which makes the calibration process for distance estimation (and thus indoor localization) more challenging when the transmission power must be adjusted for the given environment. Second, iBeacons from different vendors exhibit different maximum distances, and higher maximum TX power configuration does not necessarily lead to longer maximum transmission distances. Lastly, due to the aforementioned reasons, it will be very challenging for application developers and service

providers to design an iBeacon-based system with heterogeneous iBeacons since multiple sets of calibration parameters will be required for each pair of iBeacon and mobile device type. Again, these results imply that designing a reasonably reliable iBeacon-based localization system can be extremely challenging.

4.4. Reducing to Minimum TX Power. We now focus on the fact that not all systems fully benefit from the use of maximum transmission power. For example, if we are designing an indoor localization system based on trilateralization or fingerprinting method, we would prefer a higher transmission power to cover a larger area and/or achieve denser deployments while using fewer number of iBeacons (e.g., for cost reasons). However, when building a system based on only the proximity function, reducing the transmission power to the minimum just enough to cover the target region (e.g., in front of a store or a product for advertisement) while keeping the energy cost low to extend the lifetime of the batteries can be a reasonable system design option. With this motivation in mind, we conducted an experiment using the respective minimum TX power for the three types of beacons. The minimum TX power configuration allowed on Estimote, GELO, and Wizturn Pebble beacons was −20 dBm, −23 dBm, and −19 dBm, respectively.

Figure 5 depicts the results from this study. First, we observe the drastic drop in the maximum distance (note the different x-axis values compared to the previous figures). Estimote with −20 dBm showed a maximum distance of 8 meters and Wizturn Pebble with −19 dBm gave 8 meters while the RSSI from a GELO beacon with −23 dBm TX power was detectable within only 0.3 meters. The second observation is that while the RSSI reductions for Estimote and Wizturn Pebble approximately matched their reduction in TX power (24 and 23 dBm, resp.), this was not true for the GELO beacons. GELO beacons showed RSSI reductions of
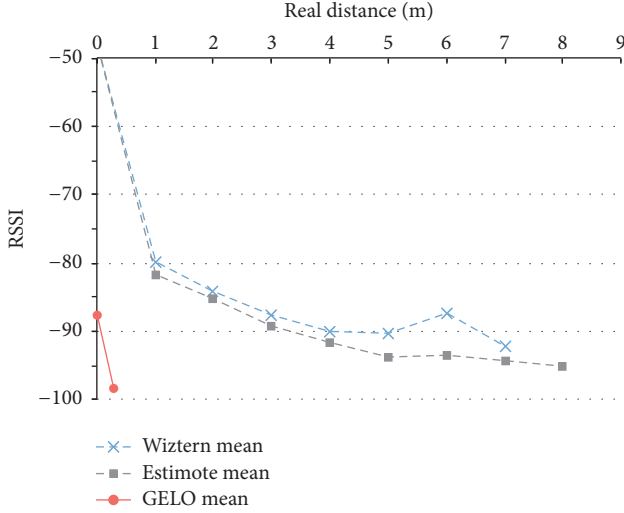
FIGURE 5: Distance versus RSSI plot when minimum transmission power is used by three different types of iBeacons (minimum TX power, 1.2 m height, LOS, and iOS).

FIGURE 6: Distance versus RSSI plot for indoor versus outdoor comparison (Estimote, 4 dBm TX power, 1.2 m height, LOS, and Android).

44–58 dBm while the TX power was reduced only by 23 dBm. Again, as our previous results concluded, these differences can cause significant complication when designing a system with multiple types of iBeacons; different TX powers, different reductions in RSSI, and different maximum distances will make the calibration process practically impossible unless only a single TX power configuration and a homogeneous iBeacon from a particular vendor is used for designing the entire system. This may be a plausible approach in the initial system design phase, but with the system scaling overtime for supporting additional regions or new use cases, considerations for heterogeneous systems are a must.

4.5. Indoors versus Outdoors. iBeacon is often referred to and is well known as an "indoor proximity/location system" because a major motivator for its development was to provide indoor location awareness where the already widely used GPS is unavailable. However, physically, there is no reason why iBeacon cannot be used outdoors. In fact, there are manufacturers (e.g., GELO) that provide durable weather-proof iBeacons for outdoor use. Through our next experiments we wanted to compare the performance of iBeacons in the indoor and outdoor environments. For this, we performed additional experiments in an unblocked 3 meters' wide corridor of a university building. Here, we use the Estimote beacons and an Android phone, again at 1.2 meters height and 4 dBm TX power so that we can make direct comparisons with our experiments performed outdoors.

Figure 6 presents our results. To our surprise, after ~25 meters in distance, the RSSI readings remained almost steady, or even increased slightly with distance. The maximum distance is shown as 50 meters only because our corridor within the building was 50 meters' long, and the actual maximum distance is projected to be longer with more physical space. To verify our unexpected results, we also conducted additional experiments with the Wizturn Pebble
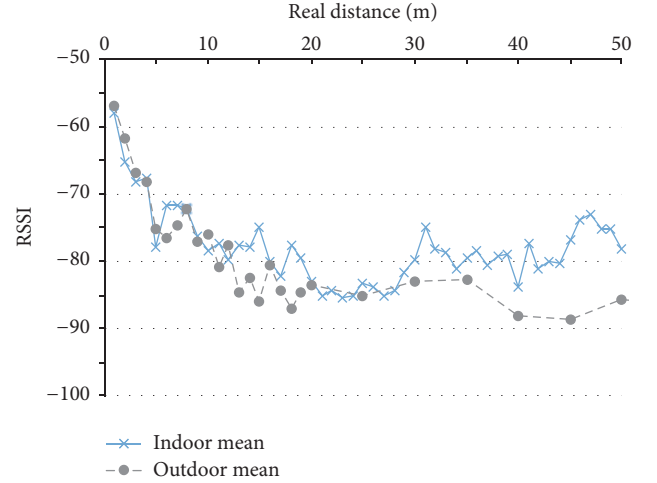
and GELO beacons and found similar results (plots omitted for brevity). The only plausible explanation for such behavior is the multipath effect caused from the walls and ceiling of the corridor. What is more important here is the fact that the RSSI value did not decay for a wide range of distances (25–50 m in our case). This means that the distance model in (1) is no longer valid, and any indoor localization method that relies on distance estimation and trilateralization can face large errors in the order of tens of meters. In bolder words, it is *not* possible to design an accurate indoor localization system with iBeacons using simple distance estimation.

4.6. Effect of WiFi on iBeacon Signal Reception. Another issue when using iBeacons for indoor applications is the interference effect of WiFi signals, which operates on the same 2.4 GHz ISM frequency band. BLE was designed to use channels 37, 38, and 39 for advertisement (out of 40 total channels and the remaining 37 channels for data communication), and their allocated frequencies are designed to avoid the most popular WiFi channels 1, 6, and 11 (c.f. Figure 7). Unfortunately, the WiFi AP deployment at our university campus used channels 1, 5, and 9, and WiFi channel 5 happens to interfere with BLE channel 38. This is not an unusual case and can occur generally in any WiFi environment. Therefore, we performed an additional experiment to quantify the consequences of WiFi-oriented collisions.

For this experiment, we placed an Estimote beacon directly under a WiFi AP and measured the RSSI readings from an Android phone 5 meters apart (c.f. Figure 8). We show a subset of observations in Figure 9 where the *x*-axis represents the sequence of measurements (e.g., time) and the *y*-axis is again RSSI. The dotted line with squares shows the readings when WiFi AP was powered off, and the solid line with circles shows the readings when the WiFi AP is in normal operation. The shaded rectangle regions, along with the disconnection in the solid line, represent the cases when
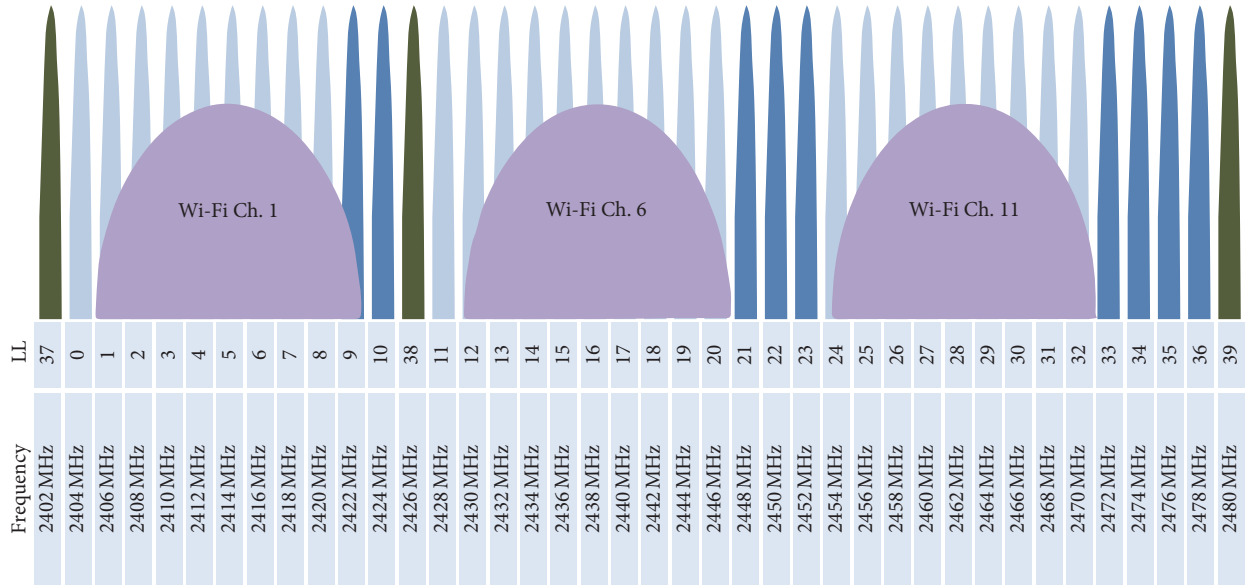
FIGURE 7: Channel frequency allocation for BLE and its relation to WiFi channels 1, 6, and 11.
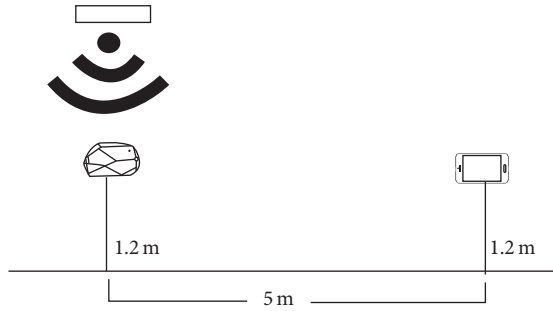

FIGURE 8: Setup for WiFi interference experiment. An Estimote iBeacon was placed vertically below an WiFi AP that uses channel 5, and the receiving Android mobile smartphone was placed 5 meters away. Both the iBeacon and the mobile device were 1.2 meters from the ground.
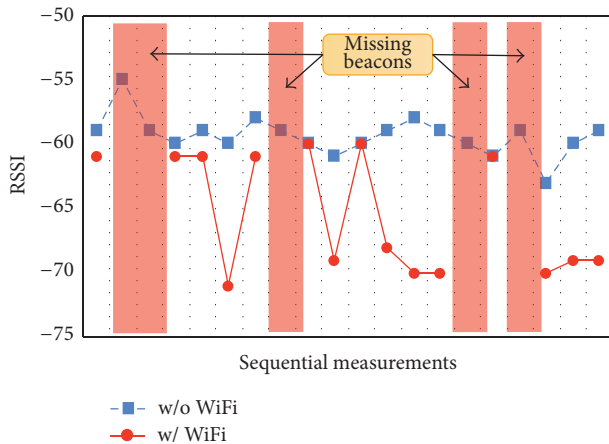

FIGURE 9: Time versus RSSI plot for a sequence of beacon reception that shows reception failures (packet losses) and reduction in RSSI (identical configuration as Figure 8).

there was no iBeacon RSSI reading on the mobile phone, due to the lack of packet reception.

We make two important observations from this experiment. First, when the WiFi AP is active, the beacon reception ratio dropped to around 75% even at a short distance of 5 meters, a distance where a 100% beacon reception can be achieved without the WiFi AP. Second, even when an iBeacon advertisement was successfully received, the RSSI readings showed significantly lower values (e.g., more than 10 dBm reduction) for ~53% of the beacons. Note that the packet loss of low-power BLE beacons under high-power WiFi interference is somewhat expected, but a consistent reduction in RSSI is an unexpected phenomenon that can be regarded as another type of wireless signal "gray region" [29]. These findings suggest that, with the widely deployed and ubiquitously utilized WiFi in today's indoor environments, an iBeacon system can be significantly affected with reduced reliability and higher estimation errors.

4.7. Effect of Obstacles. In this experiment, we examine the effect of physical obstacles on the iBeacon signal reception and compare against the LOS case. Specifically, here, we considered six different cases: three cases with obstacles of an iron door, a wooden door and a window each, one case for signal amplification using a sheet of aluminum foil, and two cases where the mobile phone is covered by hand or paper. In this experiment, we used the Estimote beacon paired with an iPhone5, and the distance between the iBeacon and the mobile phone was ~3 meters. Again, the height was 1.2 meters and the TX power was 4 dBm.

We present the results from this study in Figure 10. As expected, the iron door blocked the signal drastically with an RSSI drop of ~20 dBm, but the wooden door or window also had some (nonnegligible) effect of ~3 to 8 dBm drop, while covering the phone with a pile of paper caused a ~6 dBm drop.
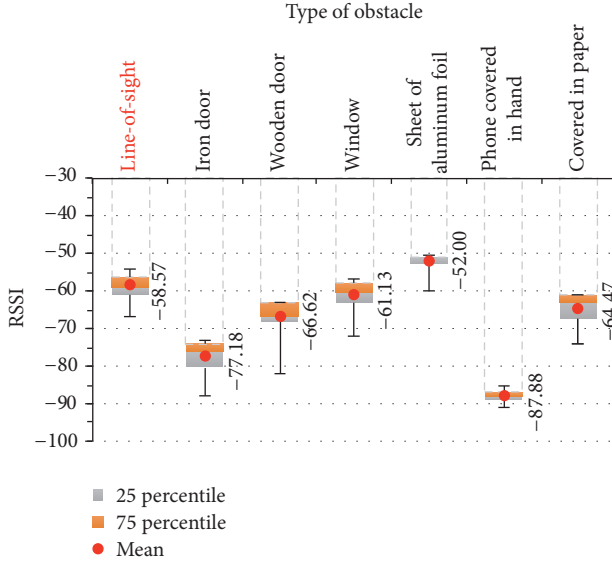
FIGURE 10: Average RSSI plot for different types of obstacles (Estimote, 4 dBm TX power, 1.2 m height, 3 m distance, and iOS).
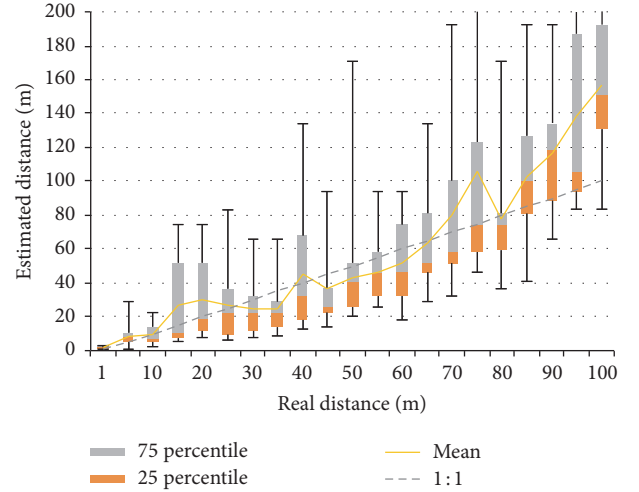

FIGURE 11: Real distance versus estimated distance plot where the previously collected RSSI data was used to curve fit the model in (1) and empirically determine the model parameter.

More interestingly, however, covering the mobile phone with the hands of the user resulted in a noticeable reduction of ~30 dBm. This suggests that, in an indoor localization system, despite undergoing a careful deployment phase and extensive calibrations, the distance and/or position estimation can have significant errors just because the user is holding the phone tightly in his/her hands or the environment changes naturally (e.g., door statuses). This is another challenge in designing an iBeacon-based system since mobile devices will inevitably be held by the users' hands and obstacle statuses will change actively in an indoor environment.

As a final note, we were able to amplify the beacon signal by using a sheet of aluminum foil behind the iBeacon which resulted in an RSSI increase of 6 dBm. While it is unlikely that iBeacon deployment will intentionally be wrapped in a sheet of aluminum foil, this result suggests that some environmental artifact or ornament can also unexpectedly cause such an effect: implying that obstacles not only reduce the RSSI levels but can also amplify the signal strength.

4.8. Distance Estimation Using the Curve-Fitted Model. Finally, we now take all the measurement data from the four iBeacon-mobile device pairs, combination of Estimote and Wizturn Pebble beacons with Android and iOS phones, and fit the data on to the model in (1) using the least-mean square method to calculate the estimated distance based on the RSSI. We note that all data are for the outdoors, LOS, 1.2-meter height, and 4 dBm TX power experiments. We then use this derived model to compare the real distance to the estimated distance as in Figure 11. One point to take away from this plot is the fact that distance estimation can show significant errors even under LOS conditions due to high signal strength variations. Furthermore, the errors increase as distance increases (e.g., as RSSI gets closer to the reception sensitivity). While we omit additional figures for brevity, we

note that we can see similar plots even with data from a single pair of iBeacon-mobile device. Quantitatively, the errors can increase from tens of meters to even hundreds of meters. This means that when designing an iBeacon-based system, a system designer should either cope with the large distance estimation errors (and hence positioning errors) or densely deploy a large number of iBeacons to reduce the errors, which can threat the "low-cost" argument in iBeacon systems. Unfortunately, as per our experimental results, neither is satisfactory.

5. Application Case Study: Automatic Attendance Checker System

As our experimental results show, the main challenge in using iBeacon signals for accurate indoor positioning is the variability of RSSI readings and its sensitivity to environment changes which result in drastic changes in signal propagation. Our findings in Section 4 show that the RSSI value (and the corresponding signal propagation model for estimating distances) varies significantly among iBeacons from different vendors (e.g., Estimote versus Wizturn Pebble versus GELO), mobile device types (iOS versus Android), height of the device installation from the ground, indoor or outdoor environmental factors, and physical obstacle types. Overall, these experiences suggest that, with such iBeacon devices, we should take these limitations and performance characteristics into consideration when designing applications and apply improvement schemes with respect to the intuitions collected from such real-world pilot deployment experiences.

In this case study, we extend the line of possible iBeacon applications by proposing an automatic attendance checker system that automatically checks the attendance of a college student to her classes in a university. However, simple proximity is not enough to accurately determine whether a student is inside the classroom or not since the beacon
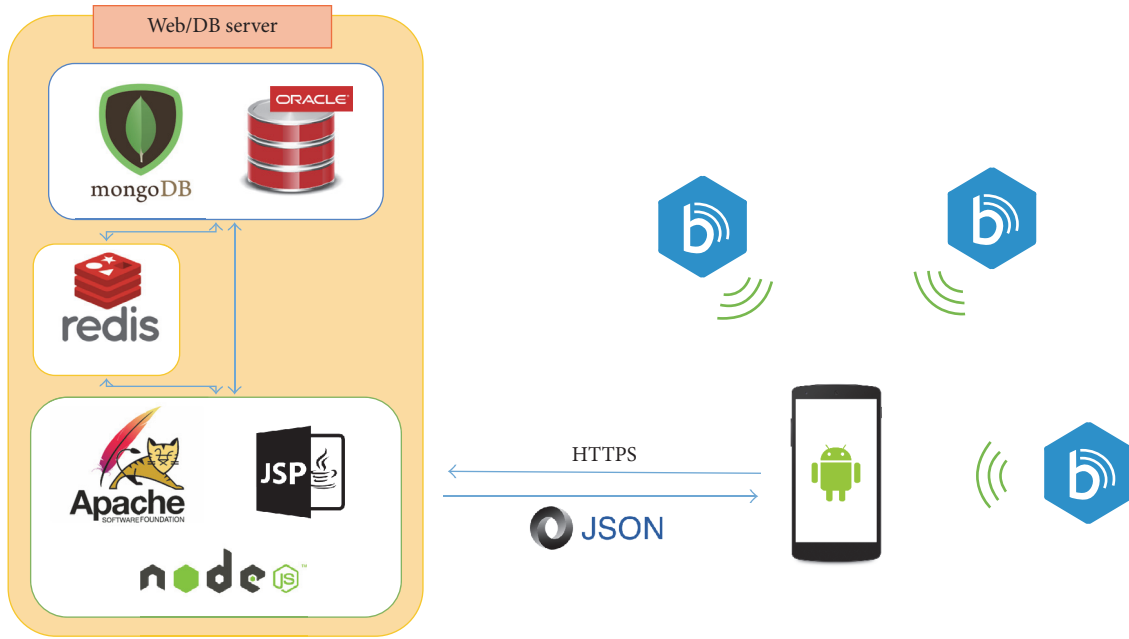
FIGURE 12: Overall system architecture of our iBeacon-based automatic attendance checker system.

signal can be received behind the walls of the classrooms as well. Thus, we use the trilateration-based position estimation to make a decision on whether the student is attending the class in the room or not. To compensate for RSSI reading errors due to unexpected obstacles (e.g., users holding their mobile phone tightly, or their phones being in a bag), we make simple yet novel geometric adjustments in trilateration calculation to improve the detection accuracy. On a system-level perspective, our system is composed of not only iBeacons and mobile devices, but also a database server which maintains and handles information about students, classes, classrooms, time-table, and attendance information. As we later show, the complete system allows for a fully automated attendance checking mechanism to take place with 100% accuracy under the circumstance that the students have enabled our application on their smartphones. Our system not only saves time wasted for manual attendance checking during classes, but also prevents attendance cheatings since it is very unlikely that a student will give his/her smartphone to a friend just for attendance checking.

5.1. System Architecture. Figure 12 shows the overview of our system architecture. Three iBeacons are deployed in each classroom with unique identification numbers (major-minor pair) for each iBeacon, and the relative *x-y* coordinate of each classroom is preconfigured. We developed an Android application for our system, which receives beacon signals in the background and communicates with the web/DB server over the Internet for attendance checking without user intervention. The server maintains all the necessary information to compute and confirm attendance in the database. The recorded attendance data can be viewed on our system's website by the faculty and staff members as well as on the students' mobile devices.

Specifically, we used an Oracle DB to maintain information regarding students, professors, classes, time tables, and, most importantly, deployed iBeacons including their *x-y* coordinates (relative, within each classroom) and identification numbers. This information is relatively static in the sense that student information changes only when they join and login for the first time, and class/time-table information changes only at the beginning of each semester for which we use web-crawling to automatically populate from the university website. To store the actual attendance information (the class, date, time, and *x-y* coordinate within classroom) for each student, we used the MongoDB to handle the frequently updated data. Figure 13 shows our database structure.

Using the aforementioned system, the automatic attendance check process operates as follows. When a student enters a classroom, the smartphone application will receive beacon advertisements from three or more iBeacons (there may be signals from nearby classrooms). At this point, the application sends the list of identification numbers (UUID, major/minor values [3]) and RSSI readings from the iBeacons to the server and queries the server for verification of attendance. Given this information, the server will first check the classroom information of each iBeacon and validates it against the classroom at which the student should be attending at the given time instance. If the classroom information indicates that the student is in a wrong classroom or if there are only two or less number of beacons detected from the target classroom, then the server returns FALSE for attendance. If no beacon is detected, attendance is FALSE by default. For all other cases, the server uses the relative *x-y* coordinates (within the target classroom) of the iBeacons to estimate a position of the mobile device and checks whether the device is within the boundaries of the classroom. If so,
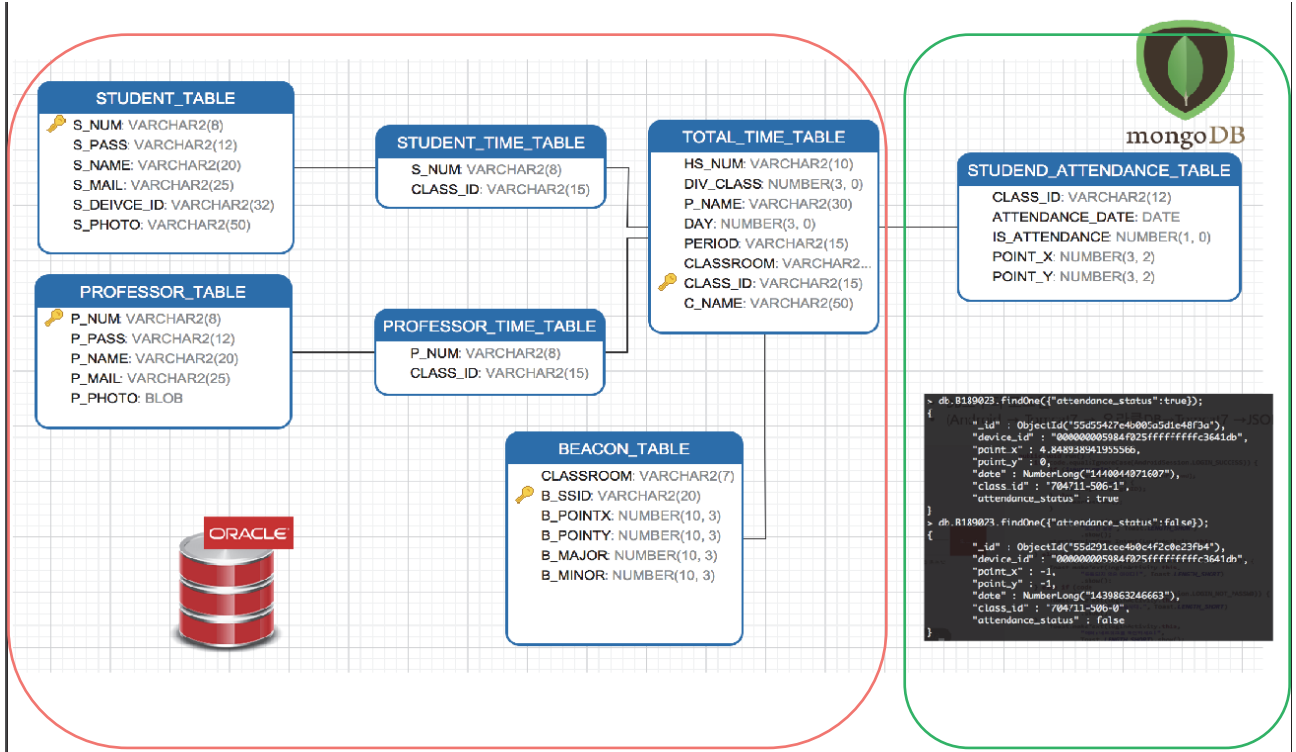
Figure 13: Databased structure in our iBeacon-based automatic attendance checker system server.

the server returns TRUE for attendance and records it in the database. This process is repeated periodically so that attendance can be checked during the class as well. This is to account for students who are a few minutes late, as well as those students who leave early during class. However, to reduce the unnecessary power consumption from the use of BLE for the duration when a student does not have a class to attend, her time-table can be locally cached on the mobile device to enable attendance checking only when needed.

Note that the *x-y* coordinates stored in the database are only relative and *local* to the target classroom. This is in contrast to other indoor positioning systems where global *x-y* coordinate within the entire target area (e.g., whole floor of the building to map) is required and is made possible given that we maintain preknowledge on which classroom a student should be at a given time. This greatly simplifies the system architecture not only in terms of estimating the position of the mobile device, but also in terms of iBeacon deployment, replacement, and reconfiguration. In other words, if systems are mostly concerned for a small geographical region, iBeacon placements can be independently (and better) customized and optimized. We provide detailed descriptions on such calibration procedures in the following section.

5.2. Geometric Adjustment for Improved Accuracy. The main challenge in using iBeacon signals for accurate indoor positioning is the variability of RSSI readings and their sensitivity to environment changes such as obstacles or user handling of the mobile device. Our findings earlier show

that RSSI values can drop significantly when a beacon signal is received behind an obstacle, and the amount of signal attenuation varies depending on the obstacle types (e.g., window, wooden/metal door, walls, etc.). Furthermore, the height of the mobile device from the ground also affects the RSSI (height of the iBeacon also introduces a high impact, but iBeacons are usually wall-mounted and its height is fixed in typical scenarios). More significantly, the signal attenuation due to holding the mobile device tightly in users' hands (which happens often) can be as much as 30 dBm. Such a high variation is a significant challenge when using trilateration for position estimation. However, we see one commonality from the aforementioned cases; the signal attenuates (RSSI is lower than the model for a given distance), and it is extremely rare to see higher RSSI than the (best-case) line-of-sight environment. Using this intuition, we use a simple yet novel geometric adjustment scheme to improve accuracy rather than trying to calibrate the model based on highly varying RSSI values.

For example, Figure 14 shows one example of an exception case where the distance estimation from three iBeacons results in three RSSI-coverage distance circles (circles drawn by the estimated distance as radii for trilateration) without any intersection; one circle is completely enclosed by the biggest circle, and a third circle is completely outside of the biggest one. In this case, standard trilateration calculation is not possible. However, our intuition is that the distance of the larger circle (larger distance from higher RSSI value) is closer to the estimation model, and the signals of the two smaller circles have been attenuated due to some reason
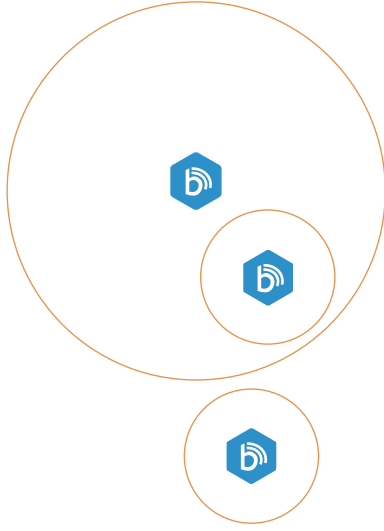
FIGURE 14: An example of possible exception scenario where trilateration calculation will fail due to high variation in RSSI reading, thus causing error.

Original distance circle
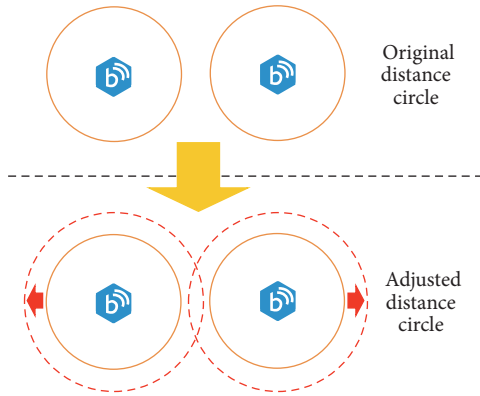
Adjusted distance circle

FIGURE 15: [Case 1] where two distance circles (circles representing the estimated distance using RSSI) from two iBeacons are nonintersecting because the sum of two distance estimations are smaller than the distance between the two iBeacons. This results in no intersection point for trilateration calculation. Thus, we gradually increase the sizes of the circles until they intersect.

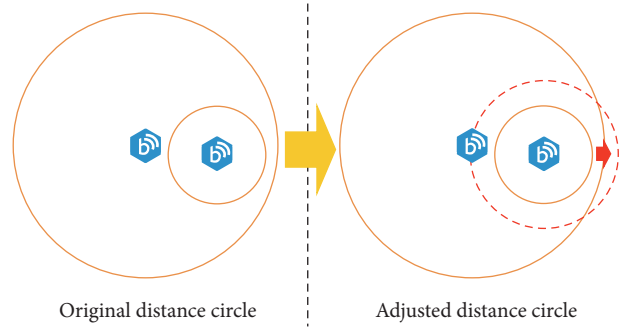Original distance circle   Adjusted distance circle

FIGURE 16: [Case 2] where a distance circle (a circle representing the estimated distance using RSSI) from one iBeacon is completely enclosed by another distance circle (from another iBeacon), resulting in no intersection point for trilateration calculation. In this case, we gradually increase the size of the inner circle until the two circles intersect.
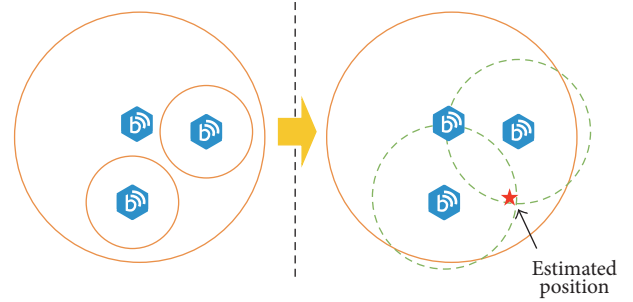

Estimated position

FIGURE 17: An example of position estimation from an exception case where the original distance circles (from distances estimated by RSSI) had two circles completely enclosed in a third circle.

such as obstacles. Based on this intuition, our approach is as follows. For each pair of circles (out of three pairs from three iBeacons), we first check whether [Case 1] two circles have no intersection points (upper figure of Figure 15) or whether [Case 2] one circle is completely enclosed by another circle (left figure of Figure 16). If neither is true, then there is nothing else to consider for that specific pair of RSSI-coverage distance circles.

If a pair of circles have no intersection points (i.e., [Case 1]), we first increase the size of the smaller circle with an increment of 1 meter until there are two intersection points or up to the point where the two circles become the same sizes. If no intersection points are created even after two circles are of the same sizes, then we increase the size of *both* circles

with an increment of 1 meter until there are two intersection points (c.f. Figure 15). Again, this adjustment comes from the assumption that the indoor RSSI levels are disrupted by obstacles, which make the RSSI levels degrade more rapidly.

In the second case where, for a pair of circles, one circle is completely enclosed by another circle (i.e., [Case 2]), we increase the size of the smaller circle with an increment of 1 meter until there are two intersection points (c.f. Figure 16). The resulting circles after the adjustments represent the *"adjusted distances"* from each iBeacon. Now, since all pairs of *"adjusted circles"* each have two intersection points, we can use these six intersection points to apply the trilateration and estimate the approximate position of the mobile device.

Figures 17 and 18 show the position estimation results from the two exception cases where the estimated distances from three iBeacons did not provide sufficient coverage for trilateration. After the geometric adjustment based on our prior observations, the resulting estimations were detected to be very close to the actual positions. As a final note, the entire position estimation process, including the geometric adjustment and trilateration, takes place at the server. The mobile device simply detects iBeacon signals during the period of classes for the student, sends the list of iBeacon
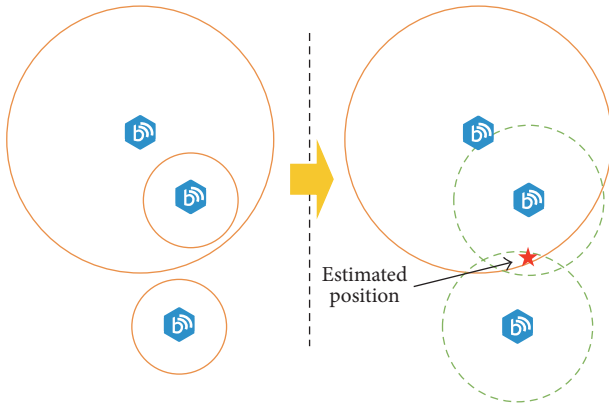
FIGURE 18: An example of position estimation from the exception case in Figure 14 where the original distance circles (from distances estimated by RSSI) had one circle enclosed in another and a third circle is nonintersecting with the other two circles.

advertisement data to the server, and queries for attendance check. Thus, there is practically no computation burden on the mobile device, and this leads to minimizing the energy usage as well.

5.3. Evaluation. We have evaluated our automatic attendance checker system with the help of four undergraduate students attending classes in three classrooms. While this case study was done in a limited scale due to challenges in getting individual consent for accessing students personal information, the system reported no false reports (neither false positive nor false negatives for attendance) for our student volunteers during the course of 1 month. We plan to scale the experiment to a larger number of students and classes in the future.

6. Related Work

There are several pieces of prior work that proposes applications and services using iBeacon devices. The work in [9] uses iBeacons for tracking luggage at the airport, and the work in [10] provides interactive experience to visitors in museums. iBeacon has been used for path planning in emergency guiding system[11] and also for tracking patients in emergency rooms [12]. Furthermore, the work in [13] proposes an indoor route guidance system, and the work in [15, 16] uses iBeacons for occupancy detection within buildings. These applications all use proximity-based on RSSI as the source of information rather than trying to pinpoint the exact and absolute location of the mobile device. There are also a number of prior works that focus on using iBeacon devices for precise indoor positioning [6, 7, 12, 23, 24]. However, none of these efforts provide an in-depth and extensive measurement study of iBeacon RSSI measurements and its variability with different environmental factors.

The work in [30] implements an Android application that collects statistics of RSSI values from nearby iBeacons and provides some measurement results but only at a fixed distance. The work in [5] provides some measurement

data regarding the transmission power and reception ratio along with the estimated distance, and the work in [31] attempts to calibrate the distance estimation model using measurement data and perform error analysis. However, none of these works provide extensive and comprehensive measurement data in the dimensions of several different types of iBeacons, mobile devices software platforms, transmission power configurations, indoor/outdoor comparisons, and WiFi interference impacts and with several different types of controlled yet practical obstacles. Furthermore, we combine the experiences from the RSSI measurements to design a well-suited and practically applicable system and propose a case study application for automatic attendance checking, which uses localized trilateration techniques along with geometric adjustments to limit the scope of error due to RSSI variability and meet the target accuracy.

7. Conclusion

The original goal of this research was to develop an improved indoor positioning system using iBeacon technologies. However, after numerous and extensive experiments, we realized that the signal variation was too high to retrieve accurate distance estimation for designing a reliable and robust localization system. Instead we decided to report on this variation and inaccuracy to clarify the misunderstanding caused by numerous sugar-coated news articles on how accurate iBeacon technology can be. Our finding shows that iBeacon RSSI values (and the corresponding signal propagation model to estimate the distance) vary significantly across iBeacon vendors, mobile device platforms, deployment height of the device, indoor/outdoor environmental factors, and obstacles. It is obvious that the accuracy and efficiency of location estimation depend heavily on the accuracy of the measured RSSI measurements and the model used to estimate the distance, not to mention the surrounding environmental factors. We believe that our work provides evidence on the challenges for designing an indoor localization system using commercial-off-the-shelf (COTS) iBeacons devices and dismantles the misunderstanding of its overestimated accuracy. Furthermore, based on the observations made in this work, our future work is to find a way to approach these errors differently and develop an iBeacon-based system that is resilient and robust to such RSSI dynamics.

Competing Interests

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

Acknowledgments

References

[1] "Bluetooth Smart," https://www.bluetooth.com/what-is-bluetooth-technology/bluetooth-technology-basics/low-energy.

[2] C. Gomez, J. Oller, and J. Paradells, "Overview and evaluation of bluetooth low energy: an emerging low-power wireless technology," *Sensors*, vol. 12, no. 9, pp. 11734–11753, 2012.

[3] Apple, "iBeacon for developers," https://developer.apple.com/ibeacon/.

[4] Google, "Eddystone-open beacon format," https://developers.google.com/beacons/.

[5] P. Martin, B.-J. Ho, N. Grupen, S. Muñoz, and M. Srivastava, "An ibeacon primer for indoor localization: demo abstract," in *Proceedings of the ACM Conference on Embedded Systems for Energy-Efficient Buildings (BuildSys '14)*, 2014.

[6] Z. Chen, Q. Zhu, H. Jiang, and Y. C. Soh, "Indoor localization using smartphone sensors and iBeacons," in *Proceedings of the IEEE 10th Conference on Industrial Electronics and Applications (ICIEA '15)*, pp. 1723–1728, IEEE, Auckland, New Zealand, June 2015.

[7] H. K. Fard, Y. Chen, and K. K. Son, "Indoor positioning of mobile devices with agile iBeacon deployment," in *Proceedings of the IEEE 28th Canadian Conference on Electrical and Computer Engineering (CCECE '15)*, pp. 275–279, Halifax, Canada, May 2015.

[8] mlb.com, "MLBAM completes initial iBeacon installations," http://m.mlb.com/news/article/67782508/mlb-advanced-media-completes-initial-ibeacon-installations.

[9] M. Kouhne and J. Sieck, "Location-based services with ibeacon technology," in *Proceedings of the 2nd IEEE International Conference on Artificial Intelligence, Modelling, and Simulation (AIMS '14)*, pp. 315–321, IEEE, Madrid, Spain, November 2014.

[10] Z. He, B. Cui, W. Zhou, and S. Yokoi, "A proposal of interaction system between visitor and collection in museum hall by iBeacon," in *Proceedings of the 10th International Conference on Computer Science and Education (ICCSE '15)*, pp. 427–430, Cambridge, UK, July 2015.

[11] L.-W. Chen, J.-J. Chung, and J.-X. Liu, "GoFAST: a group-based emergency guiding system with dedicated path planning for mobile users using smartphones," in *Proceedings of the IEEE 12th International Conference on Mobile Ad Hoc and Sensor Systems (MASS '15)*, pp. 467–468, IEEE, Dallas, Tex, USA, October 2015.

[12] X.-Y. Lin, T.-W. Ho, C.-C. Fang, Z.-S. Yen, B.-J. Yang, and F. Lai, "A mobile indoor positioning system based on iBeacon technology," in *Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '15)*, pp. 4970–4973, IEEE, Milan, Italy, August 2015.

[13] A. Fujihara and T. Yanagizawa, "Proposing an extended ibeacon system for indoor route guidance," in *Proceedings of the International Conference on Intelligent Networking and Collaborative Systems (INCOS '15)*, pp. 31–37, Taipei, Taiwan, September 2015.

[14] R.-S. Cheng, W.-J. Hong, J.-S. Wang, and K. W. Lin, "Seamless guidance system combining GPS, BLE Beacon, and NFC technologies," *Mobile Information Systems*, vol. 2016, Article ID 5032365, 12 pages, 2016.

[15] G. Conte, M. De Marchi, A. A. Nacci, V. Rana, and D. Sciuto, "Bluesentinel: a first approach using ibeacon for an energy efficient occupancy detection system," in *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings (BuildSys '14)*, pp. 11–19, Memphis, Tenn, USA, November 2014.

[16] A. Corna, L. Fontana, A. Nacci, and D. Sciuto, "Occupancy detection via ibeacon on android devices for smart building management," in *Proceedings of the Design, Automation Test in Europe Conference Exhibition (DATE '15)*, March 2015.

[17] A. LaMarca, Y. Chawathe, S. Consolvo et al., "Place lab: device positioning using radio beacons in the wild," in *Proceedings of the 3rd International Conference on Pervasive Computing (PERVASIVE '05)*, 2005.

[18] P. Bahl and V. N. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system," in *Proceedings of the INFOCOM*, pp. 775–784.

[19] Y. Chen, D. Lymberopoulos, J. Liu, and B. Priyantha, "FM-based indoor localization," in *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services (MobiSys '12)*, 2012.

[20] J. Brassil, C. Pearson, and L. Fuller, "Indoor positioning with an enterprise radio access network," *Procedia Computer Science*, vol. 34, pp. 313–322, 2014.

[21] A. Varshavsky, E. de Lara, J. Hightower, A. LaMarca, and V. Otsason, "GSM indoor localization," *Pervasive and Mobile Computing*, vol. 3, no. 6, pp. 698–720, 2007.

[22] J. Paek, K.-H. Kim, J. P. Singh, and R. Govindan, "Energy-efficient positioning for smartphones using cell-ID sequence matching," in *Proceedings of the 9th ACM International Conference on Mobile Systems (MobiSys '11)*, pp. 293–306, Washington, DC, USA, June 2011.

[23] Y. Wang, X. Yang, Y. Zhao, Y. Liu, and L. Cuthbert, "Bluetooth positioning using RSSI and triangulation methods," in *Proceedings of the IEEE 10th Consumer Communications and Networking Conference (CCNC '13)*, pp. 837–842, IEEE, Las Vegas, Nev, USA, January 2013.

[24] P. Kriz, F. Maly, and T. Kozel, "Improving indoor localization using bluetooth low energy beacons," *Mobile Information Systems*, vol. 2016, Article ID 2083094, 11 pages, 2016.

[25] S.-H. Lee, I.-K. Lim, and J.-K. Lee, "Method for improving indoor positioning accuracy using extended Kalman filter," *Mobile Information Systems*, vol. 2016, Article ID 2369103, 15 pages, 2016.

[26] Estimote beacon, http://www.estimote.com.

[27] Wizturn beacon, http://www.lime-i.com/beacon/pebble/en/.

[28] "GELO beacon," http://www.getgelo.com/.

[29] J. Zhao and R. Govindan, "Understanding packet delivery performance in dense wireless sensor networks," in *Proceedings of the ACM 1st International Conference on Embedded Networked Sensor Systems (SenSys '03)*, Los Angeles, Calif, USA, November 2003.

[30] M. Varsamou and T. Antonakopoulos, "A bluetooth smart analyzer in iBeacon networks," in *Proceedings of the 4th IEEE International Conference on Consumer Electronics (ICCE '14)*, pp. 288–292, IEEE, Berlin, Germany, September 2014.

[31] T. M. Ng, "From 'where i am' to 'here i am': accuracy study on location-based services with iBeacon technology," *HKIE Transactions Hong Kong Institution of Engineers*, vol. 22, no. 1, pp. 23–31, 2015.

Research Article
A Fast Self-Planning Approach for Fractional Uplink Power Control Parameters in LTE Networks

J. A. Fernández-Segovia,[1] S. Luna-Ramírez,[1] M. Toril,[1] and C. Úbeda[2]

[1]*Ingeniería de Comunicaciones, Universidad de Málaga, Campus de Teatinos S/N, 29071 Málaga, Spain*
[2]*Ericsson, Calle de la Retama 1, 28045 Madrid, Spain*

Correspondence should be addressed to J. A. Fernández-Segovia; jfs@ic.uma.es

Uplink Power Control (ULPC) is a key feature of mobile networks. Particularly, in LTE, Physical Uplink Shared Channel (PUSCH) performance strongly depends on Uplink Power Control configuration. In this work, a methodology for the self-planning of uplink Fractional Power Control (FPC) settings is presented. Values for nominal power and channel path-loss compensation factor are proposed. The method is designed for the planning and operational (replanning) stages. A very fast solution for FPC setting can be achieved by the combination of several simple solutions obtained by assuming some simplifications. First, the FPC planning problem is formulated analytically on a cell basis through the combination of multiple regular scenarios built on a per-adjacency basis from a live scenario. Secondly, detailed inspection of the FPC parameter values aims to identify the most important variables in the scenario impacting optimal FPC settings. Finally, regression equations can be built based on those key variables for a simple FPC parameter calculation, so computational costs are extremely reduced. Results show that network performance with the proposed FPC parameter settings is good when compared with typical FPC configurations from operators.

1. Introduction

Mobile Communication Networks have experienced strong evolution in the last years. The development of new radio access technologies has increased network capacity and quality significantly, especially with the UMTS Long-Term Evolution (LTE) [1]. Simultaneously, the appearance of the so-called *smartphones* has changed the traffic behavior carried by mobile networks, where data transmission (and not voice calls) is the traffic benchmark, and, as a consequence, data transmission enhancement has been the main focus in present networks [2]. These factors have strengthened the role of network planning when there is a desire to improve overall network performance. Before the network deployment stage, network planning aims to get the best network performance in a concrete scenario. Trade-off between network capacity and coverage is the most limiting factor for network planning [3].

Regardless of the network radio access technology, proper network planning allows the operator to identify key network areas, which eases proper network dimensioning and enables the prediction of future bottlenecks. Thus, network planning is useful to avoid or, at least, delay subsequent capital investments [4, 5]. The growth of application data traffic has led to changes in network planning approaches trying to predict how the user experience is. Whereas former approaches have focused on network performance indicators, user-centric statistics are now the preference (e.g., average and cell-edge user throughput) [6].

Cellular network planning can be divided depending on the network system to be planned: core and Radio Access Network (RAN). While the core network planning relies mainly on dimensioning processes, RAN planning comprises radio dimensioning and radio parameter configuration [3, 7, 8]. Network operators do not usually take advantage of radio parameter configuration due to the inherent complexity of finding optimal parameter settings for every cell in the network. Thus, operators usually set radio parameters to some network-wide values recommended by vendors, which work reasonably well in most cases, but some additional network improvement is discarded. To revert this situation, the

3rd Generation Partnership Project (3GPP) has defined the requirements for the automation of planning, optimization, and self-healing in mobile networks [9]. As a result, network self-planning has been identified as an important process in Self-Organizing Networks (SON) now [5, 10, 11] and in the future [12].

Power Control (PC) is one of the most impacting algorithms in network performance. Fractional Power Control (FPC) has been selected for the Physical Uplink Shared Channel (PUSCH) in LTE [13, 14]. Consequently, FPC algorithm controls LTE uplink performance, which makes its configuration an ideal candidate for self-planning purposes. In fact, the variability of radio conditions such as propagation losses and interference level makes it difficult to set an optimal value for FPC parameters. For this reason, network operators use safe network-wide recommended values. As a consequence, suboptimal performance is achieved by the network. Hence, even if this can be solved in the operational stage, the provision of proper initial FPC settings would be valuable for network operators. Additionally, the research of low computational complexity methodologies for FPC self-planning is of high interest in the development of SON algorithms.

2. Related Work and Contribution

There is a wide background regarding FPC performance in LTE uplink. First, a performance analysis of open-loop FPC is presented in [15, 16], whereas closed-loop behavior was analyzed in [17–19]. Moreover, more sophisticated Power Control schemes for LTE were assessed in later studies [20, 21]. In those schemes, interference or load conditions were taken into consideration.

A sensitivity analysis of FPC parameters is performed in [22]. The analysis relies on system-level simulations and the results suggest a suboptimal parameter configuration for noise-limited and interference-limited scenarios. Obviously, the overall problem solution is not as simple, since it is a nonseparable and nonlinear large-scale optimization problem. However, it is the start point in the search of more complex solutions. In [23], the FPC parameter settings problem in a single cell is formulated as a classical optimization problem, where average user throughput and cell throughput are taken as figures of merit for the optimization process. An extension of this analysis to a multicell scenario is done in [24] by formulating FPC as a noncooperative game model where a heuristic iterative optimization algorithm solves the problem. More conscientiously, a self-planning method for selecting the best parameter settings in FPC on a per-cell basis in an irregular LTE scenario is proposed in [25]. It is based on an exhaustive search approach using Taguchi's method over a system-level simulator. There are other studies considering tuning algorithms for the network operational stage. These self-tuning algorithms, which have been conceived for the operational stage, can also be applied in the planning stage, provided that a system model is available (e.g., a system-level simulator). For instance, a self-tuning algorithm is proposed in [26] to control interference by performing dynamical adjustment of nominal power parameter based on the overload indicator [27]. Likewise, a self-tuning algorithm for FPC is proposed in [28] based on fuzzy-reinforcement learning techniques. Most of these self-tuning algorithms need iterative evaluation of the system model for many different parameter settings, thus emulating the realistic network behavior. As a consequence, this iterative process is adequate for live networks, where performance measurements are provided. However, this is not the case of network planning, where computational cost increases with the complexity of the implemented system model. For this reason, most existing FPC planning methods rely on simple analytical models, which eases scalability and performance assessment.

A more computationally efficient planning method is presented in [29]. The method relies on an analytical model that predicts the influence of the nominal power and path-loss compensation factor on call acceptance probability for a previously defined spatial user distribution. A suboptimal solution for these parameters is computed by a local greedy search algorithm. In the same way, a computationally efficient method for self-planning Uplink Power Control parameters in LTE is presented in [30]. This method proposes a heuristic algorithm that can handle irregular scenarios at a low computational complexity. For this purpose, the parameter planning problem in a cell is formulated analytically through the combination of multiple regular scenarios built on a per-adjacency basis. However, in [30], Nonfractional Power Control is considered, assuming total propagation losses compensation.

To the authors' knowledge, few of the previous references handle irregular scenarios at a low computational cost and none of them propose some simple model with the aim of getting near-optimal FPC parameter values depending on scenario details.

In this paper, a fast method for the self-planning of FPC parameters in LTE uplink is proposed. The self-planning method determines the nominal PUSCH power, P_0, and the path-loss compensation, α, parameters in FPC. Similar to [30], to deal with scenario irregularities, the parameter setting problem is solved by the aggregation of multiple scenarios defined on a per-adjacency basis. Moreover, with the aim of minimizing computational complexity, solutions provided by the self-planning method are further analyzed and a simple model for the estimation of FPC parameter values is proposed.

Unlike [30], the decision variables in this work are P_0 and α, instead of P_0 and uplink cell load, U_{UL}. The approach in [30] is suitable for the planning stage, when performance measurements (PMs) are not available and the maximum cell loads are still design variables. However, it is limited to some first vendor releases, where α was a system constant ($\alpha = 1$). In contrast, the approach proposed here is conceived for the operational stage, when input parameters can be taken from network PMs. Thus, U_{UL} is an input parameter taken from statistics of Physical Resource Block (PRB) utilization ratio in the network management system. Likewise, FPC is considered here ($\alpha \leq 1$). Moreover, the optimization criterion is different from that used in [30].

The main contributions of this work are (a) a sensitivity analysis of FPC parameter solutions in a realistic network

implemented over a system-level simulator, (b) a thorough analysis of how FPC parameter values are related to LTE scenario characteristics and the identification of the most significant scenario parameters affecting FPC setting, and (c) a highly computationally efficient methodology to configure FPC parameters. The rest of the paper is organized as follows. In Section 3, the system model is outlined. The self-planning algorithm is presented in Section 4. Performance assessment is carried out in Section 5. Finally, concluding remarks are given in Section 6.

3. Fractional Power Control in LTE Uplink

Three physical channels are defined for LTE uplink: Physical Random Access Channel (PRACH), Physical Uplink Shared Channel (PUSCH), and Physical Uplink Control Channel (PUCCH) [31]. Attending to LTE standards, Uplink Power Control feature applies to PUSCH and PUCCH [14]. Specifically, PUSCH is used to transmit user data and control information for active users. Uplink Power Control (ULPC) behavior for PUSCH is defined as

$$
P_{\text{tx}} = \min \left\{ P_{\text{tx}_{\max}}, \underbrace{P_0 + \alpha \cdot \text{PL}}_{\substack{\text{basic open-loop} \\ \text{operating point}}} + \underbrace{\Delta_{\text{TF}} + f\left(\Delta_{\text{TPC}}\right)}_{\text{dynamic offset}} \right.
$$

$$
\left. + \underbrace{10 \cdot \log_{10} M_{\text{PUSCH}}}_{\text{bandwidth factor}} \right\},
\tag{1}
$$

where $P_{\text{tx}_{\max}}$ is the maximum User Equipment (UE) transmit power, α is the channel path-loss compensation factor, PL are the propagation losses, M_{PUSCH} is the number of PRBs assigned to the UE, and $\Delta_{\text{TF}} + f(\Delta_{\text{TPC}})$ is a dynamic term that depends on the selected modulation scheme and power control commands sent by the eNodeB (eNB).

As shown in (1), transmit power depends on three terms: the basic open-loop operating point, dynamic offset which represents closed-loop corrections, and a multiplicative factor depending on the bandwidth. It must be noted that, in open-loop behavior, the parameter α ($\alpha \in \{0, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$) represents the fraction of PL which are compensated by the UE to guarantee the nominal PUSCH power, P_0. Thus, when path-loss compensation factor is different from one, ULPC is known as Fractional Power Control (FPC).

In this work, the system model is the same as the one used in [30]; that is, uplink Signal over Interference and Noise Ratio (SINR) is based in the emulation of the uplink scheduler proposed there.

4. Self-Planning Algorithm

In this section, a self-planning methodology for FPC parameters, namely, P_0 and α, is described. General considerations regarding the algorithm are first explained in Section 4.1. The algorithm operation for regular scenarios is described in Section 4.2, and the algorithm extension for irregular

scenario is performed in Section 4.3. Finally, a detailed analysis of FPC solutions is approached in Section 4.4, with the aim of building a multivariate linear regression model with the most significant scenario parameters.

4.1. General Consideration. Mobile network performance is usually experienced as a trade-off between coverage and capacity so both characteristics cannot be optimized separately. This trade-off is known as the Coverage and Capacity Optimization (CCO) SON use case defined by 3GPP in [32]. Network coverage and capacity are usually measured with cell-edge user and cell-average throughput statistics, respectively [6, 33].

Self-planning of FPC parameter is a particular way to approach the CCO problem. On the one hand, changing P_0 in a cell i impacts coverage and capacity of cell i and its surrounding neighbors. High P_0 values force the UEs connected to cell i to transmit with higher power, increasing interference in adjacent cells. However, received signal in cell i is higher, and, thus, SINR (and, as a consequence, cell throughput) is increased. Conversely, low P_0 values decrease transmit power for UEs in the cell, reducing interference in adjacent cells, which favors coverage of surrounding cells at the expense of reducing coverage and capacity of the considered cell. Regarding α parameter, different α settings impact similarly the UE transmit power, so it can be also used to manage interference between cells.

In any case, P_0 and α best settings are both very influenced by the particular topology and radio propagation conditions in the network scenario. This is especially important when irregular scenarios (which are majority) are considered. A cell-based FPC configuration can reach the best network performance by adapting P_0 and α settings to every cell environment. As a consequence, the resulting CCO problem is a nonseparable multivariable optimization problem in which all cells are jointly optimized. In other words, the solution space is $(N_{v_{P_0}} \cdot N_{v_\alpha})^{N_c}$, where $N_{v_{P_0}}$ and N_{v_α} are the number of possible values for P_0 and α parameters and N_c is the number of cells to be planned. The large size of the solution space prevents the use of exact algorithms, which are substituted by heuristic algorithms, for example, Taguchi's method [25], greedy search [29], or simulated annealing.

In this work, the methodology described in [30] is reproduced to reduce the algorithm search space. The global multivariate optimization problem is divided into N_c independent bivariate subproblems. The following subsections describe the optimization process in a regular scenario and then the extension to irregular scenarios and later an analysis of the algorithm solution that allows optimizing computational complexity by a regression model.

4.2. Regular Scenario. To design the self-planning algorithm, a sensitivity analysis of FPC parameters is carried out over a simple regular scenario. This regular scenario consists in seven trisectorial sites, specifically one central site surrounded by six adjacent sites, as shown in Figure 1. In such scenario, FPC parameters P_0 and α are configured uniformly in all cells. Then, parameters are separately swept
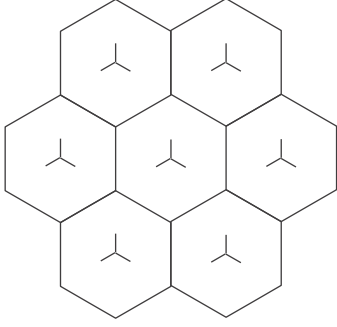
FIGURE 1: Regular scenario used in FPC parameter sensitivity analysis.

TABLE 1: Simulation parameters.

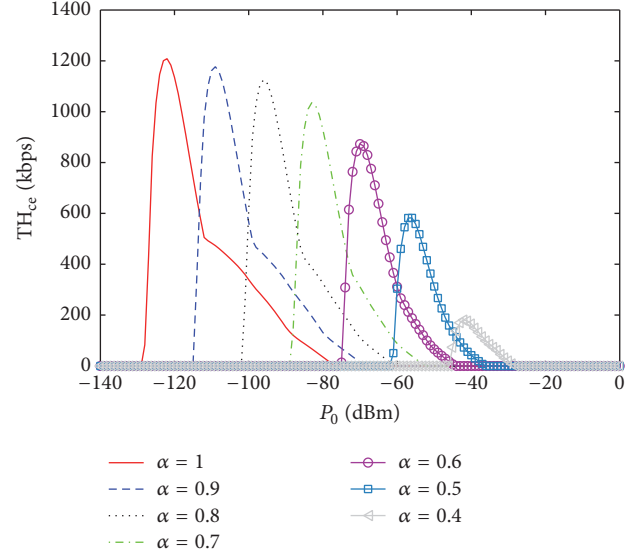| Parameters	Settings
Carrier bandwidth	10 MHz (50 PRBs)
Carrier frequency	2 GHz
Cell layout	7 eNBs, 21 sectors, regular grid (40 m resolution)
Distance attenuation	COST 231 [36]
Thermal noise density	−174 dBm/Hz
Cell radius	1.5 km (3 km intersite distance)
eNB antenna height	30 m
eNB antenna tilt	5°
eNB antenna pattern and gain	3GPP ideal [37]
Max. UE transmit power	23 dBm
Path-loss compensation factor, α	(0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1)
Cell load	100%
Traffic model	Full buffer

FIGURE 2: Cell-edge throughput performance in regular scenarios.
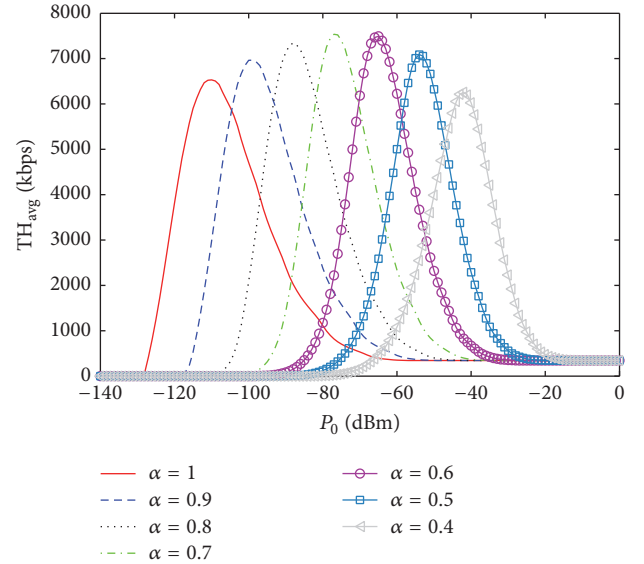
FIGURE 3: Cell-average user throughput performance in regular scenarios.

and coverage and capacity indicators are measured in one cell of the centre site. The rest of the simulation parameters are shown in Table 1.

Figures 2 and 3 show cell-edge and cell-average user throughput for cell i, $\text{TH}_{ce}(i)$ and $\text{TH}_{avg}(i)$, respectively, with different P_0 and α settings. Every curve in Figure 2 shows a similar behavior. With low P_0 values, increasing P_0 leads to an improvement in TH_{ce}. After that initial improvement, additional P_0 increases lead to TH_{ce} decreases due to intercell interference issues. Thus, an optimal P_0 value, $P_{0,opt}^{(ce)}$, can be defined (i.e., TH_{ce} is maximum for that $P_{0,opt}^{(ce)}$ value). A similar behavior is observed for TH_{avg} performance in Figure 3. Optimal P_0 has a different value when TH_{avg} performance is considered (i.e., $P_{0,opt}^{(ce)} \neq P_{0,opt}^{(avg)}$). On the other

hand, decreasing α has a high impact on coverage and capacity figures. Maximum TH_{ce} value decreases with α due to uncompensated losses, whereas maximum TH_{avg} value increases due to the interference decreases for lower α values. Moreover, curves in Figures 2 and 3 are displaced to the right due to path-loss compensation impact over UE transmit power in (1). Additionally, the difference between optimal P_0 values, $P_{0,opt}^{(avg)}$-$P_{0,opt}^{(ce)}$, is decreased for lower α values (and, thus, the best coverage and capacity performance could be reached simultaneously).

Based on the behavior observed in Figures 2 and 3, a self-planning algorithm for P_0 and α parameters in a regular scenario is designed. The proposed algorithm finds the optimal P_0 and α settings by a simple gradient search
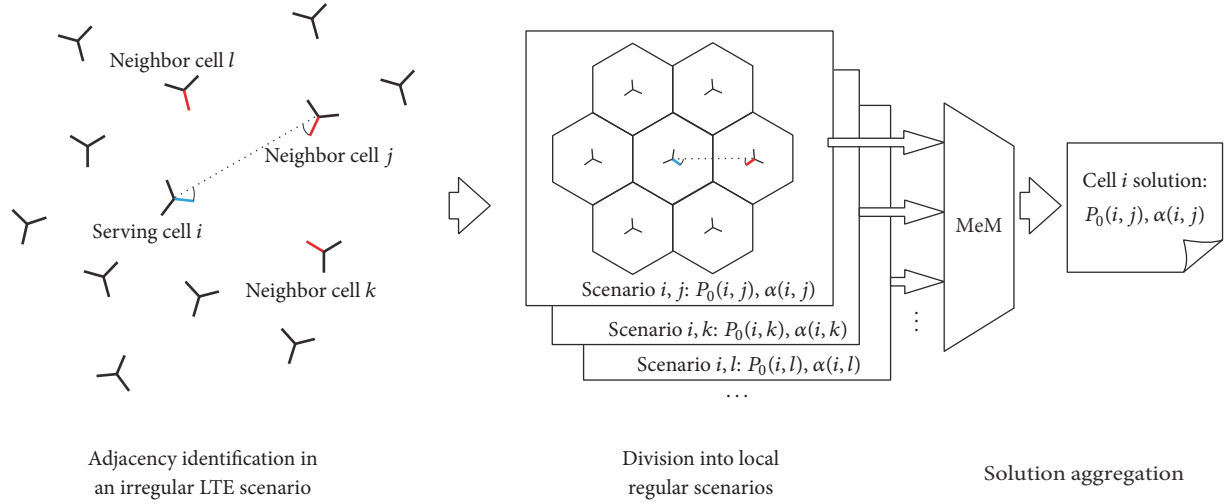
FIGURE 4: Optimization process for irregular scenarios.

which maximizes TH_{ce} and TH_{avg} according to a *trade-off* value, t, as

$$\max_{(P_0, \alpha)} \left\{ t \cdot \text{TH}_{\text{ce}}(i) + (1 - t) \cdot \text{TH}_{\text{avg}}(i) \right\}, \quad t \in [0, 1]. \quad (2)$$

Note that in regular scenarios all cells are identically configured (i.e., $P_0(i) = P_0(j)$ and $\alpha(i) = \alpha(j)$, $\forall i, j$).

4.3. Irregular Scenario. The previous section proposed a method for finding an optimal solution for FPC parameter settings in a regular scenario. However, mobile networks operate in realistic scenarios (where cells are not regular) with very different user and radio conditions along cells. Additionally, cell performance is also affected by the also irregular adjacent cells. Hence, these irregularities have to be considered to plan FPC parameters. In this work, irregularities are approached similarly to the algorithm presented in [30]. Figure 4 depicts this process in three different stages.

In the first stage, *the identification of adjacent cells in an irregular LTE scenario* is performed. To identify the most relevant adjacent cells for cell i, $N_{\text{adj}}(i)$, neighbors are sorted following the indicator NR_{UL}, defined as

$$\text{NR}_{\text{UL}}(i, j) = L(i, j) - A_H(i, j) - A_V(i, j), \quad (3)$$

where $L(i, j)$ is the path-loss between the cell under study i and cell j and $A_h(i, j)$ and $A_v(i, j)$ are the horizontal and vertical gains for the antenna of cell i to the location of eNB j. First $N_{\text{adj}}(i)$ cells in the list are selected as adjacent cells.

In the second stage, *division into several local regular scenarios* is done. One regular scenario is built per serving cell i and neighbor cell j, based on the relative geometry between both cells. Relative antenna angles, antenna tilt, and height of both cells are kept from the irregular scenario. To compute interference levels in each regular scenario, it is assumed that all cells have the same uplink cell load, equal to that of the serving cell, $U_{\text{UL}}(i)$. Recall that $U_{\text{UL}}(i)$ is an input parameter taken from network measurements. Then, neighbor cell is replicated to complete the first tier of adjacent cells resulting in a regular scenario as that shown in Figure 1. $N_{\text{adj}}(i)$ regular scenarios are built per cell. Equation (2) is solved to find optimal P_0 and α for each scenario. Note that every regular scenario has different network parameters and has to be simulated (and solved) separately. At the end of this process, there are $N_{\text{adj}}(i)$ solutions (i.e., $P_0(i, j)$ and $\alpha(i, j)$ value for every j cell being adjacent to cell i), so some final criteria must be defined to obtain the final $P_0(i)$ and $\alpha(i)$ values.

Finally, in the third stage, the *aggregation of solutions* obtained from every regular scenario is carried out. As a result, a unique solution of the pair of parameters $P_0(i)$ and $\alpha(i)$ is given for each cell i. Different aggregation methods were tested in [30], although the so-called Medium Method (MeM) achieved better trade-off results. MeM calculates $P_0(i)$ and $\alpha(i)$ as the average value of all $P_0(i, j)$ and $\alpha(i, j)$ solutions. Hence, MeM is selected as the aggregation method in this work.

4.4. Regression Model for FPC Solutions. A self-planning algorithm for FPC parameters has been presented in previous sections. The proposed algorithm reduces the complexity of exhaustive search by breaking down the problem into many regular scenarios. Nonetheless, each of these regular scenarios has still to be simulated and solved. Thus, the computational complexity of the method can be reduced by deriving a simple analytical formula giving the optimal value of FPC parameters in terms of network parameters obtained from simulations. In this section, a thorough analysis of FPC optimal settings in a regular scenario is performed by means of multivariate linear regression (MLR) model inspection. The aim is to identify which are the most relevant variables in the scenario when finding optimal P_0 and α settings in
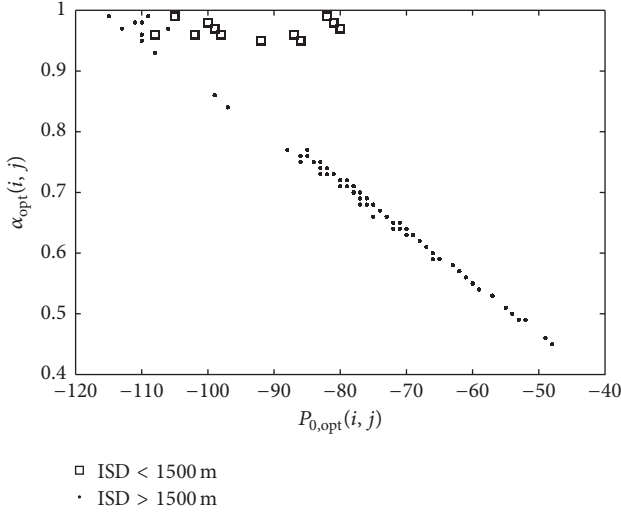
FIGURE 5: $P_{0,\text{opt}}(i, j)$ and $\alpha_{\text{opt}}(i, j)$ values.

a regular scenario. Thus, a set of regression equations for calculating optimal FPC settings is obtained as follows:

$$P_{0,\text{opt}}(c) = \beta_0 + \sum_{l=1}^{k} \beta_l \cdot p_l(c) + \epsilon,$$

$$\alpha_{\text{opt}}(c) = \beta_0 + \sum_{l=1}^{k} \beta_l \cdot p_l(c) + \epsilon, \tag{4}$$

where $P_{0,\text{opt}}(c)$ and $\alpha_{\text{opt}}(c)$ are the optimal solutions obtained from the self-planning approach for a combination c of independent variables (i.e., scenario configuration), l is used to index variables, p_l are the variables selected as candidate independent variables, β_l is the regression coefficient for each p_l, k is the number of candidate independent variables in the MLR model, and ϵ is the error term.

The first set of scenario variables to be considered in the regression analysis is as follows: (a) intersite distance, ISD, in a logarithmic scale as an indicator of propagation losses, (b) antenna height, h_b, in meters, (c) uplink cell load ratio, U_{UL}, (d) vertical antenna gain, V_G, measured in dB, as an indicator of the antenna tilt, and (e) horizontal gain, H_G, measured in dB, as an indicator of the relative angle between cells. A wide set of regular scenarios are built for different values of these variables. Specifically, 145800 different combinations of the independent variables have been tested. Note that a regular scenario can be built with a different value of the same variable for the central and surrounding cell (e.g., $h_b(i) \neq h_b(j)$ in a regular scenario). MLR model defined in (4) is, thus, built from $P_{0,\text{opt}}$ and α_{opt} solutions extracted from all the regular scenarios built according to the values set for these variables. MLR analysis has been performed for a fixed value of trade-off, $t = 0.5$.

A previous analysis of $P_{0,\text{opt}}$ and α_{opt} values obtained from the different regular scenarios is shown in Figure 5. Solutions in the figure are divided by ISD in the scenario. It is shown that there is a clear trend of optimal solutions with ISD. Parameter α_{opt} remains almost constant when ISD < 1500 m,

TABLE 2: MLR model for small-medium cells.

Regression statistics, ISD < 1500 m		
Determination coefficient	$R^{2(P_{0,\text{opt}})} = 0.9$	
	$P_{0,\text{opt}}$	
MLR analysis	$\widehat{\beta}_i$	P
Constant (β_0)	71.23	0
ISD [dB]	−5.437	0
$h_b(i)$ [m]	−3.8E − 15	1
$h_b(j)$ [m]	−2.24E − 15	1
$V_G(i)$ [dB]	−2.22E − 15	1
$V_G(j)$ [dB]	0.372	0
U_{UL}	−0.05	0.471
H_G [dB]	−2.83	1.5E − 97

whereas $P_{0,\text{opt}}$ moves in a wide range (20 dB). When ISD > 1500 m, the solution behavior is not constant, but lineal dependence between $P_{0,\text{opt}}$ and α_{opt} is shown. Thus, MLR model construction is performed in two stages: (a) small-medium cells (ISD < 1500 m) and (b) medium-large cells (ISD > 1500 m).

4.4.1. Small-Medium Cells MLR Analysis. Table 2 details MLR model analysis when ISD < 1500 m. Subscript i in the table refers to the value of such parameter for the central cell in the regular scenario (e.g., antenna height in cell i), and j subscript refers to the neighbor cell (e.g., antenna height for neighbor cell j). As shown in the table, $P_{0,\text{opt}}$ is accurately predicted when all variables are considered ($R^2 = 0.9$). As seen in Figure 5, α_{opt} remains almost constant (between 0.95 and 1), so no MLR model is built and $\alpha_{\text{opt}} = 1$ is considered.

Attending to the p values in Table 2, ISD and $V_G(j)$ prove to be the most important variables when calculating $P_{0,\text{opt}}$. With a similar methodology to that described in [34], a simplified MLR model is constructed by following a variable elimination process. Initially, all variables in Table 2 are included in the model. Then, in each iteration, the least important variable (that with the highest p value) is eliminated until the determination coefficient, R^2, remains at an acceptable level (i.e., when the determination coefficient, R^2, is less than 0.7). The resulting model is

$$\widehat{P_{0,\text{opt}}} = 64.545 - 5.437 \cdot \text{ISD} + 0.372 \cdot V_G(j),$$

$$\widehat{\alpha_{\text{opt}}} = 1, \tag{5}$$

where $V_G(j)$ refers to the vertical gain of the neighbor cell j towards the cell under study. This simplified model reaches $R^2 = 0.894$.

4.4.2. Medium-Large Cells MLR Analysis. Analogously, Table 3 shows the MLR model with all variables when ISD > 1500 m. As shown in the table, $P_{0,\text{opt}}$ and α_{opt} can be determined with relatively good accuracy attending to the determination coefficient ($R^2 > 0.75$ in both cases).

Regarding the p values in Table 3, ISD, $V_G(j)$, and $U_{\text{UL}}(i)$ prove to be the most important variables when calculating

TABLE 3: MLR analysis for medium-large cells.

Regression statistics, ISD > 1500 m				
Determination coefficient	$R^{2(P_{0,\mathrm{opt}})} = 0.76$		$R^{2(\alpha_{\mathrm{opt}})} = 0.75$	
	$P_{0,\mathrm{opt}}$		α_{opt}	
MLR analysis	$\widehat{\beta}_i$	P	$\widehat{\beta}_i$	P
Constant (β_0)	-423.53	0	3.639	0
ISD [dB]	10.527	0	-0.0893	0
$h_b(i)$ [m]	$-1.47E-15$	1	$2.91E-17$	1
$h_b(j)$ [m]	$-6.4E-16$	1	$2.03E-17$	1
$V_G(i)$ [dB]	$-3.29E-15$	1	$2.53E-17$	1
$V_G(j)$ [dB]	-0.258	$1.09E-296$	$2.44E-3$	0
U_{UL}	-17.083	0	0.144	0
H_G [dB]	-0.139	$1.3E-4$	$9.16E-4$	$3.08E-3$

$P_{0,\mathrm{opt}}$ and α_{opt}. Again, a simplified MLR model is constructed and defined as

$$\widehat{P_{0,\mathrm{opt}}} = -425.73 + 10.53 \cdot \mathrm{ISD} - 0.258 \cdot V_G(j) - 17.08$$
$$\cdot U_{\mathrm{UL}}(i),$$
$$\widehat{\alpha_{0,\mathrm{opt}}} = 3.65 - 0.0893 \cdot \mathrm{ISD} + 0.00244 \cdot V_G(j) + 0.144 \tag{6}$$
$$\cdot U_{\mathrm{UL}}(i).$$

5. Performance Analysis

Solutions for FPC settings proposed by the algorithm presented in this work are assessed in a system-level simulator implementing a realistic scenario. The simulator is the same as the one used in [30]. For the sake of clarity, the different tests are described first in Section 5.1 and the results are commented on in Section 5.2. Additionally, some comments on time complexity are presented in Section 5.3.

5.1. Analysis Setup. A static system-level LTE simulator like that in [30, 35] is used. In the simulator, the analyzed area is divided into a regular grid of points, representing potential user locations. For each network parameter setting (i.e., a FPC parameter plan), the received signal level at each base station from each point is computed by a macrocellular propagation model including log-normal slow fading (no fast fading is considered). Then, the serving cell for each point is defined as that providing the maximum signal level. Interference level is estimated by considering a nonuniform spatial user and cell load distribution following a realistic pattern extracted from a live LTE network. Then, radio link quality and efficiency are computed. Finally, different PMs are calculated by aggregating the previous measurements across all points in the scenario. Only uplink is considered here.

Despite its simplicity, the simulator is designed to make the most of available network statistics to model a live macrocellular scenario. For this purpose, the simulator includes the following features:

  (i) Delimitation of forbidden areas (i.e., points where users cannot be located) due to water resources by
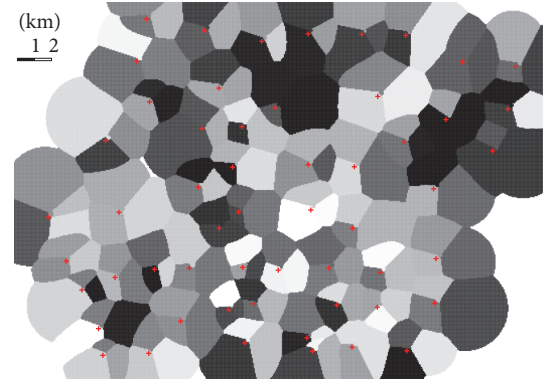

(km)
1 2

FIGURE 6: Simulated LTE network.

coastline files in Keyhole Markup Language (KML) format

 (ii) Parameterization of antenna model on a cell basis (maximum gain, horizontal/vertical beamwidth, etc.) depending on vendor equipment

(iii) Initialization of cell load distribution across the scenario with uplink utilization ratio (U_{UL}) derived from counters in the network management system of a live LTE network

(iv) Adjustment of spatial user distribution within a cell on a distance ring basis from timing advance (TA) distributions

 (v) Tuning of propagation model parameters based on the histogram of Reference Signal Received Power (RSRP) measurements

A live scenario is simulated. The scenario consists of 129 LTE cells covering a wide metropolitan area. Cells location, azimuth, antenna tilts, and uplink cell load are retrieved from network configuration management data stored in the operator network management system (NMS). In Figure 6, cell locations and services areas are represented.

Main configuration parameters of the simulated scenario are shown in Table 4. The rest of the simulation parameters were previously shown in Table 1.

Table 4: Network parameters.

Parameters	Settings
Carrier bandwidth	10 MHz (50 PRBs)
Carrier frequency	2.1 GHz
Cell layout	44 eNBs (129 sectors)
Intersite distance	0.45–4.5 km
eNB antenna height	15–54 m
eNB antenna tilt	$0°$–$13°$
Uplink cell load	5–19.2%

Three FPC parameter plans have been obtained under different trade-off conditions. FPC plans are calculated for $t = 0$ (a *capacity* oriented plan), $t = 1$ (a *coverage* oriented plan), and $t = 0.5$ (a *balanced* trade-off of capacity-coverage). Note that a FPC plan comprises $P_{0,\mathrm{opt}}(i)$ and $\alpha_{\mathrm{opt}}(i)$ values for every cell i in the scenario. As a benchmark for the solution comparison, different uniform FPC parameter settings (i.e., the same P_0 and α value for all cells) have also been simulated.

On the other hand and regarding the grade of complexity for simulations, three simulation configuration degrees are executed in ascending order of realism. In the lowest step, the simulator does not consider path-loss shadowing and uniform user spatial distribution is applied. In the second step, the simulator considers path-loss shadowing ($\sigma = 8$ dB) but still uses uniform user spatial distribution. Finally, in the third step, both path-loss shadowing and nonuniform user spatial distribution are considered. Here, it is worth noting that user spatial configurations are obtained from network realistic measurement.

To quantify solution performance, user-centric measurements are used. Overall cell-average user throughput, $\overline{\mathrm{TH}_{\mathrm{avg}}(i)}$, as the average of $\mathrm{TH}_{\mathrm{avg}}(i)$ for all cells in the scenario, is taken as a capacity indicator. Similarly, overall cell-edge user throughput, $\overline{\mathrm{TH}_{\mathrm{ce}}(i)}$, as the average of $\mathrm{TH}_{\mathrm{ce}}(i)$ for all cells in the scenario, is taken as a coverage indicator.

5.2. Results. Results are analyzed by simulator complexity and capacity/coverage trade-off first. Then, some comments on the performance of uniform parameter settings and time complexity are given.

5.2.1. Simulation Results with No Shadowing and Uniform Spatial User Distribution. Figure 7 shows the performance of the uniform FPC plans and trade-offs when no shadowing and uniform user distribution are considered. This figure can be considered as a *basic* case. Note that every plan is a point in the figure.

In the case of the capacity trade-off approach ($t = 0$), the proposed solution capacity ($\overline{\mathrm{TH}_{\mathrm{avg}}(i)}^{(\mathrm{cap})} = 18180$ kbps) slightly outperforms the maximum capacity achieved by any uniform FPC parameter settings. However, coverage ($\overline{\mathrm{TH}_{\mathrm{ce}}(i)}^{(\mathrm{cap})} = 8951$ kbps) is decreased by 7% with respect to the maximum coverage achieved by benchmark curves.

For the balanced trade-off approach ($t = 0.5$), coverage indicator ($\overline{\mathrm{TH}_{\mathrm{ce}}(i)}^{(\mathrm{bal})} = 10030$ kbps) outperforms the best
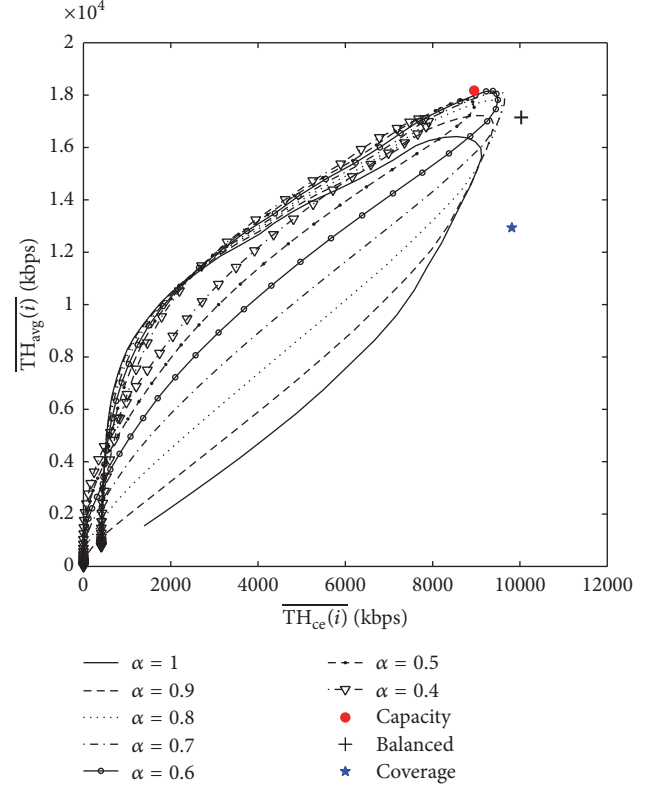

Figure 7: Simulation results, no shadowing, uniform traffic distribution.

coverage obtained by benchmark curves by 4%, but capacity decreases by 5.7% in comparison with the capacity approach. This is due to the trade-off between capacity and coverage.

Similar to the balanced trade-off approach, in the coverage approach ($t = 1$), coverage ($\overline{\mathrm{TH}_{\mathrm{ce}}(i)}^{(\mathrm{cov})} = 9816$ kbps) outperforms the best coverage obtained by any uniform FPC plan. Nonetheless, capacity is highly degraded (downgrade of 29% compared to capacity approach) as better coverage is obtained at low P_0 values, where the distance between optimal P_0 for capacity and coverage is higher as explained in Section 4.2.

As expected, balanced trade-off approach outperforms all other approaches, because it considers both capacity and coverage indicators for the construction of FPC plan.

5.2.2. Simulation Results with Shadowing and Uniform Spatial User Distribution. Figure 8 shows similar information when shadowing is included. The overall behavior is the same as that behavior shown in the basic case, Figure 7, but coverage values have decreased in all cases due to the inclusion of path-loss shadowing feature.

For the capacity approach, $\overline{\mathrm{TH}_{\mathrm{avg}}(i)}^{(\mathrm{cap})} = 17800$ kbps, performing as good as the best uniform plan. However, $\overline{\mathrm{TH}_{\mathrm{ce}}(i)}^{(\mathrm{cap})} = 5729$ kbps, which is 25% lower than the best performance achieved by benchmark curves ($\overline{\mathrm{TH}_{\mathrm{ce}}(i)}^{(\mathrm{max})} = 7652$ kbps).
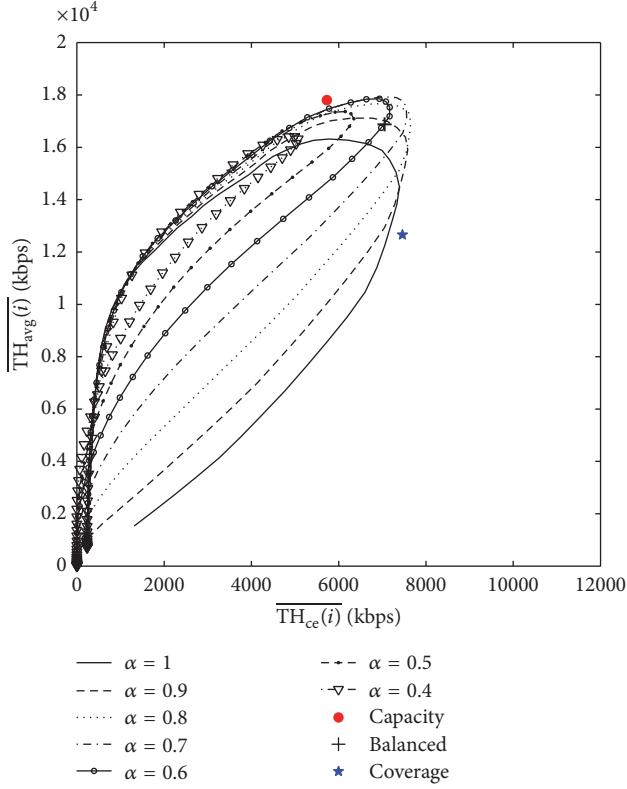
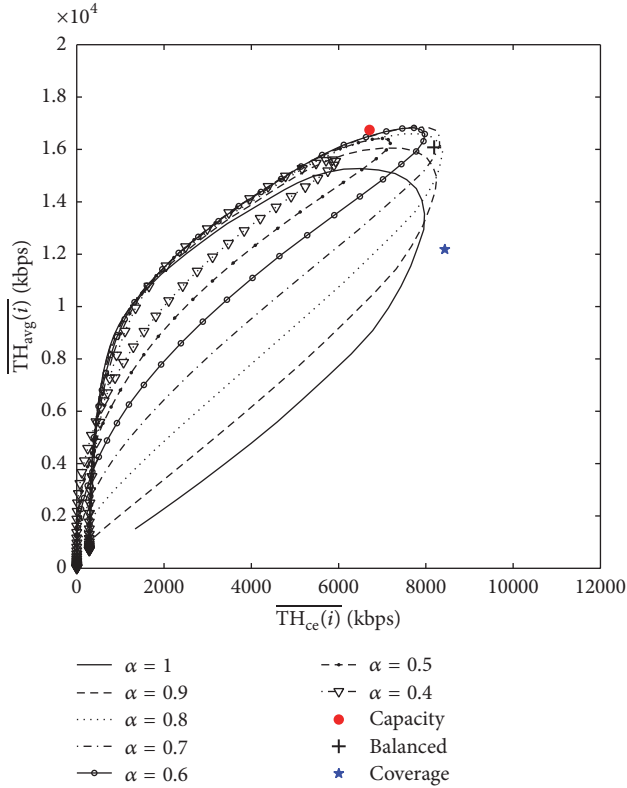FIGURE 8: Simulation results with shadowing, uniform traffic distribution.

FIGURE 9: Simulation results with shadowing and nonuniform traffic distribution.

When the balanced trade-off is considered, both coverage and capacity are degraded. Thus, capacity and coverage indicators are, respectively, 5.2% and 7.9% below the maximum achieved by other methods.

Coverage approach reaches $\overline{\text{TH}_{ce}(i)}^{(\text{cov})}$ = 7461 kbps, degrading only by 2.5% compared to the maximum achieved by other methods. However, capacity indicator $(\overline{\text{TH}_{avg}(i)}^{(\text{cov})}$ = 12650 kbps) is degraded by 29.2% regarding the maximum achieved by the capacity approach.

In this case, balanced trade-off approach is again the best solution proposed by the method. However, capacity and coverage values are below the maximum achieved by the set of uniform configurations. Nonetheless, the performance degradation remains below 8%.

5.2.3. Simulation Results with Shadowing and Nonuniform Spatial User Distribution. Finally, Figure 9 shows performance results including shadowing and nonuniform user distribution. Again, a similar behavior is observed, but some additional comments must be highlighted.

About the performance of capacity approach, capacity indicator $(\overline{\text{TH}_{avg}(i)}^{(\text{cap})}$ = 16750 kbps) is slightly degraded (0.71%) compared to the maximum achieved by all other solutions. Nonetheless, the coverage indicator $(\overline{\text{TH}_{ce}(i)}^{(\text{cap})}$ = 6708 kbps) is 19.33% below the maximum obtained by any uniform FPC plan. The balanced trade-off approach experiences a reduction of 4.7% and 1.57% for capacity and coverage indicators, respectively, compared to the other approaches. Regarding coverage approach, coverage indicator $(\overline{\text{TH}_{ce}(i)}^{(\text{cov})}$ = 8435 kbps) slightly outperforms (1.43%) the best result achieved by any other solution. Nonetheless, capacity indicator is degraded by 27.8% when compared to the best capacity performance for any other FPC plan.

Note that Figure 9 reflects the most realistic situation. Hence, performance results from the proposed FPC plans (i.e., capacity, coverage, and balanced approaches) perform similarly to the best uniform FPC plans. However, execution time when getting FPC plans with the proposed methodology is extremely decreased.

5.2.4. Uniform Parameter Settings. Figures 7, 8, and 9 show the performance for different network-wide uniform FPC settings. The impact of P_0 changes on capacity and coverage performance is the same as described in [30], and, hence, it is not detailed in this work. However, some comments on LTE network capacity and coverage regarding uniform α settings are outlined.

The different curves in those figures correspond to different α values. As expected by the analysis carried out in Section 4.2, maximum capacity is not achieved for total compensation, $\alpha = 1$. On the contrary, maximum capacity is achieved for medium α values (0.7–0.8). Regarding coverage and similar to capacity, there is a trade-off in the best value of α for coverage. Another interesting behavior is observed when α decreases. In this situation, curves are narrower. Higher and lower P_0 configurations have a similar impact on

capacity and coverage performance. This effect is produced by the decrement in $P_{0,\text{opt}}^{(\text{avg})}$-$P_{0,\text{opt}}^{(\text{ce})}$ when decreasing α described in Section 4.2.

5.3. Execution Time. All tests have been run on an Intel® Xeon© machine at 3.47 GHz with 12 GB of RAM. The execution time required to build curves from uniform FPC plans (i.e., uniform P_0 and α settings) was 974.67 seconds on average. The self-planning method proposed in this work for calculating cell-individual FPC parameters took an average of 64.5 seconds (0.5 seconds per cell), which is a gain of 15 times. However, if MLR models are used, execution time is almost instantaneous, which is quite useful in self-planning tools.

It is also worth mentioning that a uniform FPC requires resimulation of the whole scenario when a new cell is added to an existing (i.e., already planned) scenario. On the contrary, the self-planning method proposed in this work only needs to solve $1 + N_{\text{neigh}}(i)$ subproblems.

6. Conclusion

In this work, a very fast approach for the self-planning of FPC parameters in LTE uplink has been proposed. The method deals with irregular scenarios by dividing the large-scale multivariable optimization problem into multiple simple optimization problems, where a regular scenario is assumed on a per-adjacency basis. In those simplified scenarios, optimal FPC settings are calculated by an exhaustive search method. Additionally, optimal solutions have been analyzed and a set of regression equations have been given for the estimation of optimal FPC settings and, hence, reducing complexity and time execution. Finally, per-adjacency solutions are aggregated by averaging them into one solution per cell in the scenario. As a consequence, the proposed methodology is able to cope with complex irregular scenarios at a very low computational cost, which is very useful for self-planning tools. Performance assessment has been carried out over a system-level simulator implementing a real scenario. Results show that the proposed solutions work reasonably well when compared with uniform FPC plans designed by an exhaustive search process. Capacity performance is degraded by 4.7% and coverage performance by 1.57%. The main gain is in the execution time, which allows this approach to be used in frameworks where fast planning is required at a very low computational cost.

The proposed FPC self-planning methodology can be implemented as part of a centralized SON service. The system would retrieve eNB needed information from the NMS (location, azimuth, antenna tilt, uplink cell load, etc.). On the one hand, the method would perform self-configuration of every eNB in the network in the deployment stage. On the other hand, the method would reconfigure FPC parameter settings when a new eNB is added/removed from the network during the operational stage.

Competing Interests

The authors declare that they have no competing interests.

References

[1] 3GPP TS 36.301, LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Long Term Evolution (LTE) Physical Layer; General Description. V8.3, April 2009.

[2] NSN: Understanding Smartphone Behavior in the Network, Nokia Solutions and Networks Smart Labs, white paper, 2013.

[3] A. R. Mishra, *Advanced Cellular Network Planning and Optimisation*, John Wiley & Sons, New York, NY, USA, 2006.

[4] A. Hoikkanen, "Economics of 3G long-term evolution: the business case for the mobile operator," in *Proceedings of the 4th IEEE and IFIP International Conference on Wireless and Optical Communications Networks (WOCN '07)*, pp. 1–5, July 2007.

[5] S. Hamalainen, H. Sanneck, and C. Sartori, *LTE Self-Organising Networks (SON)*, John Wiley & Sons, New York, NY, USA, 2011.

[6] P. Oliver-Balsalobre, M. Toril, S. Luna-Ramírez, and J. M. Ruiz Avilés, "Self-tuning of scheduling parameters for balancing the quality of experience among services in LTE," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, no. 1, pp. 1–12, 2016.

[7] J. Lempiainen and M. Manninen, *Radio Interface System Planning for GSM/GPRS/UMTS*, Springer, New York, NY, USA, 2001.

[8] J. Laiho, A. Wacker, and T. Novosad, *Radio Network Planning and Optimisation for UMTS*, John Wiley & Sons, New York, NY, USA, 2002.

[9] 3GPP, "Telecommunication management; self-organizing networks (SON); concepts and requirements. V9," 3GPP TS 32.500, 2009.

[10] NGMN Use Cases Related to Self Organising Network, Overall Description, http://www.ngmn.org.

[11] J. Ramiro and K. Hamied, *Self-Organizing Networks: Self-Planning, Self-Optimization and Self-Healing for GSM, UMTS and LTE*, John Wiley & Sons, New York, NY, USA, 2011.

[12] L. Jorguseski, A. Pais, F. Gunnarsson, A. Centonza, and C. Willcock, "Self-organizing networks in 3GPP: standardization and future trends," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 28–34, 2014.

[13] J. F. Whitehead, "Signal-level-based dynamic power control for co-channel interference management," in *Proceedings of the 43rd IEEE Vehicular Technology Conference (VTC '93)*, pp. 499–502, May 1993.

[14] 3GPP TS 36.213, Physical Layer Procedures, V8.6, September 2009.

[15] W. Xiao, R. Ratasuk, A. Ghosh, R. Love, Y. Sun, and R. Nory, "Uplink power control, interference coordination and resource allocation for 3GPP E-UTRA," in *Proceedings of the IEEE 64th Vehicular Technology Conference (VTC '06)*, pp. 1–5, IEEE, Montreal, Canada, September 2006.

[16] A. M. Rao, "Reverse link power control for managing inter-cell interference in orthogonal multiple access systems," in *Proceedings of the IEEE 66th Vehicular Technology Conference (VTC-Fall '07)*, pp. 1837–1841, Baltimore, Md, USA, September 2007.

[17] A. Simonsson and A. Furuskär, "Uplink power control in LTE—overview and performance: principles and benefits of utilizing rather than compensating for SINR variations," in *Proceedings of the 68th Semi-Annual IEEE Vehicular Technology (VTC '08)*, pp. 1–5, September 2008.

[18] B. Muhammad and A. Mohammed, "Performance evaluation of uplink closed loop power control for LTE system," in *Proceedings of the IEEE 70th Vehicular Technology Conference Fall (VTC '09)*, pp. 1–5, Anchorage, Alaska, USA, September 2009.

[19] B. Muhammad and A. Mohammed, "Uplink closed loop power control for LTE system," in *Proceedings of the 6th International Conference on Emerging Technologies (ICET '10)*, pp. 88–93, IEEE, Islamabad, Pakistan, October 2010.

[20] S. Yang, Q. Cui, X. Huang, and X. Tao, "An effective uplink power control scheme in CoMP systems," in *Proceedings of the IEEE 72nd Vehicular Technology Conference Fall (VTC-Fall '10)*, pp. 1–5, Ottawa, Canada, September 2010.

[21] N. J. Quintero, *Advanced power control for UTRAN LTE uplink [Ph.D. thesis]*, Department of Electronic Systems, Aalborg University, 2008.

[22] C. Ú. Castellanos, D. L. Villa, C. Rosa et al., "Performance of uplink fractional power control in UTRAN LTE," in *Proceedings of the IEEE 67th Vehicular Technology Conference (VTC '08)*, pp. 2517–2521, IEEE, Singapore, May 2008.

[23] C. Suh, A. T. Koc, and S. Talwar, "Tradeoff power control for cellular systems," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '09)*, pp. 1–6, Honolulu, Hawaii, USA, December 2009.

[24] S. Xu, M. Hou, K. Niu, Z.-Q. He, and W.-L. Wu, "Coverage and capacity optimization in LTE network based on non-cooperative games," *The Journal of China Universities of Posts and Telecommunications*, vol. 19, no. 4, pp. 14–42, 2012.

[25] A. Awada, B. Wegmann, I. Viering, and A. Klein, "Optimizing the radio network parameters of the long term evolution system using Taguchi's method," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 8, pp. 3825–3839, 2011.

[26] C. Ú. Castellanos, F. D. Calabrese, K. I. Pedersen, and C. Rosa, "Uplink interference control in UTRAN LTE based on the overload indicator," in *Proceedings of the IEEE 68th Vehicular Technology Conference (VTC '08)*, pp. 1–5, IEEE, Calgary, Canada, September 2008.

[27] 3GPP, "Overload indicator handling for LTE, TSG RAN WG1 #50bis Meeting. V8.6," 3GPP R1-074349, ETSI, Sophia Antipolis Cedex, France, 2007.

[28] M. Dirani and Z. Altman, "Self-organizing networks in next generation radio access networks: application to fractional power control," *Computer Networks*, vol. 55, no. 2, pp. 431–438, 2011.

[29] K. Majewski and M. Koonert, "Analytic uplink cell load approximation for planning fractional power control in LTE networks," *Telecommunication Systems*, vol. 52, no. 2, pp. 1081–1090, 2013.

[30] J. Á. Fernández-Segovia, S. Luna-Ramírez, M. Toril, A. B. Vallejo-Mora, and C. Úbeda, "A computationally efficient method for self-planning uplink power control parameters in LTE," *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 1, article 80, 2015.

[31] 3GPP TS 36.211, LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation. V12.3, Octubre 2014.

[32] 3GPP TS 32.521, Technical Specification Group Services and System Aspects; Telecommunications Management; Self-Organizing Networks (SON) Policy Network Resource Model (NRM) Integration Reference Point (IRP); Requirements. V9, March 2010.

[33] P. Patras, A. Banchs, and P. Serrano, "A control theoretic approach for throughput optimization in IEEE 802.11e EDCA WLANs," *Mobile Networks and Applications*, vol. 14, no. 6, pp. 697–708, 2009.

[34] J. A. Fernández-Segovia, S. Luna-Ramírez, M. Toril, and J. J. Sánchez-Sánchez, "Estimating cell capacity from network measurements in a multi-service LTE system," *IEEE Communications Letters*, vol. 19, no. 3, pp. 431–434, 2015.

[35] V. Buenestado, M. Toril, S. Luna-Ramirez, J. M. Ruiz-Aviles, and A. Mendo, "Self-tuning of remote electrical tilts based on call traces for coverage and capacity optimization in LTE," *IEEE Transactions on Vehicular Technology*, 2016.

[36] European Commission, *COST Action 231: Digital Mobile Radio Towards Future Generation Systems: Final Report*, European Commission, Brussels, Belgium, 1999.

[37] 3GPP TR 36.814, Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Further Advancements for E-UTRA Physical Layer Aspects, V9, March 2010.

Research Article

WSN-Based Height Estimation of Moving Object in Surveillance Systems

Jaeseok Shim[1] and Yujin Lim[2]

[1]*Department of Computer Science, University of Suwon, San 2-2, Wau-ri, Bongdam-eup, Hwaseong, Gyeonggi-do 445-743, Republic of Korea*
[2]*Department of Information Technology Engineering, Sookmyung Women's University, Cheongpa-ro 47-gil 100, Yongsan-gu, Seoul 04310, Republic of Korea*

Correspondence should be addressed to Yujin Lim; yujin91@sookmyung.ac.kr

In the WSN- (wireless sensor network-) based surveillance system to detect undesired intrusion, all detected objects are not intruders. In order to reduce false alarms, human detection mechanism needs to determine if the detected object is a human. For human detection, physical characteristics of human are usually used. In this paper, we use the physical height to differentiate an intruder from detected objects. Using the measured information from sensors, we estimate the height of the detected object. Based on the height, if the detected object is decided as an intruder, an alarm is given to a control center. The experimental results indicate that our mechanism correctly and fast estimates the height of the object without complex computation.

1. Introduction

Conventional surveillance systems address a detection problem of an abnormal event in monitored space by deploying video systems. However, video systems are expensive in terms of hardware, storage, and communications. In particular, collecting multiple video streams imposes high demands on storage, online monitoring, and video analysis. The WSN (wireless sensor network) technology provides means to develop such a system with low costs [1]. The WSN can run standalone or it can be used to trigger external surveillance system like a video system. For example, in a hybrid system, when WSN detects and notifies an abnormal event, WSN triggers a video surveillance system. Then, camera is activated and zooms into the monitored space. The hybrid system optimizes accuracy, while keeping costs low. WSN is ideal for monitoring application because it is a fully automated system which does not require any human intervention and easily connected to external system so it can notify the user of the undesired situation.

There are a lot of monitoring applications based on WSN technology and they are interested in monitoring specific events or objects [2–5]. Not only do they need to figure out occurrence or existence of events or objects, but also they need to find out where the events have occurred or the objects are present. Examples of that kind of applications include fire monitoring systems, surveillance systems, livestock monitoring, and protection systems. The target application aimed by this paper is surveillance systems for home, office, or factory. In surveillance systems, it is very important to detect undesired intrusion while nobody is present in office, home, or factory. Typically, WSN-based surveillance systems start with motion detecting. Once any motion in the monitored space is detected, because all moving objects are not intruders, human detection mechanism needs to determine if the detected object is a human in order to reduce false alarms to video system. For human detection, physical characteristics of human are usually used. Height, speed, and body heat are one of these physical characteristics. When the height of a detected object is within a reasonable range for normal adults, for example, between 150 cm and 200 cm, we can consider the object as a human. In the same manner, we can check whether the mobility speed or the body heat of the object is within a reasonable range. In this paper, we use the physical height to differentiate an intruder from detected objects [6, 7]. Using the ranging information from sensors, we present a

height estimation mechanism for the detected object [8]. In order to reduce the estimation error of the height of the detected object, we extend the height estimation mechanism with a compensation algorithm using interpolation method. If the finally estimated height is in the height range for normal adults, it is determined that the object is a human and an alarm is given to activate video surveillance system. Our mechanism makes WSN-based surveillance system self-configuring because it can configure a range for normal adults in a dynamic way.

The rest of the paper is organized as follows. In Section 2, we introduce the related studies for 3D localization. In Section 3, we propose a height estimation mechanism with an error compensation algorithm for human detection. Then, we show the performance of the proposed mechanism in Section 4. Finally, the conclusion remarks are given in Section 5.

2. Related Work

In a localization system, most algorithms work on 2D (2-dimensional) plane, that is, x and y plane. In the 2D localization algorithms, the process of estimation is less complex and faster than the process of 3D (3-dimensional) localization algorithms [9]. The 2D localization algorithms provide good accuracy on flat terrains but they are difficult to estimate on rough terrains. Thus, mapping these positions estimated by the 2D algorithms to the real world can cause errors. By using 3D with one extra plane, that is, z plane, 3D localization algorithms provide more accurate results on rough terrains. In 2D space, at least three anchor nodes determine a coordinate system. In 3D space, four anchor nodes at least are required.

3D DV-Hop algorithm [10] expands the traditional range-free DV-Hop algorithm into 3D space. In the algorithm, the minimum hop counts between the detected object and the anchor nodes are calculated by using mobile agents. Then, the algorithm calculates the average per-hop distance of the object by using the minimum hop counts and the average per-hop distance of the anchor nodes. Based on the calculated distance, the position of the object is estimated. 3D Centroid algorithm [11] expands the traditional Centroid algorithm. All anchor nodes send their position information within their transmission range. An object collects the signals from anchor nodes and it selects randomly four anchor nodes in range to form a series of tetrahedrons. For each tetrahedron, the algorithm calculates the barycenter. Finally, the average coordinate of these barycenters is used as the estimated position of the object.

In 3D-ADAL (3-dimensional azimuthally defined area localization algorithm) [12], a mobile beacon moves in the whole network and broadcasts beacon messages within its transmission range. First, each sensor node estimates its position in the local xy-plane. Then, each node focuses on the local yz-plane of the mobile beacon and determines its altitude by using the information of tilt in the beacon message. In [13], the loss of wireless signal strength between the object and an anchor node is converted to the distance first, and then the maximum-likelihood estimation method is used to calculate the 3D coordinate of the object.

In other words, RSSI (received signal strength indication) value between the object and anchor node is converted to the distance. For 3D positioning, the object needs to communicate with four adjacent anchor nodes at least. When obstacles are present, the attenuation of wireless signal does not meet linear relationship with the distance any longer, so the compensation algorithm like the maximum-likelihood estimation method is used.

In 3D UMP (unitary matrix pencil) [14], it is based on TOA (time-of-arrival) estimation of UWB (ultrawideband) signal using UMP algorithm. The UMP algorithm is proposed to measure the distance between two nodes. The algorithm is extended to the applications of UWB-based WSN to reduce computational load and improve time resolution. The estimation results are used by multilateral localization algorithm for 3D position computation. In SDI (space distance intersection) [15], each sensor node measures a set of distances with the mobile beacon by using TOA techniques of UWB signal.

In [16], a closed-form solution for 3D localization using AOA (angle of arrival) is presented. It reduces the amount of estimation bias that is caused by the measurement noise and sensor position errors. Besides it achieves asymptotically the CRLB (Cramer-Rao lower bound) performance and maintains a bias level close to the maximum-likelihood estimator. Reference [17] applies the WIV (weighted instrumental variable) technique for AOA positioning. The WIV technique uses an IV matrix to reduce the noise correlation between the regressor and regressand.

The conventional 3D localization techniques require at least four anchor nodes to calculate the position of the detected object. Besides, they have high processing overhead. However, our surveillance system needs the only height information of the detected object to determine if the object is human, instead of the position of the object. Thus, we propose a lightweight height estimation mechanism with low processing overhead and our mechanism requires at least three anchor nodes. Based on the height estimation of the object, we can implement fast and accurate human detection system.

3. Proposed Mechanism

In order to estimate the height of a target object, we firstly estimate the distance from a ceiling to the target. Given the distance, the height of the target is calculated by subtracting the distance from the height of the monitored space where the target locates. We assume that sensors are deployed on a ceiling in the monitored space. The sensors measure the distances from themselves to the target. In Figure 1, the base of upside-down triangular pyramid indicates a ceiling of the monitored space. Three sensors mounted on a ceiling are vertices A, B, and C. The target is located at the apex (P_x) of the pyramid. The distances measured by sensors A, B, and C are d_x, e_x, and f_x. We propose an algorithm to estimate the height of the pyramid, x. Given a, b, c, d_x, e_x, and f_x, our algorithm estimates the height x.

The base area is the area of the triangular base in the pyramid. Using Heron's formula, we calculate the area of a
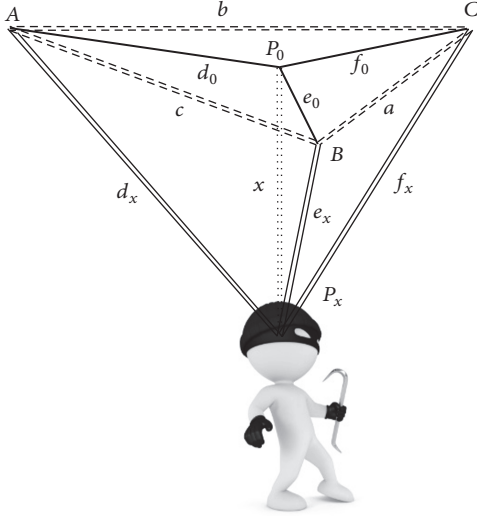
FIGURE 1: Symbols used by our algorithm.

triangle, $S(abc)$ when the lengths of all three sides, a, b, and c are known [18]. The base area whose sides have lengths a, b, and c is calculated using Heron's formula, as below.

$$S(abc)$$
$$= (a + b + c)(-a + b + c)(a - b + c)(a + b - c). \tag{1}$$

In other words, the base area with three vertices A, B, and C is

$$4\text{Area}_{\text{Base}} = \sqrt{S(abc)}$$
$$= \sqrt{S(ae_0 f_0)} + \sqrt{S(bf_0 d_0)} + \sqrt{S(cd_0 e_0)}. \tag{2}$$

The lateral area is the sum of areas of the side faces, that is, ΔBCP_x, ΔCAP_x, and ΔABP_x, in the pyramid. The lateral area is

$$4\text{Area}_{\text{Lateral}} = \sqrt{S(ae_x f_x)} + \sqrt{S(bf_x d_x)} + \sqrt{S(cd_x e_x)}. \tag{3}$$

The height of the pyramid x meets the base at a point, called P_0 (we can find P_0 by using triangulation). Using d_0, e_0, and f_0 in $d_x^2 = d_0^2 + x^2$, $e_x^2 = e_0^2 + x^2$, and $f_x^2 = f_0^2 + x^2$, (3) is rewritten as follows:

$$4\text{Area}_{\text{Lateral}} = \sqrt{S(ae_0 f_0) + 4a^2 x^2}$$
$$+ \sqrt{S(bf_0 d_0) + 4b^2 x^2} \tag{4}$$
$$+ \sqrt{S(cd_0 e_0) + 4c^2 x^2}.$$

By treating $S(ae_0 f_0)$, $S(bf_0 d_0)$, and $S(cd_0 e_0)$ as constants, (4) is rewritten as follows:

$$\sqrt{S(ae_0 f_0) + 4a^2 x^2} + \sqrt{S(bf_0 d_0) + 4b^2 x^2}$$
$$+ \sqrt{S(cd_0 e_0) + 4c^2 x^2} > \sqrt{4a^2 x^2} + \sqrt{4b^2 x^2} \tag{5}$$
$$+ \sqrt{4c^2 x^2} = 2(a + b + c)x.$$

Using $\sqrt{a + b} \leq \sqrt{a} + \sqrt{b}$ ($a, b \geq 0$), we rewrite (5) as follows:

$$\sqrt{S(ae_0 f_0) + 4a^2 x^2} + \sqrt{S(bf_0 d_0) + 4b^2 x^2}$$
$$+ \sqrt{S(cd_0 e_0) + 4c^2 x^2} \leq \sqrt{S(ae_0 f_0)} + \sqrt{4a^2 x^2}$$
$$+ \sqrt{S(bf_0 d_0)} + \sqrt{4b^2 x^2} + \sqrt{S(cd_0 e_0)} + \sqrt{4c^2 x^2} \tag{6}$$
$$= 2(a + b + c)x + \sqrt{S(abc)}.$$

Thus, the lateral area is

$$\frac{1}{2}(a + b + c)x < \text{Area}_{\text{Lateral}}$$
$$\leq \frac{1}{2}(a + b + c)x + \frac{1}{4}\sqrt{S(abc)}. \tag{7}$$

We define the function $f(x)$ as follows:

$$f(x) = \frac{1}{4}\left(\sqrt{S(ae_0 f_0) + 4a^2 x^2} + \sqrt{S(bf_0 d_0) + 4b^2 x^2}\right.$$
$$\left. + \sqrt{S(cd_0 e_0) + 4c^2 x^2}\right). \tag{8}$$

Using (7), we define the function $g(x)$ to estimate the lateral area as follows:

$$g(x) = \frac{1}{2}(a + b + c)x. \tag{9}$$

We can estimate x_{est} when $g(x_{\text{est}}) = f(x)$ as follows:

$$x_{\text{est}} = \frac{2f(x)}{a + b + c}$$
$$= \frac{\sqrt{S(ae_x f_x)} + \sqrt{S(bf_x d_x)} + \sqrt{S(cd_x e_x)}}{2(a + b + c)}. \tag{10}$$

The estimation error is the difference between $f(x)$ and $g(x)$. By tracking limit as x to infinity, the estimation error converges to 0 as follows:

$$\lim_{x \to \infty}(f(x) - g(x)) \approx 0. \tag{11}$$

But, when $x = 0$, the error is maximized

$$f(0) - g(0) = \frac{1}{4}\sqrt{S(abc)}. \tag{12}$$

To reduce the estimation error, we improve (7) using $\sqrt{a + b} \leq \sqrt{a} + \sqrt{b}$ ($a, b \geq 0$) as follows:

$$\frac{1}{2}(a + b + c)x + \frac{1}{4}\sqrt{S(abc)}$$
$$= \frac{1}{4}\left(\sqrt{S(abc)} + \sqrt{4(a + b + c)^2 x^2}\right) \tag{13}$$
$$\geq \frac{1}{4}\sqrt{S(abc) + 4(a + b + c)^2 x^2}.$$
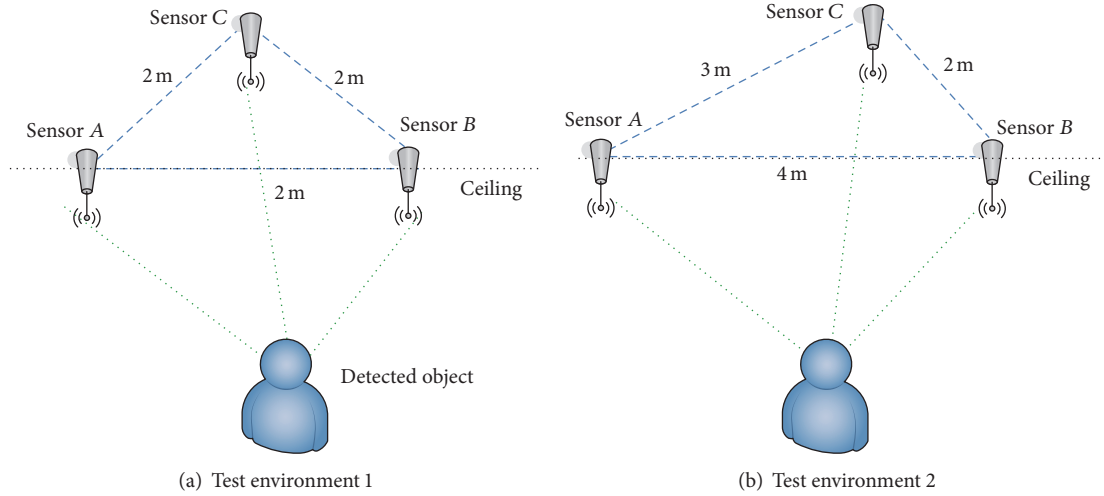
(a) Test environment 1

(b) Test environment 2

FIGURE 2: Test environments for evaluating performance.

Using (13), we improve the function $g(x)$ to estimate the lateral area as follows:

$$h(x) = \frac{1}{4}\sqrt{S(abc) + 4(a+b+c)^2 x^2}. \qquad (14)$$

Using (14), we can estimate $x_{\text{est\_new}}$ when $h(x_{\text{est\_new}}) = f(x)$ as follows:

$$
\begin{aligned}
x_{\text{est\_new}} &= \frac{\sqrt{16f(x)^2 - S(abc)}}{2(a+b+c)} \\
&= \frac{\sqrt{\left(\sqrt{S(ae_xf_x)} + \sqrt{S(bf_xd_x)} + \sqrt{S(cd_xe_x)}\right)^2 - S(abc)}}{2(a+b+c)}.
\end{aligned}
\qquad (15)
$$

As (7), we can get the following relationship:

$$
\begin{aligned}
\frac{1}{2}(a+b+c)x &< h(x) \\
&\leq \frac{1}{2}(a+b+c)x + \frac{1}{4}\sqrt{S(abc)}.
\end{aligned}
\qquad (16)
$$

Finally, when x increases on infinity or decreases on 0, the estimation error converges to 0 as follows:

$$
\begin{aligned}
\lim_{x\to\infty}(f(x) - h(x)) &\approx 0, \\
f(0) - h(0) &= 0.
\end{aligned}
\qquad (17)
$$

As a result, given a, b, c, d_x, e_x, and f_x, we estimate the height of the target by using (15) while minimizing the estimation error.

4. Performance Analysis

To evaluate the performance of our mechanism, we implement sensors on the mote modules [19] equipped with DMS (distance measurement sensor) and communications between them are based on IEEE 802.15.4 standard [20]. We deploy three sensors on a ceiling in a laboratory in our college building and we devise two test environments, as shown in Figure 2. Firstly, Figure 2(a) shows the *test environment 1*. In the environment, we use an equilateral triangle which is a triangle that has equal length (2 m) on all three sides. Figure 2(b) shows the *test environment 2*. We use a triangle with 2 m × 3 m × 4 m. We have two test scenarios for each test environment. In the first scenario, there is no obstacle between sensors and the detected object. In the second scenario, there is an obstacle, such as ceiling fixture, between sensor A and the detected object. In both test environments, we set the distance between the ceiling and a detected object (x) to 0.5 m or 1 m, respectively. For each experiment, we measure 1000 samples.

Figure 3 shows the estimation accuracy of our mechanism in the *test environment 1*. We plot a normal distribution with mean μ and standard deviation σ. Each dot indicates $x_{\text{est\_new}}$ calculated by using (15) based on each measured sample. Figures 3(a) and 3(b) indicate the estimation accuracy when there is no obstacle between the sensors and the object. In the subfigures, $\pm 2\sigma$ shows a 95% confidence interval. When $x = 0.5$ m, 946 samples are correctly estimated as $x \approx x_{\text{est\_new}}$, and the estimation accuracy is 94.6%. When $x = 1$ m, the accuracy is 95%. Besides, we are 99% confident that the estimation accuracies are 99.6% and 99.5%, respectively. Figures 3(c) and 3(d) indicate the estimation accuracy when there is the obstacle. In the subfigures, the effect of an obstacle on the estimation is analyzed. When $x = 0.5$ m, the estimation accuracy is 91.2% or 99% with a 95% or 99% level of confidence, respectively. When $x = 1$ m, the accuracy is 94.6% or 99.4%. The performance slightly decreases due to the environmental impact such as obstacles. To compensate the impact, more sensors could be deployed or error compensation mechanism could be used.

Figure 4 shows the estimation accuracy of our mechanism in the *test environment 2*. The figure shows the effect of the deployment of sensors on the estimation. As shown

(a) $x = 0.5$ m, no obstacle

(b) $x = 1$ m, no obstacle

(c) $x = 0.5$ m, obstacle

(d) $x = 1$ m, obstacle
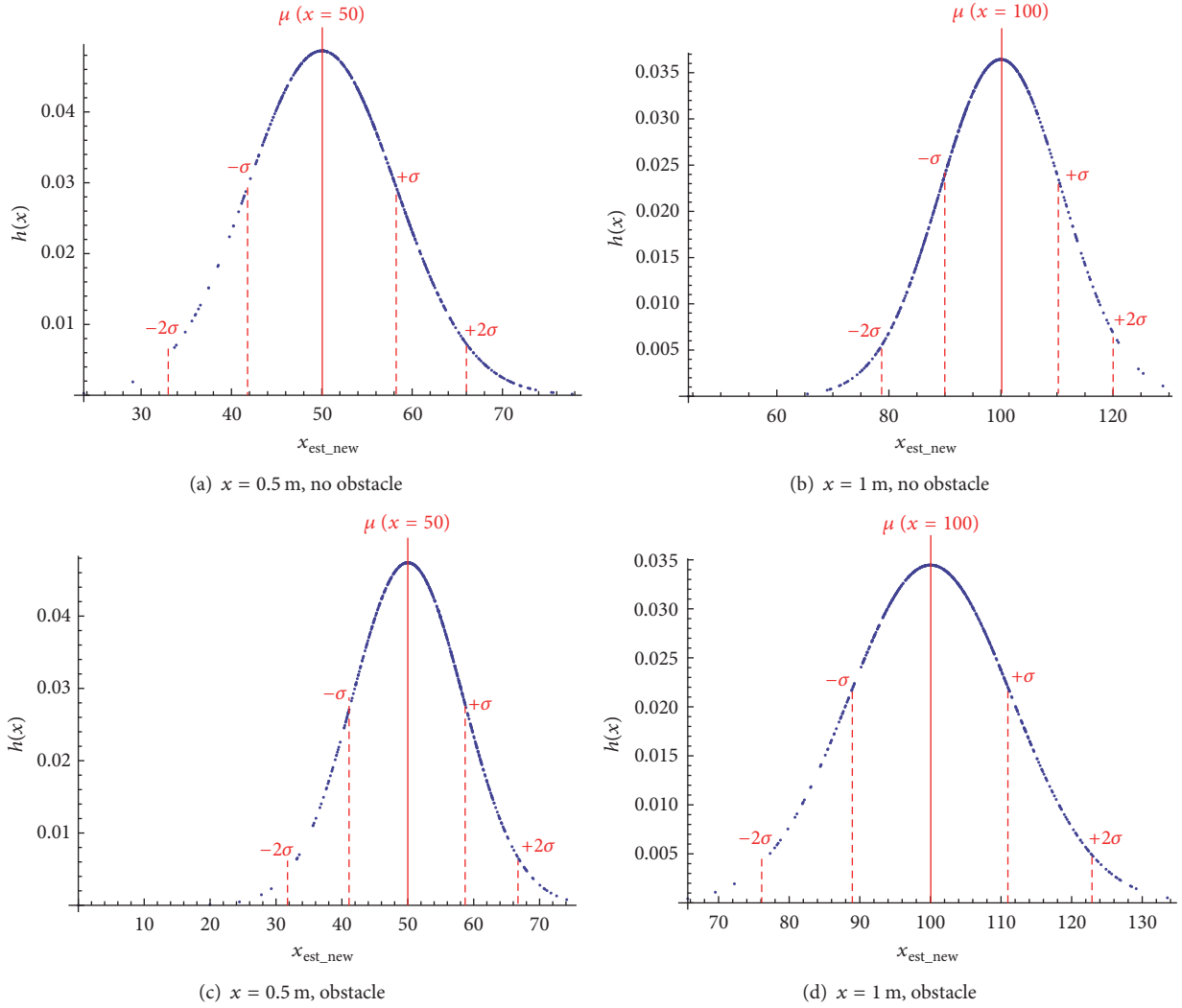
FIGURE 3: Estimation accuracy in *test environment 1*.

in Figures 4(a) and 4(b), when there is no obstacle between the sensors and the object, the accuracies are 95.6% and 94.8% with a 95% level of confidence. Besides, we are 99% confident that the accuracies are 98.8% and 99.2%. When $x = 0.5$ m with the obstacle, as shown in Figure 4(c), the estimation accuracy is 93.4% or 98.8% with a 95% or 99% level of confidence, respectively. When $x = 1$ m with the obstacle, as shown in Figure 4(d), the accuracy is 89.6% or 99.2%.

Figure 5 shows the estimation error of our mechanism in the *test environment 2* with the obstacle. In the figure, the effect of the number of sensors on the estimation is analyzed. Because our mechanism operates based on the area of a triangle, we increase the number of sensors from 3 to 5 when $x = 1$ m. In the figure, estimation error 0% means that the height of the object is estimated exactly. We measure 100 samples for the figure and the x-axis indicates the number of the experimental samples. As shown in the figure, the accuracy is about 99% when three sensors are deployed. When four or five sensors are deployed, we randomly select three sensors to form a series of triangles. For

each triangle, our mechanism estimates the distance. Then, the average distance of the estimated distances is used as the final estimated distance between the ceiling and the object. The estimation accuracies are over 99% when four or five sensors are deployed. As a result, the performance is more affected by obstacle than the deployment of sensors or the number of sensors. The experimental results indicate that our mechanism correctly and fast estimates the distance from a ceiling to an object without complex computation.

5. Conclusion

In the WSN-based surveillance systems, they need to figure out the occurrence or existence of events or objects. To estimate the position of the events or objects, 3D localization techniques have been used. However, the conventional 3D localization techniques require at least four anchor nodes to calculate the position of the detected object. Besides, they cause high processing overhead. However, our surveillance system needs the only height information of the detected
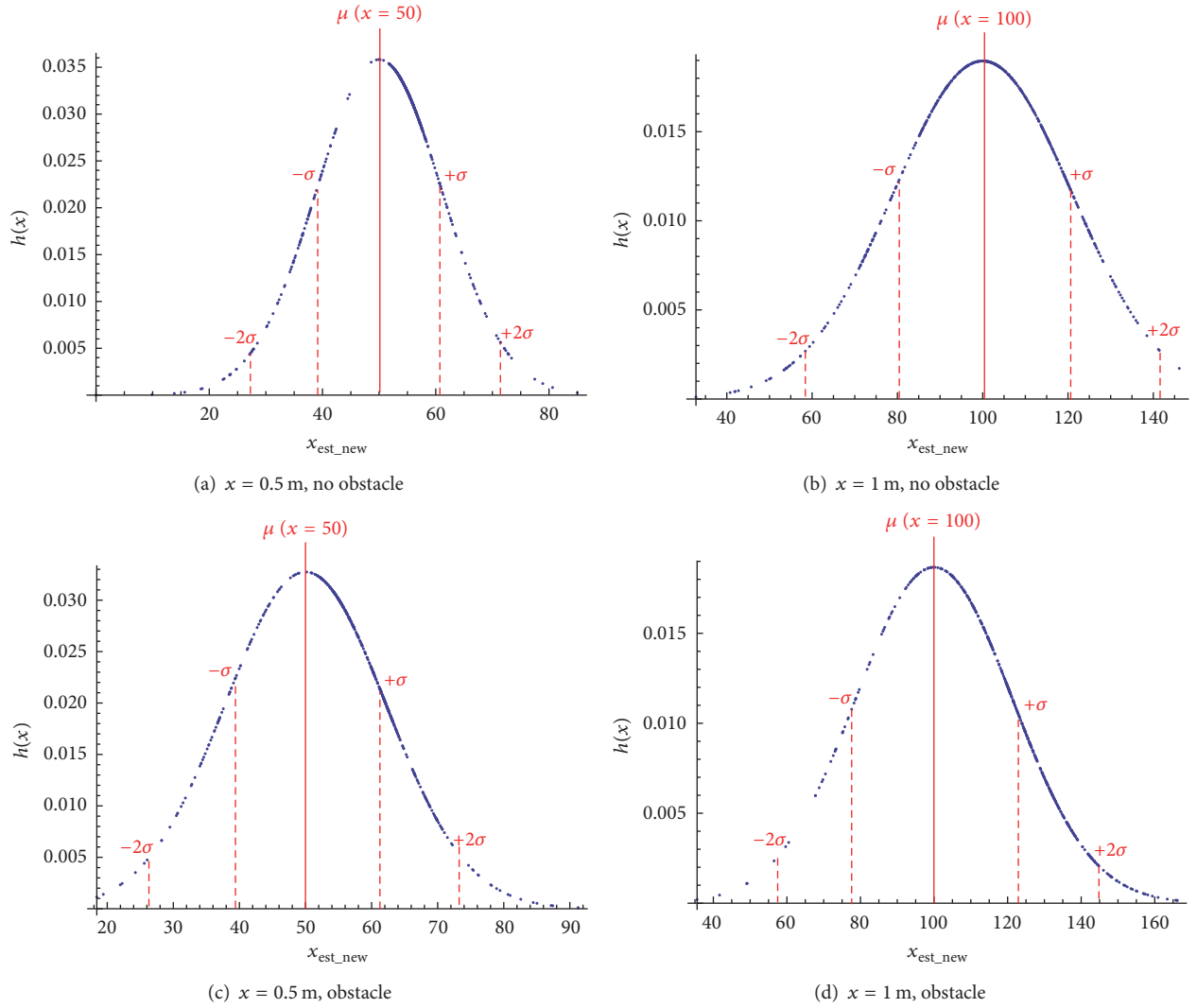
(a) $x = 0.5$ m, no obstacle

(b) $x = 1$ m, no obstacle

(c) $x = 0.5$ m, obstacle

(d) $x = 1$ m, obstacle

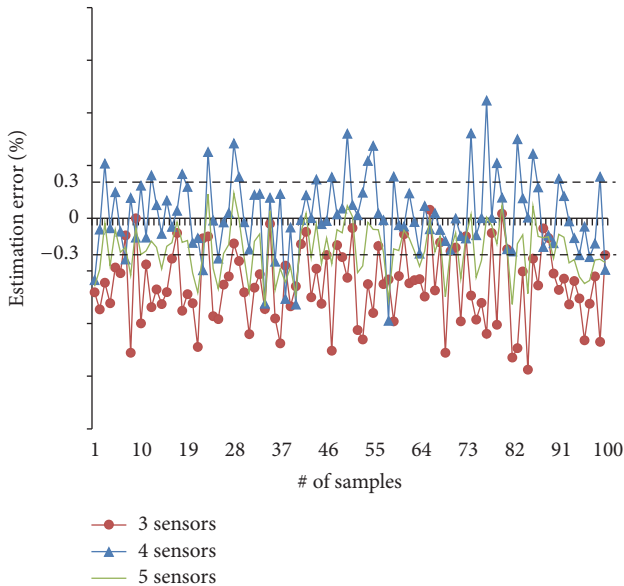FIGURE 4: Estimation accuracy in *test environment 2*.

FIGURE 5: Estimation error with varying the number of sensors.

object to determine if the object is human, instead of the position of the object. We propose a height estimation mechanism with low processing overhead. On the top of the mechanism, in this paper, we revise a compensation algorithm using interpolation method to reduce the estimation error of the height. Through implementation of our mechanism, we verify the performance of the height estimation. The experimental results indicate that our mechanism correctly and fast estimates the height of the object without complex computation. Through performance analysis, we show that the estimation accuracy is more affected by obstacle than the deployment of sensors or the number of sensors. For the future work, we plan to perform experiments with a practical scenario and to improve the estimation mechanism to compensate the estimation error or the environmental impact.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

References

[1] M. Walchli and T. Braun, "Building intrusion detection with a wireless sensor network," in *Ad Hoc Networks*, vol. 28 of *Lecture Notes of the Institute for Computer Sciences*, pp. 607–622, Springer, Berlin, Germany, 2010.

[2] R. Lara, D. Benítez, A. Caamaño, M. Zennaro, and J. L. Rojo-Álvarez, "On real-time performance evaluation of volcano-monitoring systems with wireless sensor networks," *IEEE Sensors Journal*, vol. 15, no. 6, pp. 3514–3523, 2015.

[3] S. K. Dash, J. P. Sahoo, S. Mohapatra, and S. P. Pati, "Sensor-cloud: assimilation of wireless sensor network and the cloud," in *Advances in Computer Science and Information Technology Networks and Communications*, vol. 84 of *Springer Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pp. 455–464, 2012.

[4] L. Buttyán, D. Gessner, A. Hessler, and P. Langendoerfer, "Application of wireless sensor networks in critical infrastructure protection: challenges and design options," *IEEE Wireless Communications*, vol. 17, no. 5, pp. 44–49, 2010.

[5] F. Viani, M. Salucci, P. Rocca, G. Oliveri, and A. Massa, "A multisensor WSN backbone for museum monitoring and surveillance," in *Proceedings of the IEEE 6th European Conference on Antennas and Propagation (EUCAP '12)*, pp. 51–52, IEEE, Prague, Czech Republic, March 2012.

[6] T. Suzuki, A. Khan, M. Kobayashi, and W. Takita, "Collaboration in routing and velocity measurement function for mobile ad hoc networks," in *Proceedings of the IEEE International Conference on Future Generation Communication and Networking (FGCN '07)*, pp. 108–113, Jeju Island, Korea, December 2007.

[7] E. Zamora and W. Yu, "Recent advances on simultaneous localization and mapping for mobile robots," *IETE Technical Review*, vol. 30, no. 6, pp. 490–496, 2013.

[8] J. Shim and Y. Lim, "Height estimation for applications in surveillance systems," *IETE Technical Review*, vol. 32, no. 5, pp. 321–329, 2015.

[9] S. Samanta, P. U. Tembhare, and C. R. Pote, "A survey on 3D localization in wireless sensor networks," *International Journal of Computational Engineering Research*, vol. 3, no. 1, pp. 90–94, 2013.

[10] L. Wang, J. Zhang, and D. Cao, "A new 3-dimensional DV-Hop localization algorithm," *Journal of Computational Information Systems*, vol. 8, no. 6, pp. 2463–2475, 2012.

[11] H. Chen, P. Huang, M. Martins, H. C. So, and K. Sezaki, "Novel centroid localization algorithm for three-dimensional wireless sensor networks," in *Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '08)*, pp. 1–4, October 2008.

[12] E. Guerrero, H. Wang, J. Alvarez, and L. Rivero, "A three-dimensional range-free localization algorithm based on mobile beacons for wireless sensor networks," *Computer Aided Drafting, Design and Manufacturing*, vol. 2010, no. 10, pp. 83–92, 2010.

[13] Q. Liu, P. Ren, and Z. Zhou, "Three-dimensional accurate positioning algorithm based on wireless sensor networks," *Journal of Computers*, vol. 6, no. 12, pp. 2582–2589, 2011.

[14] C. Liu, H. Jiang, and D.-L. Zeng, "Unitary matrix pencil algorithm for range-based 3D localization of wireless sensor network nodes," *Journal of Networks*, vol. 7, no. 9, pp. 1384–1390, 2012.

[15] Q. Shi, H. Huo, T. Fang, and D. Li, "A 3D node localization scheme for wireless sensor networks," *IEICE Electronics Express*, vol. 6, no. 3, pp. 167–172, 2009.

[16] Y. Wang and K. C. Ho, "An asymptotically efficient estimator in closed-form for 3-D AOA localization using a sensor network," *IEEE Transactions on Wireless Communications*, vol. 14, no. 12, pp. 6524–6535, 2015.

[17] H.-J. Shao, X.-P. Zhang, and Z. Wang, "Efficient closed-form algorithms for AOA based self-localization of sensor nodes using auxiliary variables," *IEEE Transactions on Signal Processing*, vol. 62, no. 10, pp. 2580–2594, 2014.

[18] D. M. Mitchell, "A heron-type formula for the reciprocal area of a triangle," *Mathematical Gazette*, vol. 89, no. 516, pp. 494–516, 2005.

[19] C. Suh, J.-E. Joung, and Y.-B. Ko, "New RF models of the TinyOS simulator for IEEE 802.15.4 standard," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '07)*, pp. 2238–2242, Hong Kong, March 2007.

[20] IEEE 802.15 WPAN™ Task Group, IEEE standard, http://www.ieee802.org/15/pub/TG4.html.

Research Article

Random-Access Technique for Self-Organization of 5G Millimeter-Wave Cellular Communications

Jasper Meynard Arana, Joo Pyo Han, and Yong Soo Cho

School of Electrical and Electronics Engineering, Chung-Ang University, Seoul, Republic of Korea

Correspondence should be addressed to Yong Soo Cho; yscho@cau.ac.kr

The random-access (RA) technique is a key procedure in cellular networks and self-organizing networks (SONs), but the overall processing time of this technique in millimeter-wave (mm-wave) cellular systems with directional beams is very long because RA preambles (RAPs) should be transmitted in all directions of Tx and Rx beams. In this paper, two different types of preambles (RAP-1 and RAP-2) are proposed to reduce the processing time in the RA stage. After analyzing the correlation property, false-alarm probability, and detection probability of the proposed RAPs, we perform simulations to show that the RAP-2 is suitable for RA in mm-wave cellular systems with directional beams because of the smaller processing time and high detection probability in multiuser environments.

1. Introduction

As the demand for capacity in mobile broadband communications increases drastically every year, wireless communication industries are preparing to support up to a thousandfold increase in total mobile traffic by 2020 [1–3]. Millimeter-wave (mm-wave) communications with possible multigigabit-per-second data rates have attracted much attention as a candidate technology for the 5G era [4, 5]. Highly directional beamforming antennas are necessary at both the base station (BS) and mobile station (MS) to compensate for the high attenuation in the mm-wave frequency band and to extend its transmission range. With the small wavelength corresponding to mm-wave frequencies, antenna arrays can be easily installed in the MS.

A misalignment between transmit (Tx) and receive (Rx) beams may cause a significant loss in the received power, especially for systems with narrow beams. Beam alignment in mm-wave communication systems is necessary to find the best beam pair from all possible beam pairs for maximum beamforming efficiency. Currently, switched beamforming techniques with a set of predefined angles are used for Tx-Rx beamforming because the multiple analog chains at mm-wave frequencies are costly, and sampling analog signals at

GHz rates consumes a substantial amount of power [6]. The best beam pair is determined by selecting a beam pair with the maximum array gain. The Tx-Rx beamforming technique using the 60 GHz unlicensed spectrum with a bandwidth of 2.16 GHz has already been standardized in IEEE 802.11ad to provide multigigabit-per-second data rates [7]. However, the standard is mainly designed for indoor communications (wireless LAN) and is not adequate for cellular communication.

Random access (RA) is a key procedure in cellular networks, enabling an MS to initiate communications and time alignment to a BS. RA is also one of the key procedures in self-organizing networks (SONs) [4]. In [8], a RA procedure for mm-wave cellular systems is described, where preambles are transmitted/received repeatedly in multiple directions from the MS/BS. Individual Tx beams are transmitted by the MS until all of the Tx beams are transmitted. The Rx beam sweep is performed at the BS for each Tx beam to measure the signal-to-noise ratio (SNR) for every Tx-Rx pair. However, the total RA duration will be very long because the RA preamble (RAP) should be long enough to be detected by the BS in large coverage areas, and multiple preambles should be transmitted for all directions of Tx and Rx beams. In [9], a RA technique with beamforming, referred to as an adaptive

spatial RA, is proposed, where beams are generated depending on the number of MSs and their locations. In this technique, the probability of a collision can be significantly reduced by exploiting the space domain efficiently in the RA stage. However, this technique cannot be employed in mm-wave cellular systems with switched beamforming because a fixed set of predefined angles is used in switched beamforming, and the previous collision statistic is required to adjust the beams in the next frame.

In this paper, we propose a new technique that can reduce the processing time in the RA stage. We design an RAP for mm-wave cellular systems with switched beamforming. Unlike the conventional RA preamble, which carries six bits (a total of 64 preamble IDs), the proposed RA preamble carries the information on beam ID (BID; optimal downlink Tx beam) as well as the preamble ID (PID; one of 64 signatures), all at the physical layer. The preambles are designed such that the additional information on the BID does not affect the number of available PIDs. We propose two different types of preambles (RAP-1 and RAP-2) that satisfy the constant-amplitude zero-autocorrelation (CAZAC) property for RA in mm-wave cellular systems. After analyzing the properties of the proposed preambles, false-alarm and detection probabilities are derived when the proposed preambles are used for RA. Using simulations, we show that the proposed technique can significantly reduce the processing time in the RA stage, and the proposed RAP-2 is appropriate for RA in mm-wave cellular systems with directional beams because it provides high detection probability in multiuser environments.

The remainder of the paper is organized as follows. In Section 2, we discuss an RA procedure for mm-wave cellular systems with directional beams. We propose an RA procedure that can reduce the processing time using the proposed RAP. In Section 3, we propose two different types of RAPs for mm-wave cellular systems and discuss their correlation properties. In Section 4, we describe a detection technique for the proposed RAP and derive the false-alarm probability and detection probability when the proposed RAPs are used for RA. In Section 5, we evaluate the performance of the proposed RA technique by performing computer simulations using a simple model of an mm-wave cellular system after verifying the properties of the proposed RAP. In Section 6, we conclude the paper.

2. RA Procedure in mm-Wave Cellular Systems

RA is generally performed when an MS turns on from sleep mode, performs a handover from one cell to another, or loses uplink time synchronization. The RA procedure in Long-Term Evolution (LTE) consists of five steps. In the first step, downlink synchronization and cell searching are performed by receiving synchronization channel and broadcast channel [10]. After acquiring downlink synchronization and receiving system information including information on parameters specific to RA, the MS performs RA preamble transmission in Step 2. The MS selects one of 64 predefined sets of preambles and transmits it using the time-frequency resources indicated by the system information. When the BS successfully receives the RA preamble, it sends a RA response (RAR) indicating
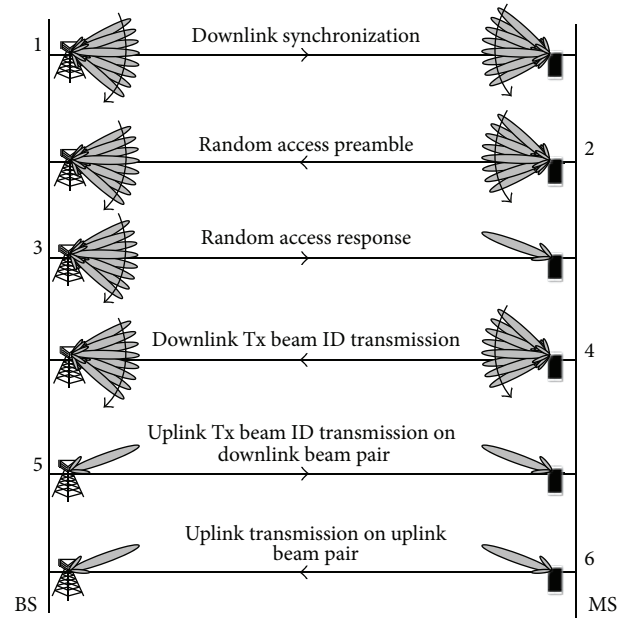

Figure 1: Conventional random-access procedure in mm-wave cellular systems.

a preamble index, uplink timing advance (TA), and uplink resource-allocation information in Step 3. The MS can then determine whether its RA attempt has been successful by matching the preamble index. In Step 4, if the preamble index matches, the MS uses the TA information to adjust its uplink timing and transmits a RA message including the MS identity in the resource allocation provided by the previous step. In Step 5, the BS transmits the MS identity to the MS. If the MS cannot decode its own MS identity, it will automatically exit the procedure and resume the RA process.

In mm-wave cellular communication systems with directional beams, the RA procedure needs to be modified because the best beam pair for uplink transmission is unknown in the initial state. Figure 1 shows the initial RA procedure for an LTE-based mm-wave cellular communication system with directional antennas. Hereafter, we refer to this procedure as the conventional RA procedure. Here, we assumed that channel reciprocity does not hold; that is, the best beam pair in the downlink is not the same as the best beam pair in the uplink because of the different characteristics of the RF circuitry of the transmitter and receiver.

The first step starts with the downlink synchronization. The BS transmits synchronization signals, a primary synchronization signal (PSS), and secondary synchronization signal (SSS), for synchronization at the MS [11]. Then, the BS transmits a system information block (SIB) on a physical broadcast channel (PBCH), which carries configuration parameters of physical RA channels (PRACHs) such as the preamble root index. The same signals are transmitted repeatedly in individual beams until all of the Tx and Rx beams are swept, because the best downlink beam pair is not yet known. The MS is synchronized with the BS and determines the best downlink beam pair (downlink Tx beam and Rx beam) in this step. Upon decoding the SIB transmitted on the best downlink

beam pair, the MS will select a preamble signature for RA. In Step 2, the selected preamble signature will be transmitted repeatedly from the MS by sweeping the Tx and Rx beams. If the preamble signature is detected, the BS will send an RAR in Step 3. The RAR contains the necessary information for uplink data transmission at the MS, such as TA, temporary identifier, and resource-allocation information. The BS needs to transmit the same RAR repeatedly by sweeping downlink Tx beams because the BS does not yet know the best downlink Tx beam. At this step, the MS knows the best downlink beam pair. Therefore, the RAR is received at the MS with the best downlink Rx beam determined from Step 1. After decoding the RAR, the MS completes uplink synchronization with the BS. In Step 4, the MS transmits the information of the best downlink Tx beam ID, obtained from Step 1, to the BS in a message format after performing a timing adjustment. At this step, the BS determines the best uplink beam pair (uplink Tx beam and Rx beam) using the received message. Note that the best downlink beam pair that is estimated in Step 1 is generally not the same as the best uplink beam pair. In Step 5, the BS sends the information on the best uplink Tx beam ID to the MS. In this step, the downlink message is transmitted using the best downlink Tx beam because the BS has obtained the information from the previous step (Step 4). In Step 6, the MS can send any uplink message using the best uplink Tx beam. Step 6 completes the information exchange (between BS and MS) on the best beam selected for the uplink and downlink. This procedure may be repeated multiple times until successful RA is achieved.

However, the overall RA duration is too long, especially for a moving terminal because the synchronization parameters and best beam pairs need to be reselected periodically. In the proposed procedure, the duration of RA is reduced by transmitting a modified version of the RA preamble used in LTE systems. The proposed RA preamble carries the additional information of the best downlink Tx beam ID at the physical layer as well as preamble signature. The proposed RA procedure is shown in Figure 2. Step 1 in Figure 2 is the same as the first step in the conventional RA procedure. The MS is synchronized with the BS and determines the best downlink beam pair (downlink Tx beam and Rx beam). The MS will select a preamble signature for RA after decoding the SIB transmitted on the best downlink beam pair. In Step 2, the proposed RA preamble is transmitted repeatedly from the MS by sweeping the Tx and Rx beams. The proposed preamble carries the information on the best downlink Tx beam ID and selected preamble ID at the physical layer. Then, the BS detects the best downlink Tx beam ID and preamble ID from the received RA preamble. The BS also determines the best uplink beam pair (uplink Tx beam and Rx beam) from the received preamble. In Step 3, the BS sends an RAR to the MS. The BS does not need to transmit the same RAR repeatedly by sweeping downlink Tx beams because the BS has received the best downlink Tx beam ID from Step 2. The RAR contains the information on the best uplink Tx beam ID as well as the information required for uplink transmission at the MS, such as TA, all at the MAC layer. The RAR is received at the MS with the best downlink Rx beam determined from Step 1. In Step 4, the MS can send any uplink message using the
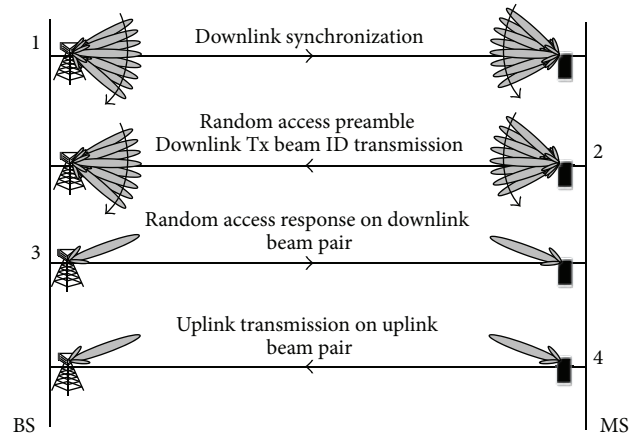

Figure 2: Proposed random-access procedure in mm-wave cellular systems.

best uplink Tx beam. In this procedure, the steps required for RA and beam selection are reduced using the proposed RA preamble. In the following sections, we describe the preamble design and detection for RA in mm-wave cellular communication systems with directional beams.

3. Proposed RA Preamble

In this section, we discuss an RA preamble that is suitable for mm-wave cellular systems with directional beams. The RA preamble waveform should have good detection probability while maintaining a low false-alarm rate, low collision probability, and low peak-to-average power ratio (PAPR) and allow accurate timing estimation. Because the proposed preamble can be viewed as a modified version of the RAP used in LTE systems, we will start with a short summary on LTE RAP.

The PRACH in LTE systems is generated from the cyclic shift of the prime-length Zadoff-Chu (ZC) sequence, where the root of the sequence is unique to each cell. The 64 different preambles are generated by the cyclic shifts of the base sequence with a length of 839. The MS selects one of 64 available PIDs and transmits the RAP in PRACH [11]. The cyclically shifted ZC sequences have a constant envelope, which ensures low PAPR property. The sequences have an ideal cyclic autocorrelation, which is important for obtaining an accurate timing estimation at the BS. In addition, the cross correlation between different preambles based on the cyclic shifts of the same ZC sequence is zero as long as the cyclic shift is larger than the maximum round-trip propagation delay in the cell plus the maximum delay spread of the channel. Therefore, there is no interference when multiple RA attempts are performed simultaneously using the preambles derived from the same ZC sequence. To handle different cell sizes, the value of the cyclic shift is broadcast as part of the system information. The set of cyclic shifts that produce a zero correlation of the ZC sequence is defined as a zero correlation zone (ZCZ) in the LTE specification [12].

In this paper, we propose two types of RAPs (RAP-1 and RAP-2) for mm-wave cellular communication systems.

The RAP-1 consists of two preambles, each one being constructed with the existing RAP structure in LTE systems. Both preambles share the same root index. The PID and BID are mapped onto cyclic shifts of the first and second preambles, respectively. The best downlink Tx beam ID that is obtained at the MS from Step 1 in the proposed RC procedure will be assigned to the value of BID. Here, uth root ZC sequence is defined by

$$x_u(n) = e^{-j(\pi u n(n+1)/N_{ZC})}, \quad 0 \le n \le N_{ZC} - 1, \tag{1}$$

where N_{ZC} denotes the ZC sequence length. From uth root ZC sequence, the PID and BID preambles with ZCZ of length $N_{CS} - 1$ in RAP-1 are defined as follows:

$$x_{1,u}^{b,v}(n) = \left\{ x_{1,u}^{v}(n), x_{1,u}^{b}(n) \right\}, \tag{2}$$

where the two preambles are

$$\begin{aligned} x_{1,u}^{v}(n) &= x_u\left((n + C_v) \bmod N_{ZC}\right), \\ x_{1,u}^{b}(n) &= x_u\left((n + C_b) \bmod N_{ZC}\right) \end{aligned} \tag{3}$$

and cyclic shifts are defined as

$$C_v = \begin{cases} vN_{CS} & v = 0, 1, \ldots, \left\lfloor \dfrac{N_{ZC}}{N_{CS}} \right\rfloor - 1, \ N_{CS} \ne 0 \\ 0 & N_{CS} = 0 \end{cases}$$

$$C_b \tag{4}$$

$$= \begin{cases} bN_{CS} & b = 0, 1, \ldots, \left\lfloor \dfrac{N_{ZC}}{N_{CS}} \right\rfloor - 1, \ N_{CS} \ne 0, \ b < N_B \\ 0 & N_{CS} = 0. \end{cases}$$

Here, N_{CS}, C_v, C_b, and N_B denote the length of the cyclic shift in the ZCZ, the cyclic shift mapped onto the PID, the cyclic shift mapped onto the BID, and the number of beams, respectively. $u, b \in \mathbf{N}_B$, $v \in \mathbf{N}_V$, $\mathbf{N}_B = \{0, 1, \ldots, N_B - 1\}$ and $\mathbf{N}_V = \{0, 1, \ldots, \lfloor N_{ZC}/N_{CS} \rfloor\}$ denote the root index uniquely assigned to each cell, BID, PID, set of BIDs, and set of PIDs, respectively. The RAP-1 satisfies the CAZAC property because the structure of the preamble is the same as the RAP in LTE.

The RAP-2 can be viewed as a product of two Chu sequences with different root indices. The first root index is used for cell identification in the same way that the root index in PRACH is used. The BID is mapped onto the second root index. The PID is mapped onto the cyclic shift of the sequence. The RAP-2 is defined as

$$\begin{aligned} x_{2,u}^{b}(n) &= e^{j(\pi/N_{ZC})(un(n-1)+bn(n+1))}, \\ x_{2,u}^{b,v}(n) &= x_{2,u}^{b}\left((n + C_v) \bmod N_{ZC}\right). \end{aligned} \tag{5}$$

Here, the subscripts 1 and 2 denote RAP-1 and RAP-2, respectively. RAP-2 is designed such that the PID is mapped onto the sequence (with cell ID) in conjunction with its BID. Note that, in RAP-1, the PID is not associated with the BID because they are mapped onto different preambles. The autocorrelation function and frequency-domain version of RAP-2 are given by

$$\begin{aligned} R_{2,u}^{b}(l) &= \sum_{n=0}^{N_{ZC}-1} x_{2,u}^{b}(n) \left[x_{2,u}^{b}(n+l) \right]^* \\ &= \sum_{n=0}^{N_{ZC}-1} e^{j(\pi/N_{ZC})(2r_{ub}ln)} e^{j(\pi/N_{ZC})(ul(l-1)+bl(l+1))} \\ &= N_{ZC} e^{j(\pi/N_{ZC})(ul(l-1)+bl(l+1))} \delta\left(\bmod\left(l, N_{ZC}\right)\right), \end{aligned} \tag{6}$$

$$X_{2,u}^{b}[k] = x_{r_{ub}}^{*}\left[r_{ub}^{-1}k \right] X_{2,u}^{b}[0], \tag{7}$$

where

$$X_{2,u}^{b}[0] = \sqrt{N_{ZC}} \left(\frac{r_{ub}\alpha}{N_{ZC}} \right) \frac{1 - j^{N_{ZC}}}{1 - j} e^{-(j2\pi/N)r_{ub}\alpha\beta^2},$$

$$r_{ub} = u + b, \quad \alpha = \frac{N_{ZC}+1}{2}, \quad \beta = \frac{N_{ZC}-1}{2}. \tag{8}$$

Here, r_{ub}^{-1} and $\delta(\cdot)$ denote the multiplicative inverse of r_{ub} and the Kronecker delta function, respectively. $(r_{ub}\alpha/N_{ZC})$ denotes the Legendre symbol with a value of -1, 1, or 0. The frequency-domain version of RAP-2 in (7) can be obtained using the DFT property of the Chu Sequence [13]. The amplitude of the frequency-domain version of RAP-2 is given by $\sqrt{N_{ZC}}$, with a constant value for all values of n. As can be seen from (6), RAP-2 has a zero-autocorrelation property that has a peak amplitude of N if l is equal to zero. From (6) and (7), we can see that RAP-2 also satisfies the CAZAC property. However, this property holds only when the root index, r_{ub}, and sequence length are coprime.

To further analyze the proposed preambles, we examine their correlation properties. The correlation function of the first preamble in RAP-1 is given by

$$\begin{aligned} R_{1,u}\left(v, v'\right) &= \sum_{n=0}^{N_{ZC}-1} x_{1,u}^{v}(n) \left[x_{1,u}^{v'}(n) \right]^* \\ &= \sum_{n=0}^{N_{ZC}-1} e^{-j(\pi u(n+C_v)(n+C_v+1)/N_{ZC})} e^{j(\pi u(n+C_{v'})(n+C_{v'}+1)/N_{ZC})} \\ &= N_{ZC} e^{j(\pi u(C_{v'}(C_{v'}+1)-C_v(C_v+1))/N_{ZC})} \delta\left(\bmod\left(v, v'\right)\right). \end{aligned} \tag{9}$$

In (9), we observe the interference between preambles when two preambles with PIDs (v and v'), transmitted from MSs in the same cell, are arrived at the BS. The correlation function of the second preamble in RAP-1 is given by

$$
\begin{aligned}
R_{1,u}\left(b,b'\right) &= \sum_{n=0}^{N_{ZC}-1} x_{1,u}^{b}\left(n\right)\left[x_{1,u}^{b'}\left(n\right)\right]^{*} \\
&= \sum_{n=0}^{N_{ZC}-1} e^{-j(\pi u(n+C_b)(n+C_b+1)/N_{ZC})} e^{j(\pi u(n+C_{b'})(n+C_{b'}+1)/N_{ZC})} \quad (10) \\
&= N_{ZC} e^{j(\pi u(C_{b'}(C_{b'}+1)-C_b(C_b+1))/N_{ZC})} \delta\left(\mathrm{mod}\left(b,b'\right)\right).
\end{aligned}
$$

(10) shows the interference between preambles when two preambles with BIDs (b and b'), transmitted from MSs in the same cell, are arrived at the BS. The results show that the preambles (RAP-1) in the same cell (same root index) are orthogonal, implying that there is no interference when BIDs or PIDs in the same cell are different. When two preambles with PIDs (v and v'), transmitted from MSs in different cells (different root indices), arrive at the BS, the interference between preambles is given by

$$
\begin{aligned}
R_{1,u,u'}\left(v,v'\right) &= \sum_{n=0}^{N_{ZC}-1} x_{1,u}^{v}\left(n\right)\left[x_{1,u'}^{v'}\left(n\right)\right]^{*} \\
&= \sum_{n=0}^{N_{ZC}-1} e^{-j(\pi u(n+C_v)(n+C_v+1)/N_{ZC})} e^{j(\pi u'(n+C_{v'})(n+C_{v'}+1)/N_{ZC})} \quad (11) \\
&= e^{j(\pi u' C_{v'}{}^2-u C_v{}^2-(u'-u)^{-1}(u'C_{v'}-uC_v)^2/N_{ZC})}\Sigma^{u,u'}.
\end{aligned}
$$

When two preambles with BIDs (b and b'), transmitted from MSs in different cells, arrive at the BS, the interference between preambles is given by

$$
\begin{aligned}
R_{1,u,u'}\left(b,b'\right) &= \sum_{n=0}^{N_{ZC}-1} x_{1,u}^{b}\left(n\right)\left[x_{1,u'}^{b'}\left(n\right)\right]^{*} \\
&= \sum_{n=0}^{N_{ZC}-1} e^{-j(\pi u(n+C_b)(n+C_b+1)/N_{ZC})} e^{j(\pi u'(n+C_{b'})(n+C_{b'}+1)/N_{ZC})} \quad (12) \\
&= e^{j(\pi u' C_{b'}{}^2-u C_b{}^2-(u'-u)^{-1}(u'C_{b'}-uC_b)^2/N_{ZC})}\Sigma_{u,u'},
\end{aligned}
$$

where

$$
\begin{aligned}
&\Sigma_{u,u'} \\
&= \sqrt{N_{ZC}} \begin{cases} \left(\dfrac{\left|\left(u'-u\right)\right|\alpha}{N_{ZC}}\right)\dfrac{1-j^{N_{ZC}}}{1-j}e^{-(j2\pi/N_{ZC})(u'-u)\alpha\beta^2}, & u' > u \\[2mm] \left(\dfrac{\left|\left(u'-u\right)\right|\alpha}{N_{ZC}}\right)\dfrac{1+j^{N_{ZC}}}{1+j}e^{-(j2\pi/N_{ZC})(u'-u)\alpha\beta^2}, & u' < u, \end{cases} \quad (13)
\end{aligned}
$$

$$
\alpha = \frac{N_{ZC}+1}{2}, \ \beta = \frac{N_{ZC}-1}{2}.
$$

Equations (11) and (12) are derived using Gauss sum expression [14]. The results show that the preambles (RAP-1), which are transmitted from MSs in different cells, are not orthogonal, and the amplitude of the correlation is given by $\sqrt{N_{ZC}}$. In a similar manner, the correlation functions of RAP-2 with different PIDs or BIDs, which are transmitted from MSs in the same cell or different cells, are given by

$$
R_{2,u}\left(v,v'\right) = \sum_{n=0}^{N_{ZC}-1} x_{2,u}^{b,v}\left(n\right)\left[x_{2,u}^{b,v'}\left(n\right)\right]^{*} = N_{ZC} e^{(j\pi/N_{ZC})((C_v-C_{v'})u(C_v+C_{v'}-1))} e^{(j\pi/N_{ZC})((C_v-C_{v'})b(C_v+C_{v'}+1))}\delta\left(\mathrm{mod}\left(v,v'\right)\right), \quad (14)
$$

$$
R_{2,u}\left(b,b'\right) = \sum_{n=0}^{N_{ZC}-1} x_{2,u}^{b,v}\left(n\right)\left[x_{2,u}^{b',v}\left(n\right)\right]^{*} = \sum_{n=0}^{N_{ZC}-1} e^{(j\pi/N)((b-b')(n+C_v)(n+C_v+1))} = \begin{cases} N_{ZC}, & b = b' \\ \Sigma^{b,b'}, & b \neq b', \end{cases} \quad (15)
$$

$$
R_2\left(u,u'\right) = \sum_{n=0}^{N_{ZC}-1} x_{2,u}^{b,v}\left(n\right)\left[x_{2,u'}^{b,v}\left(n\right)\right]^{*} = \sum_{n=0}^{N_{ZC}-1} e^{(j\pi/N)((u-u')(n+C_v)(n+C_v-1))} = \begin{cases} N_{ZC}, & u = u' \\ \Sigma^{u,u'}, & u \neq u'. \end{cases} \quad (16)
$$

The results show that the preambles (RAP-2), which are transmitted from MSs in different cells, are not orthogonal regardless of their BID and PID values, and the amplitude of the correlation is given by $\sqrt{N_{ZC}}$. When two preambles (RAP-2) with different BIDs (b and b') are transmitted from MSs in the same cell, the amplitude of the correlation is given by $\sqrt{N_{ZC}}$. However, the preambles (RAP-2) with different PIDs in the same cell are orthogonal, implying that the detection probability at the BS will be high when preambles (RAP-2) with different PIDs are transmitted from MSs in the same cell.

4. Preamble Detection

In this section, we describe a detection technique for RAP in mm-wave cellular systems. As discussed in Step 2 of the proposed RA procedure, the proposed preamble will be transmitted repeatedly from the MS by sweeping Tx and Rx beams. The preamble received at the BS contains the information on the best downlink Tx BID (obtained from Step 1) and PID selected from the MS. The BS needs to detect the BID and PID from the received preambles in multiuser and multicell environments. When the RAPs are received

from multiple MSs in neighboring cells, the received signal at the serving BS with bth Rx beam is given in the frequency domain as follows:

$$
\begin{aligned}
Y_{t,b}^{s}(k) &= \sum_{i=1}^{N_I} \sum_{m=1}^{N_M} G_{\mathrm{Tx},c}^{i,m} G_{\mathrm{Rx},b} X_{t,u_{i,m},c}^{b_{DL}}(k) H_{c,b}^{i,m}(k) \\
&\quad + W(k) \\
&= \sum_{m=1}^{N_M} G_{\mathrm{Tx},c}^{s,m} G_{\mathrm{Rx},b} X_{t,u_{i,m},c}^{b_{DL}}(k) H_{c,b}^{s,m}(k) \\
&\quad + \sum_{i=1,i\neq s}^{N_I} \sum_{m=1}^{N_M} G_{\mathrm{Tx},c}^{i,m} G_{\mathrm{Rx},b} X_{t}^{u_{i,m}}(k) H_{c,b}^{i,m}(k) \\
&\quad + W(k).
\end{aligned} \tag{17}
$$

Here, $G_{\mathrm{Tx},c}^{i,m}$, $G_{\mathrm{Rx},b}$, and $H_{c,b}^{i,m}(k)$ are the Tx beam gain at cth beam of mth MS in ith cell, the Rx beam gain at bth beam of the BS, and the channel frequency response between the cth beam of the mth MS in the ith cell and the bth beam of the serving BS, s, where $c \in \mathbf{N_C}$, $\mathbf{N_C} = \{0.1, \ldots, N_C - 1\}$. $X_{t,u_{i,m},c}^{b_{DL}}(k)$ and $W(k)$ denote the RAP transmitted from the mth MS in ith cell carrying the best TX downlink beam, $b_{DL} \in N_B$, which is transmitted from cth MS beam and additive white Gaussian noise (AWGN) at kth subcarrier, respectively. Further, $u_{i,m}$ and $u_{s,m}$ denote Chu's root indices of ith cell and serving cell for mth MS, respectively. N_M and N_I denote the number of MSs and neighboring cells, respectively. $t = \{1, 2\}$ denotes the type of proposed RAP. The received signal can be classified as two parts: the signal received from MSs in the serving cell and interference signals from MSs in neighboring cells.

The preamble-detection process at the BS is composed of a power-delay profile (PDP) computation and signature detection. The PDP computation of the received signal with a specific beam is given by

$$
\left| z_{t,u}^{b,v}(l) \right|^2 = \left| \sum_{n=0}^{N_{ZC}-1} y(n) \left[x_{t,u}^{b,v}(n+l) \right]^* \right|^2, \quad t = 1, 2, \tag{18}
$$

where $z_{t,u}^{b,v}(l)$ represents the correlation function between the received signal $y(n)$ and reference RAP $x_{t,u}^{v,b}(n)$ with lag l. In (18), the variables for Tx and Rx beams are not included for notational simplicity. The signature detection is made by searching the PDP peaks above a detection threshold over a search window corresponding to the cell size. The detection threshold is determined by the false-alarm probability. From the PDP peaks, the PID and TA can be computed as

$$
\begin{aligned}
\mathrm{PID} &= \left\lfloor \frac{N_{\mathrm{peak}}}{N_{\mathrm{CS}}} \right\rfloor + 1, \\
\mathrm{TA} &= \mathrm{mod}\left(N_{\mathrm{peak}}, N_{\mathrm{CS}}\right) - 1,
\end{aligned} \tag{19}
$$

where N_{peak} denotes the position of the PDP peak. The preamble detection can be performed in the frequency domain, as shown in Figure 3. First, the received signals are demapped onto the corresponding subcarriers in the frequency domain. Next, the known preamble sequences (RAP-1 or RAP-2) in the time domain are converted to the frequency domain using a discrete Fourier transform (DFT) operation. After multiplication of these two signals, the result is transformed to the time domain by inverse DFT (IDFT). Then, the PID and BID are detected using the PDP in the time domain.

A preamble transmission is detected if the PDP peak is greater than the detection threshold. This process will be performed repeatedly for each Tx-Rx beam pair after evaluating the PDP of each Tx-Rx beam pair with the known RAP. Finally, the BID is detected as

$$
\bar{b} = \begin{cases} \displaystyle\arg\max_{b \in N_B} \left| z_{t,u}^{b,v}(l) \right|^2 \\ \displaystyle\arg\max_{b \in N_B} \left| z_{t,u}^{b,v}(l) \right|^2. \end{cases} \tag{20}
$$

The preamble-detection probability depends on the detection threshold setting. In LTE systems, the detection threshold is determined by the target false-alarm probability [15, 16]. Here, the false-alarm probability is defined as the probability that the receiver detects a preamble transmission when the received signal is purely noise. When there is no preamble transmission, the PDP of the received signal follows a central chi-square distribution with two degrees of freedom (DoFs), with the mean given by the noise-floor level. The false-alarm probability can be easily obtained from the cumulative distribution function (CDF) by changing the threshold value. Then, the detection threshold can be found by setting the target false-alarm probability. The false-alarm probability for an LTE system with multiple antennas can be derived by accumulating PDPs obtained from multiple antennas. The detection probability can be found using the detection threshold obtained from the false-alarm probability. Here, the detection probability is defined as the probability that the receiver correctly detects the RAP transmitted from the MS. When the proposed RAP is received in mm-wave cellular systems with directional beams, the false-alarm probability in an AWGN channel is given by [15, 16]

$$
P_t^{\mathrm{FA}} = \left[1 - \left[\left(1 - e^{-T_r} \sum_{b=0}^{N_B-1} \frac{1}{b!} \left(T_r\right)^b \right)^{N_{ZC}} \right]^{N_M} \right]^n, \tag{21}
$$

where

$$
n = \begin{cases} 2 & \text{if } t = 1 \\ 1 & \text{if } t = 2. \end{cases} \tag{22}
$$

Here, T_r denotes the threshold value normalized by the noise-floor level. The false-alarm probabilities are given for the cases of RAP-1 and RAP-2. Note that, in the cases of RAP-1 consisting of two symbols, the false-alarm probabilities of PID and BID are multiplied to obtain the total false-alarm probability. From (21), we can obtain the plot of the false-alarm probability by changing the threshold value. From
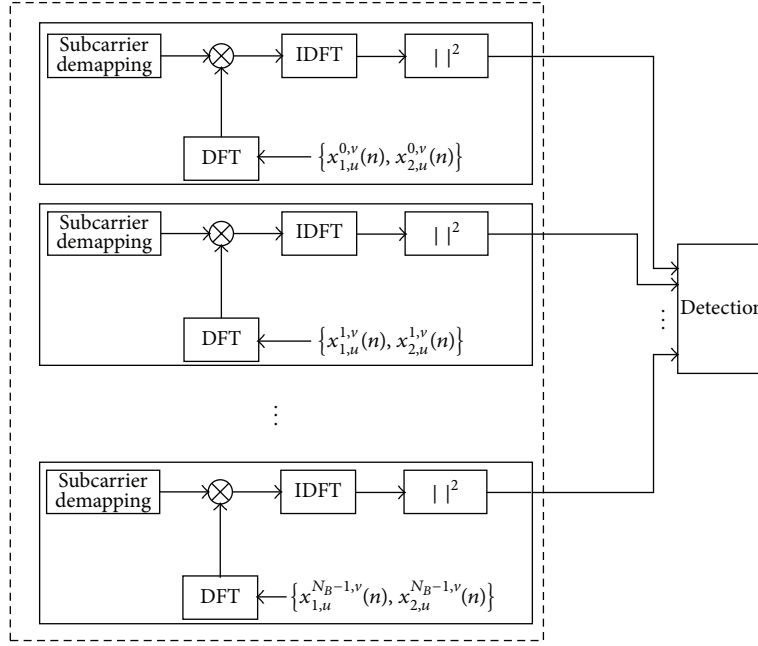
FIGURE 3: Detection process of proposed preamble.

the plot, we determine the detection threshold by setting the target false-alarm probability. Using the value of the detection threshold, the detection probabilities of RAP-1 and RAP-2 can be obtained as [15, 16]

$$P_1^D = \left(1 - \frac{1}{N_B}\right)^{N_{MS}-1} \left[Q_1\left(\sqrt{\delta}, \sqrt{\lambda}\right)\right]^2,$$

$$P_2^D = Q_1\left(\sqrt{\delta}, \sqrt{\lambda}\right), \tag{23}$$

where δ and λ denote the noncentrality parameter and detection threshold, respectively. $Q_n(\cdot)$ represents the Marcum Q-function. The performance of the proposed RAP in mm-wave cellular systems with beamforming can be evaluated using the false-alarm probability and detection-probability functions.

5. Simulation

We evaluated the performance of the proposed RAP by performing computer simulations using a simple model of a mm-wave cellular system with directional beams. Before applying the proposed RAP to the mm-wave cellular system, we verified the properties of the proposed RAP described in Section 2. Figures 4 and 5 show the correlation properties of RAP-1 and RAP-2, respectively. Here, we assumed that the sequence length is 839, and the PID and BID of MS1 under the test are set to 44 and 2, respectively. The values of the PID and BID of MS2 are assumed to vary. Scenario 1 refers to the case where two MSs (MS1 and MS2) are located in the serving cell. In Scenario 2, it is assumed that MS1 is located in the serving cell and MS2 is located in the neighboring cell. As can be seen from Figure 4(a), the correlation value of RAP-1 in Scenario 1 is zero, with the exception of the case where MS2 has the same PID = 44. In this case, the magnitude of

the correlation becomes 839 (N_{ZC}), as given by (9). As can be seen from Figure 4(b), we can obtain the same result, with the exception of the case where BID = 2, as given by (10). However, the correlation value of RAP-1 in Scenario 2 is 28.97 ($\sqrt{N_{ZC}}$) for all values of PID and BID, as given by (11) and (12), respectively. We can confirm this result from Figures 4(a) and 4(b). As can be seen from Figure 5(a), the correlation value of RAP-2 in Scenario 1 is zero, except for the case where MS2 has the same PID = 44. In this case, the magnitude of the correlation becomes 839, as given by (14). As can be seen from Figure 5(b), the correlation value of RAP-2 in Scenario 1 becomes 839 only when we use the same value of BID (2). However, the correlation value becomes 28.97 when we use another value of BID, as given by (15). The correlation value of RAP-2 in Scenario 2 is 28.97 for all values of PID and BID, as given by (16).

Figures 6(a) and 6(b) show false-alarm probabilities of RAP-1 and RAP-2, respectively. We obtain the false-alarm probability using both the analytic solution in (21) and the simulation when the number of MSs varies. The PDP received at the BS is compared to the normalized threshold value, which varies from 10 to 30. The threshold value for preamble detection, which satisfies the required false-alarm probability, can be found from Figure 6. The value of the false-alarm probability is set to be less than 0.1%, as in LTE systems. As can also be seen in Figure 6, the threshold value satisfying the required false-alarm probability is sensitive to the number of users in the cell. The threshold value needs to be increased to avoid the misdetection of RAP as the number of MSs increases. We also see that the simulation results agree well with the analytical results in (21). From these results, we determined the threshold values for the detection probability as follows: 22 for 1 MS, 24 for 3 MSs, and 26 for 5 MSs in the case
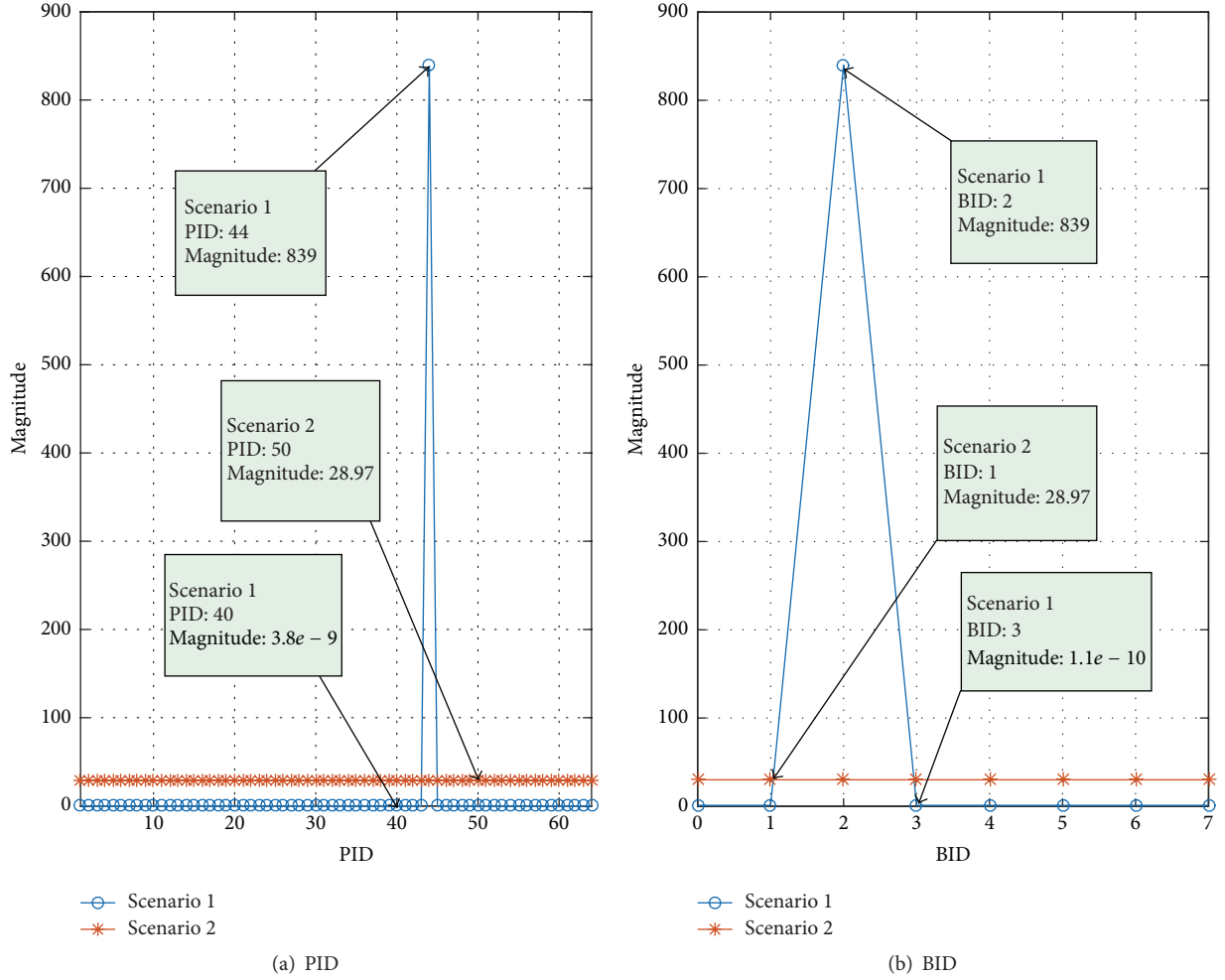
(a) PID

(b) BID

FIGURE 4: Correlation property of RAP-1.

of RAP-1, and 26 for 1 MS, 30 for 3 MSs, and 38 for 5 MSs in the case of RAP-2.

Figures 7(a) and 7(b) show the detection probabilities of RAP-1 and RAP-2, respectively. The threshold values selected from the previous results are used for preamble detection. For the simulation, we used a three-dimensional spatial channel model (3D SCM), which is often used for the system simulation of cellular systems, with the SNR varying from −40 to −18 dB with a step size of 2 dB. We set the carrier frequency, subcarrier spacing, and FFT size to 28 GHz, 270 KHz, and 2048, respectively. The number of antenna elements is 16 at both the BS and MS, all with a uniform rectangular array (URA). The positions of the MSs are randomly placed in a cell, and the best Tx-Rx beam directions are derived from their respective positions. PIDs are assigned as [40 44 26 17 10] when the number of MSs is five. As can be seen from Figure 7, the detection probability tends to degrade as the number of MSs increases. That is because of the increased interference as the number of RAPs received at the BS increases.

However, the detection probability of RAP-1 in Figure 7(a) degrades more significantly compared with that

of RAP-2. In the case of RAP-1, the detection probability cannot reach 100%, even in high SNR regions when the number of MSs is larger than three. This may result in system failure because of the misdetection of the PID or BID. This phenomenon results from the structure of RAP-1 being composed of two preambles. Because the PID and BID are detected separately, it is difficult to determine the relationship between the PID and BID. It is not possible to know whether the detected PID and BID are from the same source (MS) in the case of RAP-1. On the other hand, the detection probability of RAP-2 in Figure 7(b) is 100% in higher SNR regions when the number of MSs is five. This is because the PID and BID in RAP-2 are related to each other. Once the preamble is detected, the PID and BID are found simultaneously in the case of RAP-2. From Figure 7, it can also be seen that simulation results agree well with the analytical results in (23). From these results, we can see that the RAP-2 is appropriate for the detection of RAP in multiuser environments.

Table 1 compares the processing time and number of complex multiplications when we used the conventional and proposed RA procedures. Here, we excluded the downlink
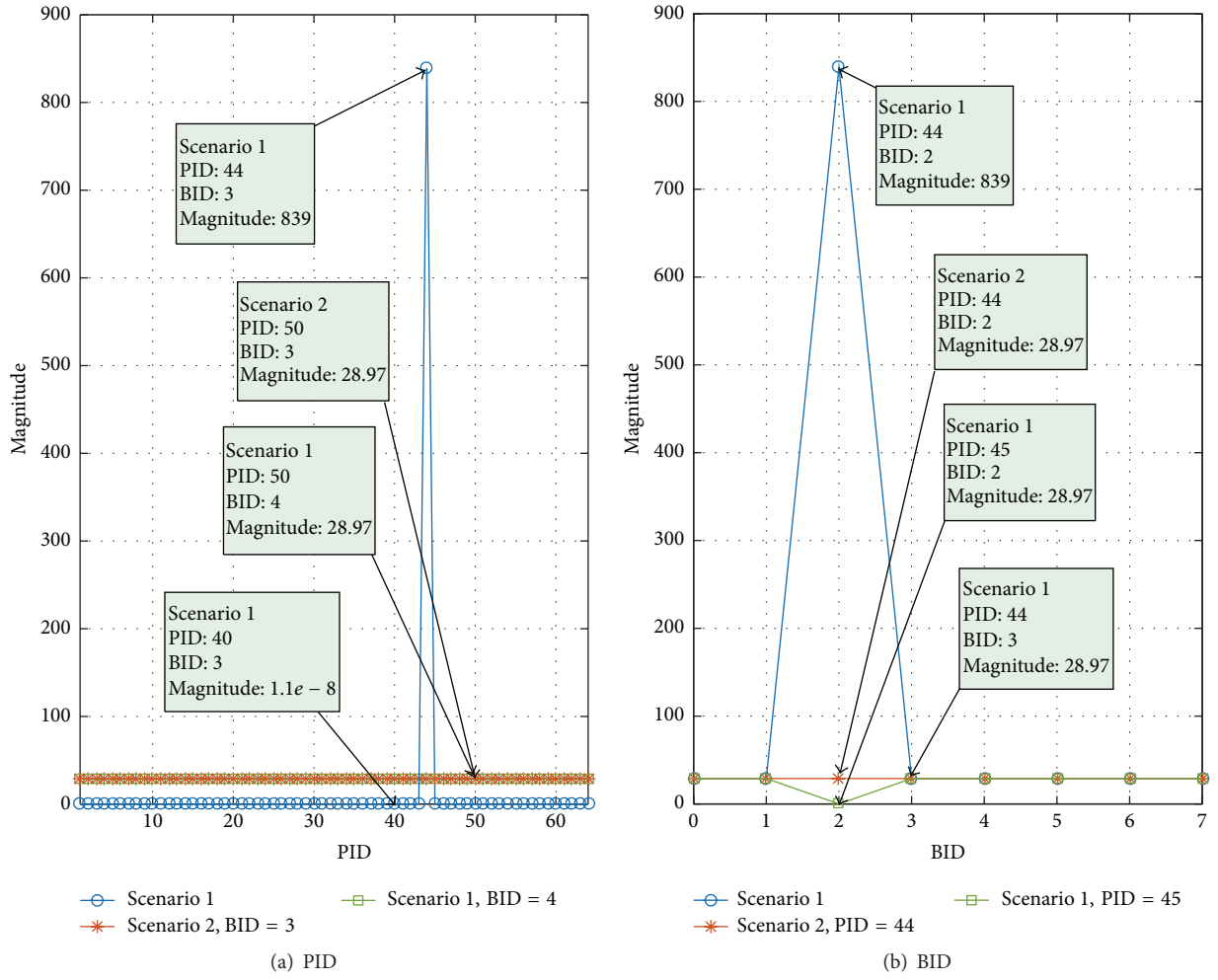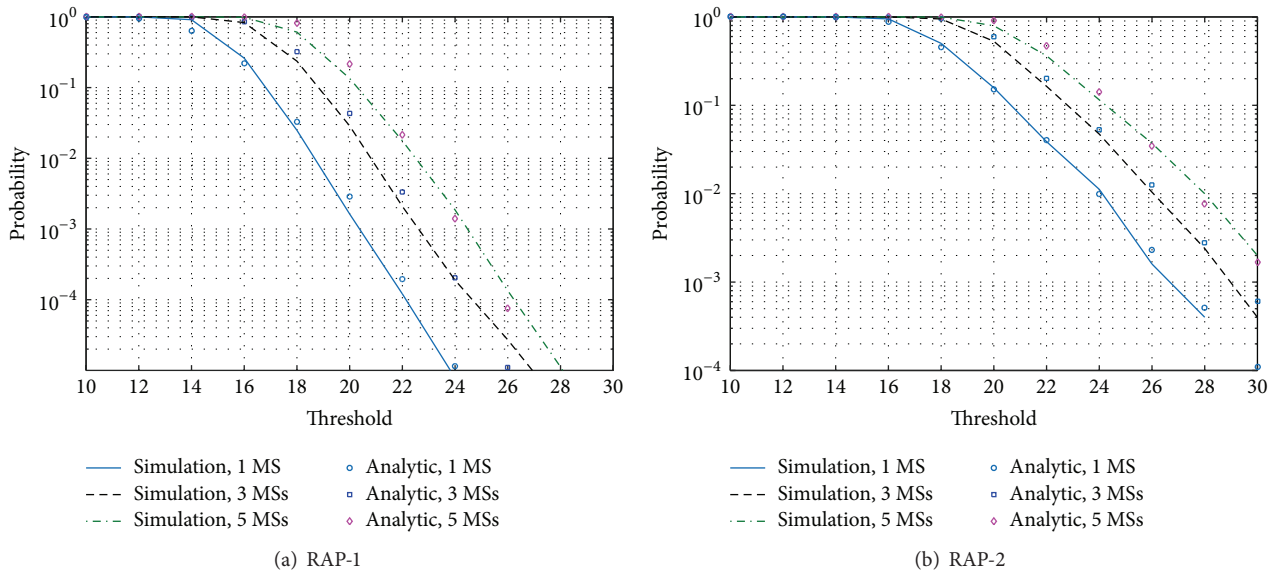
Figure 5: Correlation property of RAP-2.

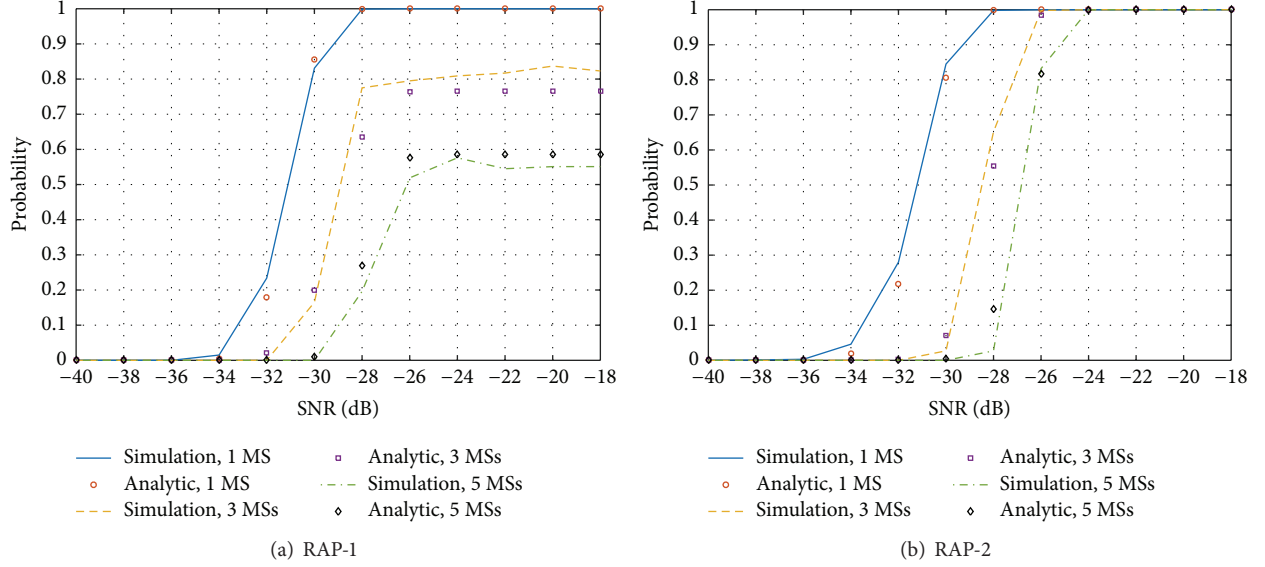Figure 6: False-alarm probabilities of RAP-1 and RAP-2.

(a) RAP-1

(b) RAP-2

Figure 7: Detection probabilities of RAP-1 and RAP-2.

Table 1: Comparison of conventional and proposed RA procedures.

Item	Conventional RA procedure	Proposed RA procedure	
		RAP-1	RAP-2
Processing time (μs)	$T_{\text{RAP}} \cdot (2 \cdot N_B \cdot N_C + N_B)$ (108, 800)	$2 \cdot T_{\text{RAP}} \cdot (N_B \cdot N_C)$ (102, 400)	$T_{\text{RAP}} \cdot (N_B \cdot N_C)$ (51, 200)
Number of complex multiplications	$N_B \cdot N_C \cdot N_{\text{ZC}}^2$ (45, 050, 944)	$2 \cdot N_B \cdot N_C \cdot N_{\text{ZC}}^2$ (90, 101, 888)	$N_B \cdot N_C \cdot N_{\text{ZC}}^2$ (45, 050, 944)

synchronization step (Step 1) because this step is common to both procedures. In addition, we do not consider the steps that do not require beam sweeping (Step 5 and Step 6 in the conventional RA procedure, and Step 3 and Step 4 in the proposed RA procedure) because the processing time and complexity are minimal. In the conventional RA procedure, the processing time required for Step 2 and Step 4 is given by the product of the number of BS beams (N_B) and the number of MS beams (N_C). In addition, the processing time required for Step 3 is given by the number of BS beams. For the proposed RA procedure, the processing time required for Step 2 is given by the product of the number of BS beams and the number of MS beams. Here, an example is shown in parenthesis when $N_B = 8$ and $N_C = 8$. The RAP duration T_{RAP} is assumed to be $800\,\mu$s, as in LTE systems. As can be seen in Table 1, the RAP-1 requires twice the processing time when compared to RAP-2 because two sequences are used in RAP-1. Note that RAP-2 requires half of the processing time when compared with the conventional technique. Next, we compared the computational complexities required for preamble detection. In order to detect the RAP, we require complex multiplications equal to N_{ZC}^2 for a correlator with the size of N_{ZC}. The total number of complex multiplications required for preamble detection in Step 2 (conventional technique and RAP-2) is given by $N_B \cdot N_C \cdot N_{\text{ZC}}^2$ because the detection process needs to be performed in all directions.

The number of complex multiplications required for the RAP-1 is twice that of the conventional one because the RAP-1 uses two preambles.

6. Conclusion

In this paper, we proposed two different types of preambles (RAP-1 and RAP-2) to reduce the processing time required for RA in mm-wave cellular systems with directional beams. Using simulations, we showed that the proposed RA procedure with the preamble (RAP-1 or RAP-2) can reduce the processing time, compared with the conventional RA procedure. However, in the case of RAP-1, the detection probability decreases significantly as the number of MSs increases, because the PID and BID are detected separately. In the case of RAP-2, a 100% detection probability can be achieved in high SNR regions even when the number of MSs is five, because the PID in RAP-2 is designed in association with the BID. Also, 6% and 53% reduction gains in processing time can be obtained when RAP-1 and RAP-2 are used, respectively. Hence, the RAP-2 was shown to be better suited for RA in mm-wave cellular systems with directional beams because of the shorter processing time, high detection probability in multiuser environments, and reasonable computational complexity (similar to the conventional one).

Competing Interests

Acknowledgments

References

[1] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: potentials and challenges," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 366–385, 2014.

[2] T. S. Rappaport, F. Gutierrez, E. Ben-Dor, J. N. Murdock, Y. Qiao, and J. I. Tamir, "Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor Urban cellular communications," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 4, pp. 1850–1859, 2013.

[3] W. Hong, K.-H. Baek, Y. Lee, Y. Kim, and S.-T. Ko, "Study and prototyping of practically large-scale mmWave antenna systems for 5G cellular devices," *IEEE Communications Magazine*, vol. 52, no. 9, pp. 63–69, 2014.

[4] M. Bennis, S. M. Perlaza, P. Blasco, Z. Han, and H. V. Poor, "Self-organization in small cell networks: a reinforcement learning approach," *IEEE Transactions on Wireless Communications*, vol. 12, no. 7, pp. 3202–3212, 2013.

[5] A. T. Nassar, A. I. Sulyman, and A. Alsanie, "Radio capacity estimation for millimeter wave 5G cellular networks using narrow beamwidth antennas at the base stations," *International Journal of Antennas and Propagation*, vol. 2015, Article ID 878614, 6 pages, 2015.

[6] B. Li, Z. Zhou, H. Zhang, and A. Nallanathan, "Efficient beamforming training for 60-GHz millimeter-wave communications: a novel numerical optimization framework," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 2, pp. 703–717, 2014.

[7] S. K. Yong, P. Xia, and A. Valdes-Garcia, *60 GHz Technology for Gbps WLAN and WPAN: From Theory to Practice*, John Wiley & Sons, New York, NY, USA, 2011.

[8] C. Jeong, J. Park, and H. Yu, "Random access in millimeter-wave beamforming cellular networks: issues and approaches," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 180–185, 2015.

[9] M. Lott, "Adaptive random access with beam-forming in 4G mobile networks," in *Proceedings of the IEEE 63rd Vehicular Technology Conference (VTC-Spring '06)*, pp. 435–439, Melbourne, Australia, July 2006.

[10] E. Dahlman, S. Parkvall, and J. Skold, *4G LTE/LTE-Advanced for Mobile Broadband*, Elsevier, Oxford, UK, 2011.

[11] S. Sesia, I. Toufik, and M. Baker, *LTE—The UMTS Long Term Evolution: From Theory to Practice*, John Wiley & Sons, Chichester, UK, 2011.

[12] LTE Standard, *LTE; Evolved Universal Terrestrial Radio Access (EUTRA); Physical Channels and Modulation (3GPP TS 36.211 Version 12.6.0 Release 12)*, 2015.

[13] B. M. Popovic, "Efficient DFT of Zadoff-Chu sequences," *Electronics Letters*, vol. 46, no. 7, pp. 502–503, 2010.

[14] B. C. Berndt, R. J. Evans, and K. S. Williams, *Gauss and Jacobi Sums*, Canadian Mathematical Society Series of Monographs and Advanced Texts, John Wiley & Sons, New York, NY, USA, 1998.

[15] O. Altrad and S. Muhaidat, "A new mathematical analysis of the probability of detection in cognitive radio over fading channels," *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, no. 1, article 159, 11 pages, 2013.

[16] K. Simon, *Probability Distribution Involving Gaussian Random Variables: A Handbook for Engineers and Scientists*, Springer, Berlin, Germany, 2002.

Research Article

Adaptive Access Class Barring Method for Machine Generated Communications

Jaesung Park[1] and Yujin Lim[2]

[1]*Department of Information Security, University of Suwon, San 2-2, Wau-ri, Bongdam-eup, Hwaseong, Gyeonggi-do 445-743, Republic of Korea*
[2]*Department of Information Technology Engineering, Sookmyung Women's University, Cheongpa-ro 47-gil 100, Yongsan-gu, Seoul 04310, Republic of Korea*

Correspondence should be addressed to Yujin Lim; yujin91@sookmyung.ac.kr

Cellular network is provisioned to serve traffic demands generated by human being. The random access channel used for nodes to compete for a connection with an eNB is limited. Even though machines generate very small amount of data traffic, the signaling channel of a network becomes overloaded and collisions occur to fail the access if too many MTC (Machine Type Communication) devices attempt to access network. To tackle the issue, 3GPP specifies an access class barring but leaves a specific algorithm as an implementation issue. In this paper, we propose an adaptive access barring method. Generally, an eNB does not know the number of MTC devices in its coverage area. Thus, it is difficult to control the barring factor by predicting the number of MTC devices in a service area of a cell. On the contrary, we control the barring factor based on the prediction of access intensity which can be measured at an eNB. Simulation results show that since the proposed method can manipulate the barring factor autonomously according to the access intensity, it is superior to the original method in terms of the access success probability and the collision probability.

1. Introduction

MTC (Machine Type Communication) application is one of automated applications which provide connectivity between a tremendous number of embedded devices in LTE-A (Long Term Evolution-Advanced) [1, 2]. These embedded devices support a wide range of applications such as smart grid/smart metering, e-health, surveillance and security, and consumer electronics. Industry reports expect the significant potential of MTC applications to generate considerable revenues. In MTC, a large number of devices generate a huge amount of signaling or data flow to access RAN (Random Access Network) in a short time. Concurrent accesses from a large number of devices result in significant congestion. The system suffers from packet collision, long access delay, or low service quality of application. To alleviate the congestion, system needs to control the excessive connectivity from devices [3]. Thus, self-adaptation is a key building block of MTC application because it adaptively controls the strategic parameters to maximize the system performance in dynamic environment.

When a device is activated to send its data, it tries to reserve uplink contention-based channels, that is, RACHs (Random Access CHannels). In LTE-A, multichannel slotted ALOHA systems are used to model the random access channel. The two kinds of random access traffic arises from MTC devices: the traffic which is occurred by an external event and the traffic which periodically occurred to communicate between a device and a system in a synchronized manner. For example, for electricity meter reading, a device needs to report periodically electricity usage.

When a device tries to send data, it follows a random access 4-step procedure to reserve uplink channel in Figure 1. Firstly, a device selects a preamble randomly from a set of random access preambles. The selected preamble is transmitted to allow eNB to estimate the transmission timing of a device. In the second step, RAR (Random Access Response) is sent by eNB. It involves a timing advance command to adjust a device transmit timing and assigns uplink resources to a device. In the third step, a device transmits a message with identity information. If a preamble collision occurred in the first
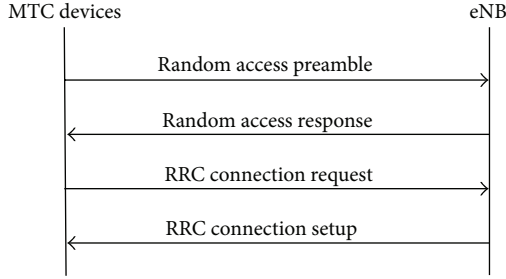
FIGURE 1: The random access procedure.

step, the colliding devices transmit identity information within the same time-frequency resource block. In this case, collision occurs at eNB. In the final step, eNB transmits a contention-resolution message. If packet collision occurs and eNB decodes one of collided packets, eNB acknowledges the device whose packet is successfully received. Unacknowledged devices fail the random access in current access opportunity and they will try again in the next access opportunity.

Under slotted ALOHA scheme, if the number of devices is controlled to access a system at the same time, system performance can be improved. Thus, eNB needs to optimally maintain the number of devices which simultaneously access a system. 3GPP (3rd Generation Partnership Project) has discussed some solutions such as ACB (Access Class Barring) and EAB (Extended Access Barring) [4]. Using ACB, devices defer their random access requests with a probability p. In other words, eNB periodically broadcasts barring parameters in SIB (System Information Block). Barring parameters involve a barring factor ($0 \leq p \leq 1$) and a barring duration. Each device generates a random number between 0 and 1 and compares the number with p. If the random number is equal to or greater than p, the access is barred for the barring duration. In EAB, devices are classified into a set of ACs (Access Classes). An eNB periodically broadcasts SIB with a barring bitmap which indicates the barred ACs. When a device initiates a random access procedure, a device checks whether its AC is barred. The barred devices postpone their random access procedures until the next SIB is changed.

In ACB and EAB, it is crucial to control the strategic parameters such as barring factor or barring bitmap. Even though many researchers try to solve the control problem, it is still open issue for further research [5]. Therefore, we propose a control method to determine a barring factor autonomously in ACB scheme to improve system performance. The experimental results show that our method improves system performances in terms of access success probability and collision probability.

The rest of the paper is organized as follows: in Section 2, we describe the related work. In Section 3, we present our control method to manipulate a barring factor in ACB. In Section 4, simulation results are summarized. In Section 5, conclusions and future research plan are drawn.

2. Related Work

There are N orthogonal preambles in one random access opportunity (time slot). If more than one device send the same preamble in one slot, a collision happens. When a collision happens, a device tries again with a randomly chosen preamble in a new slot. As the number of devices increases, the congestion could be worse and system performance deteriorates.

To alleviate the congestion due to random access collision, control mechanisms have been proposed. Many mechanisms are presented to regulate the access opportunity of devices [6]. Backoff-based mechanisms distribute the access attempts from devices in time slots by using backoff window. Admission control-based mechanisms reject the access attempts based on the average admission rate [7]. ACB-based mechanisms distribute the access attempts based on the barring factor. It effectively reduces the collision probability under large transmissions at the same time. Among them, we focus on ACB-based mechanism to autonomously manipulate the barring factor. There is a trade-off between barring factor p and LTE performance like resource utilization or average access delay. If the barring factor is smaller, fewer devices are allowed to access RACH. However, the access delay increases and resource utilization decreases. Thus, eNB needs to effectively control the barring factor under the current network condition. In [8], the barring factor is determined to maximize the expected number of successfully accessed devices in each slot. In [9], the factor is controlled to minimize the total service time which is the time for all devices to successfully access to eNB. In [10], an analytical model is proposed to minimize the access delay of all the devices. However, these mechanisms suffer from the high computational complexity to optimize the performance metrics. To reduce the complexity, approximation algorithms are also proposed to control the barring factor.

In an approximate approach, the current traffic load or predicted traffic load is mostly used to adaptively control the barring factor. In [11], the probability of packet transmission is calculated based on the current traffic load. Then, eNB determines the barring factor to improve network throughput. In [12], cooperative mechanism to control congestions among cells is proposed. eNBs cooperate to decide the barring factor in each cell for global stabilization and access load sharing. In [13], the number of contending devices is estimated using the Markov Chain with collision status. With the estimated load, the barring factor is determined. These conventional mechanisms estimate the number of contending devices by using traditional traffic prediction techniques.

3. Proposed Method

In this section, we describe an adaptive access barring method. The purpose of the method is to control the barring factor according to the estimated access intensity of MTC devices. The access intensity is defined as the number of random access attempts made during a certain time period.

An optimal predictor should know the exact correlation structure between the past measurements and the future

Barring factor controller

Access intensity measurement

X_i

Access intensity quantization

L_{i-1}, L_i

Transition matrix manipulation

$E[L_{i+1}]$

Barring factor generation

$p = 1 - \dfrac{E[L_{i+1}]}{M}$

At the end of a time slot

eNB

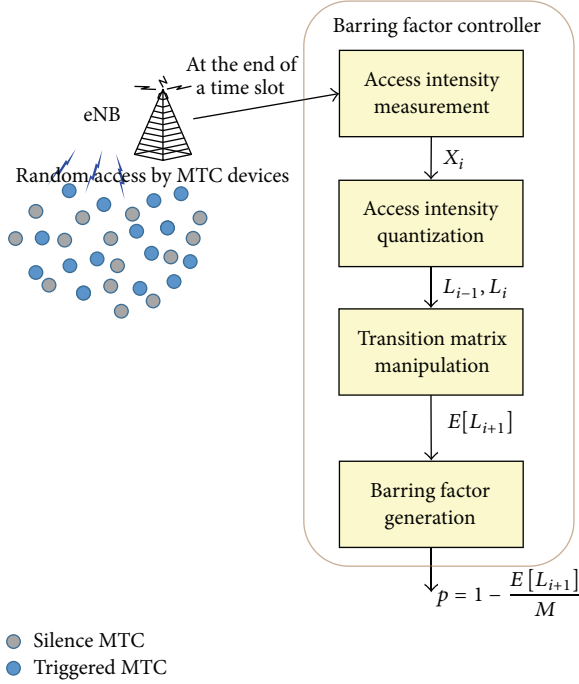Random access by MTC devices

○ Silence MTC
● Triggered MTC

FIGURE 2: Barring factor controller.

measurements. However, it is not possible to know the future measurements. Thus, we design a practical predictor based on the past measurements. Figure 2 shows the block diagram of the barring factor controller. The controller is composed of four distinct modules: access intensity measurement module, access intensity quantization module, transition matrix manipulation module, and barring factor generation module.

The access intensity measurement module measures the access intensity during a certain time period. We assume that time is divided into a fixed time slot, the size of which is t_s. An eNB measures the access intensity during each time slot. We denote by X_i the access intensity during the ith time slot.

The access intensity quantization module discretizes the measured access intensity into M access intensity levels. We denote by A_M the maximum number of access intensities that a system allows. Then, X_i is an integer and can be any value in $[0, A_M]$. However, as A_M becomes larger, atomizing X_i with the finest granularity demands more system resources for an eNB to maintain X_is for predicting future access intensity without improving system performance. For example, if A_M is sufficiently large, $X_i = x$ and $X_i = x + 1$ do not mean much to an eNB. Thus, we divide the range $[0, A_M]$ into M access intensity levels. The height of each level is $\Delta = A_M/M$, where M is an operating parameter controlling the granularity of X_i. We assign quantized values from $\Delta/2$ to $(2M-1)\Delta/2$ to each access intensity level starting from the level corresponding to $[0, \Delta]$. Then, we approximate X_i to the quantized value by a mapping function L. In other words, if we denote by Q_i the quantized value of X_i, $Q_i = L(X_i)$. We also number the access intensity levels from zero to $M-1$ and denote by L_i the access intensity level that Q_i belongs to.

In the transition matrix manipulation module, the access intensity for the next time slot is estimated. An eNB maintains M by $M + 1$ state transition matrix $A[M][M + 1]$. The transition matrix is used to approximate the correlation structure between X_{i-1} and X_i. To achieve the goal, an eNB uses the transition matrix to keep track of the relation between the access intensity measured during the previous time slot (Q_{i-1}) and the access intensity measured during the current time slot (Q_i). In the state transition matrix, a state corresponds to the index number of an access intensity level. At the end of each time slot i, an eNB quantizes X_{i-1} and X_i into $Q_{i-1} = L(X_{i-1})$ and $Q_i = L(X_i)$. Since L_i is the access intensity level that Q_i belongs to, the eNB updates the state transition matrix as follows:

$$A[L_{i-1}][L_i] = A[L_{i-1}][L_i] + 1,$$
$$A[L_i][M + 1] = A[L_i][M + 1] + 1. \tag{1}$$

Then, using the transition matrix, the transition matrix manipulation module predicts the access intensity level of the next time slot $i + 1$ as

$$E[L_{i+1} \mid L_i] = \sum_{j=1}^{M} jPr(L_j \mid L_i), \tag{2}$$

where $Pr(L_{i+1} = j \mid L_i) = A[L_i][L_j]/A[L_i][M + 1]$.

Using $E[L_{i+1} \mid L_i]$, the barring factor generation module determines the barring factor p as follows:

$$p = 1 - \frac{E[L_{i+1} \mid L_i]}{M}. \tag{3}$$

4. Performance Analysis

We use the traffic model for smart electric metering MTC application as an experimental scenario [1]. Smart meters support various applications such as automatic meter reading, energy demand management, and microelectric generation management. In the traffic model, the household density in urban area of London is 35,670 households/cell. Periodical reporting of meter readings ranges from 5 mins to 24 hours [14]. The reporting cycle affects the access intensity of meters in a cell. We set reading frequency to 5 mins. In the experiments, we set A_M to the maximum number of devices which attempt the random access in a time slot. For example, we set A_M to 30 for the 100 ms time slot. The parameters for the random access channel of LTE-A networks follow [1].

An eNB serves a number of N devices in a cell. Each device generates a single access request during a period of time T. In other words, all devices activate sending data between $t = 0$ and $t = T$. In our experimental scenario, T is set to 5 mins. The access intensity follows the distribution $p(t)$. The number of devices which generate access requests in the ith access opportunity, that is, access intensity, is defined as follows:

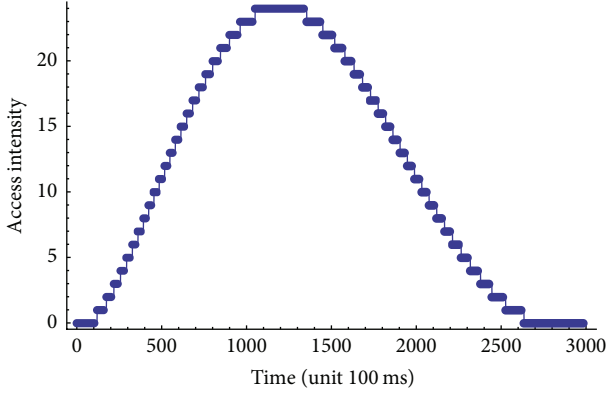$$N_{new}^i = N \int_{t_i}^{t_{i+1}} p(t)\, dt, \tag{4}$$

FIGURE 3: The access intensity of smart electric metering application.

where t_i is the time of the ith access opportunity and $p(t)$ follows the Beta distribution with the values of $\alpha = 3$ and $\beta = 4$ [1]:

$$p(t) = \frac{t^\alpha (T - t)^{\beta - 1}}{T^{\alpha + \beta - 1} \text{Beta}(\alpha, \beta)}, \qquad (5)$$

where $\text{Beta}(\alpha, \beta)$ is the Beta function. Figure 3 shows the access intensity for smart electric metering application.

We adopt the congestion coefficient to enable or disable a congestion control mechanism like ACB [15]. The congestion coefficient is the inverse of the ratio of the number of successful devices to the number of contending devices. When the congestion coefficient exceeds a specific threshold, the congestion control is activated. We set the threshold to 0.4. It is deactivated when the congestion coefficient goes below the threshold. Once congestion control is activated, eNB manipulates the barring factor to regulate the access intensity.

Figure 4 shows the variance of barring factor with different number of access intensity levels (M). We can see that the barring factor is adaptively controlled according to the access intensity. We analyze the variance of barring factor in in three aspects. The first is when the time slot is changed, 500 ms and 1000 ms. The curve when the time slot is 1000 ms has less spikes than it when the time slot is 500 ms. The second is when the number of access intensity levels is changed, 5, 10, 15, and 20. As the number of levels increases, the curves look like staircase. The last is when the access intensity within the traffic model is changed, increase period and decrease period. In the asymmetric curve of the Beta distribution in Figure 3, uphill slope is steeper than downhill slope. Thus, the curve of the barring factor in gently decreased period of access intensity is stable more than that in steeply increased period.

Figure 5 shows the variance of barring factor when the access intensity increases. In the figure, N indicates 35,670 households/cell, that is, the household density in urban area of London. We set the number of access intensity levels to 10. In steeply increased period of access intensity, our method rapidly reacts. In gently decreased period, the method slowly reacts. However, if the average of access intensity is too

(a) Time unit = 500 ms
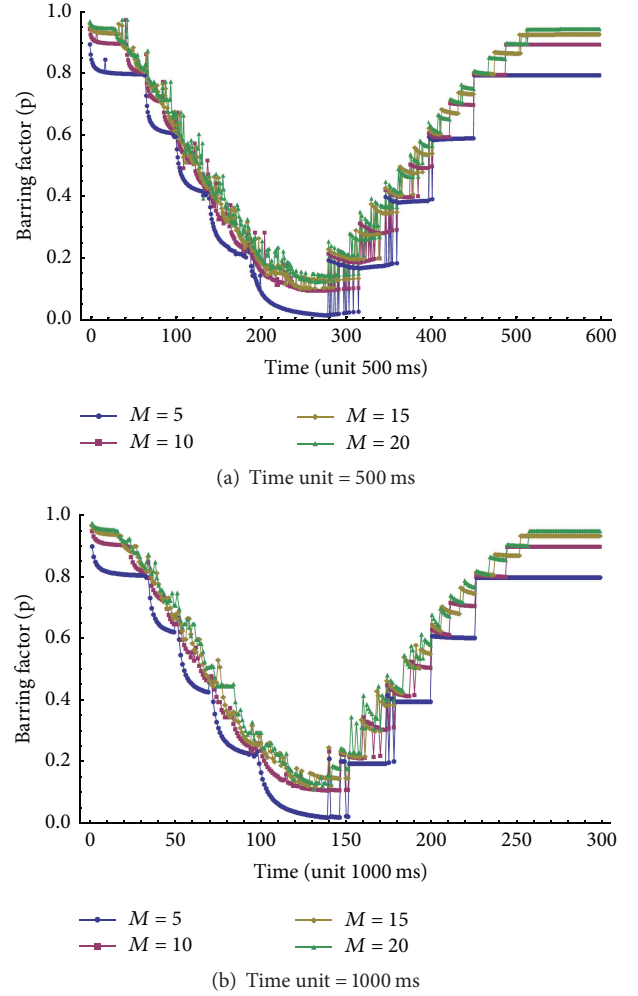

(b) Time unit = 1000 ms

FIGURE 4: The value of barring factor by varying the number of access intensity levels.

low (in the figure, when the total number of devices = $0.5N$), the colliding devices decrease and the barring factor correspondingly reacts.

Figure 6 shows the access success probability which is the successful completion probability of the random access in the maximum number of preamble transmissions. In the experiments, the time slots are changed from 60 ms to 200 ms. For comparison, we use the original ACB [4]. In the original ACB, once the congestion control is activated, the barring factor is set to 0.1. When our method is used, the average of success probability is about 18% better than that of the original ACB. Figure 7 shows collision probability which is the ratio between the number of occurrences when two or more devices with the same preamble try to access RACH and the total number of access opportunities. When our method is used, the average of collision probability is about 12% better than that of the original ACB.

Figures 6 and 7 show that there is an optimal range on performance when M is between 3 and 4. As shown in Figure 8, when M increases, the difference of access intensities between adjacent levels decreases and the effect of
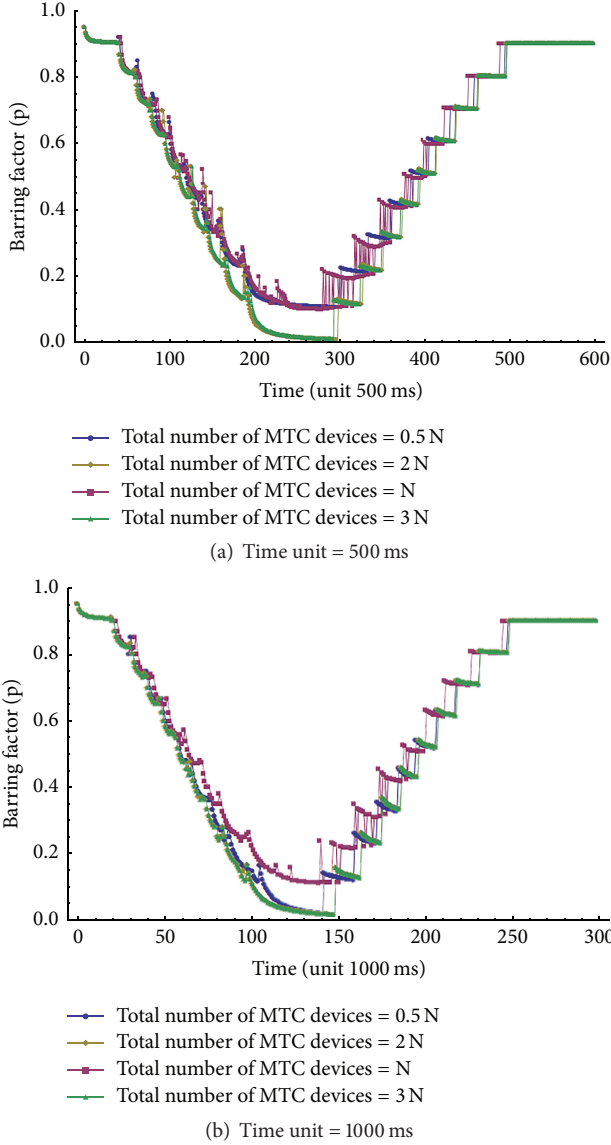
(a) Time unit = 500 ms

(b) Time unit = 1000 ms

FIGURE 5: The value of barring factor by varying the number of devices.

FIGURE 6: The success probability by varying time slots.

FIGURE 7: The collision probability by varying time slots.

changing the barring factor decreases. When M decreases, the difference of access intensities between adjacent levels increases and the changing the access intensity does not finely affect the barring parameter adjustment. It indicates that the performance can be improved by controlling the number of access intensity levels.

Figure 9 shows the variance of the number of successful devices for each time slot. The successful device indicates the device which successfully accesses RACH in a time slot. When the access intensity is low, the performances of our method and the original ACB are similar. However, when the access intensity is high, the number of successful devices in our method is 3.5 times larger than that in the original ACB. In other words, our method efficiently controls the barring factor to improve the system performance.
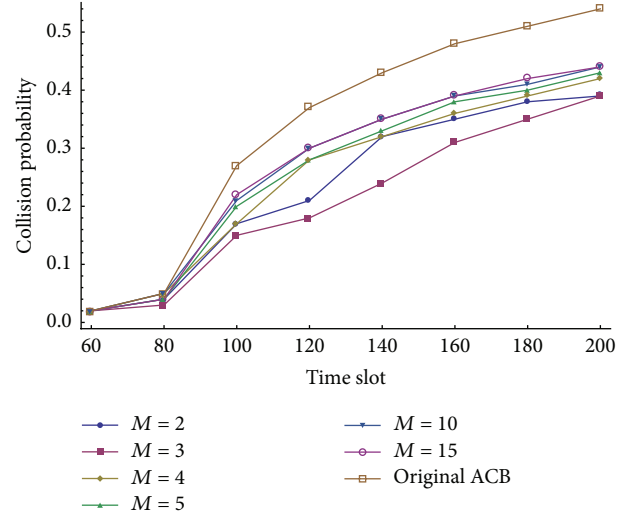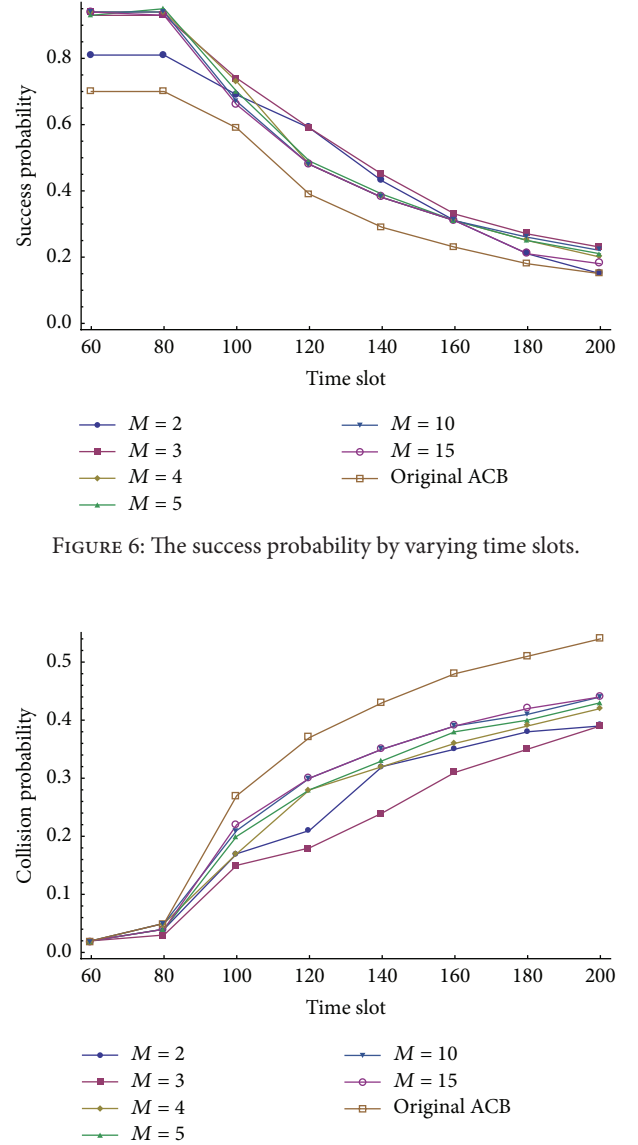
5. Conclusion

In this paper, we proposed a dynamic access barring method that controls the barring factor according to the access intensity. By measuring the access intensity at the end of each time interval, we construct the transition probability among the access intensity. Then, at the end of each interval, eNB calculates the expected access intensity based on the current access intensity. Simulation results show that the proposed method outperforms the original ACB method in terms of the success probability and the collision probability because it can control the barring factor even if the traffic intensity changes over time. Simulation results also suggest that there is an optimal value for the number of access intensity levels which is the operating parameter of the proposed method. Since the number of access intensity levels influences the success probability, we leave it as a future work to devise
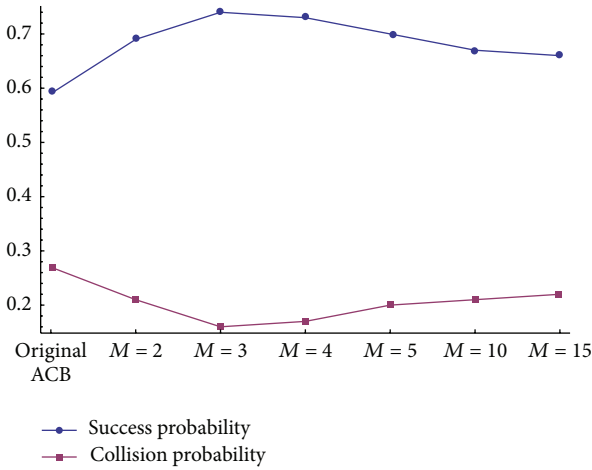
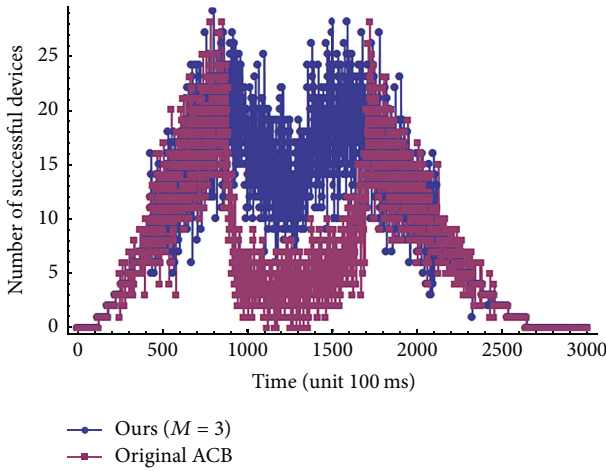FIGURE 8: The performance by varying the number of access intensity levels.

FIGURE 9: The performance comparison between our method and the original ACB.

a systematic method that finds the optimal value for the operating parameter by mathematical analysis.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

References

[1] 3rd Generation Partnership Project, Technical specification group radio access network; study on RAN improvements for machine-type communications, 3GPP TR 37.868 V11.0.0, 2011.

[2] 3rd Generation Partnership Project, "Evolved universal terrestrial radio access (E-UTRA); radio resource control (RRC); protocol specification," 3GPP TS 36.331 V12.7.0, 2015.

[3] P. Jain, P. Hedman, and H. Zisimopoulos, "Machine type communications in 3GPP systems," *IEEE Communications Magazine*, vol. 50, no. 11, pp. 28–35, 2012.

[4] M. Hasan, E. Hossain, and D. Niyato, "Random access for machine-to-machine communication in LTE-advanced networks: issues and approaches," *IEEE Communications Magazine*, vol. 51, no. 6, pp. 86–93, 2013.

[5] M.-Y. Cheng, G.-Y. Lin, H.-Y. Wei, and A. C.-C. Hsu, "Overload control for machine-type-communications in LTE-advanced system," *IEEE Communications Magazine*, vol. 50, no. 6, pp. 38–45, 2012.

[6] R.-G. Cheng, J. Chen, D.-W. Chen, and C.-H. Wei, "Modeling and analysis of an extended access barring algorithm for machine-type communications in LTE-A networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 2956–2968, 2015.

[7] A. Ksentini, Y. Hadjadj-Aoul, and T. Taleb, "Cellular-based machine-to-machine: overload control," *IEEE Network*, vol. 26, no. 6, pp. 54–60, 2012.

[8] Z. Wang and V. W. S. Wong, "Optimal access class barring for stationary machine type communication devices with timing advance information," *IEEE Transactions on Wireless Communications*, vol. 14, no. 10, pp. 5374–5387, 2015.

[9] S. Duan, V. Shah-Mansouri, and V. W. S. Wong, "Dynamic access class barring for M2M communications in LTE networks," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM '13)*, pp. 4747–4752, Atlanta, Ga, USA, December 2013.

[10] S. Duan, V. Shah-Mansouri, Z. Wang, and V. Wong, "D-ACB: adaptive congestion control algorithm for bursty M2M traffic in LTE networks," *IEEE Transactions on Vehicular Technology*, no. 99, p. 1, 2016.

[11] G. Wang, X. Zhong, S. Mei, and J. Wang, "An adaptive medium access control mechanism for cellular based machine to machine (M2M) communication," in *Proceedings of the IEEE International Conference on Wireless Information Technology and Systems (ICWITS '10)*, pp. 1–4, Honolulu, Hawaii, USA, September 2010.

[12] S.-Y. Lien, T.-H. Liau, C.-Y. Kao, and K.-C. Chen, "Cooperative access class barring for machine-to-machine communications," *IEEE Transactions on Wireless Communications*, vol. 11, no. 1, pp. 27–32, 2012.

[13] H. He, Q. Du, H. Song, W. Li, Y. Wang, and P. Ren, "Traffic-aware ACB scheme for massive access in machine-to-machine networks," in *Proceedings of the IEEE International Conference on Communications (ICC '15)*, pp. 617–622, London, UK, June 2015.

[14] Verizon, Smart grid traffic behaviour discussion, 3GPP R2-102340, RAN WG2 Meeting #69b, 2010.

[15] Intel Corporation, "Further performance evaluation of EAB information update mechanisms," 3GPP R2-120270, RAN WG2 Meeting #77, 2012.

Research Article

Adaptive Power Allocation and Splitting with Imperfect Channel Estimation in Energy Harvesting Based Self-Organizing Networks

Kisong Lee[1] and JeongGil Ko[2]

[1]Department of Information and Telecommunication Engineering, Kunsan National University, Gunsan 573-701, Republic of Korea
[2]Department of Software and Computer Engineering, Ajou University, Suwon 16499, Republic of Korea

Correspondence should be addressed to JeongGil Ko; jgko@ajou.ac.kr

As miniature-sized embedded computing platforms are ubiquitously deployed to our everyday environments, the issue of managing their power usage becomes important, especially when they are used in energy harvesting based self-organizing networks. One way to provide these devices with continuous power is to utilize RF-based energy transfer. Previous research in RF-based information and energy transfer builds up on the assumption that perfect channel estimation is easily achievable. However, as our preliminary experiments and many previous literature in wireless network systems show, making perfect estimations of the wireless channel is extremely challenging due to their quality fluctuations. To better reflect reality, in this work, we introduce an adaptive power allocation and splitting (APAS) scheme which takes imperfect channel estimations into consideration. Our evaluation results show that the proposed APAS scheme achieves near-optimal performances for transferring energy and data over a single RF transmission.

1. Introduction

As we slowly enter the era of the Internet of Things (IoT), we will start to experience various embedded computing systems being introduced to our everyday lives. In particular, it is important to maintain a long sensing and operational lifetime in self-organizing networks (SONs). Given that SONs are typically meant to tackle applications with little or no human intervention, their operational durations can determine the overall system's self-conguration, self-optimization, and self-healing performance [1, 2]. For this, a decade of research in the wireless embedded systems domain has introduced a number of schemes for optimizing energy efficiency on resource limited computing platforms [3–5].

In addition to these schemes that focus on conserving the power usage, another direction of research is to gather energy. This energy gathering can take two different forms where, in the first, an explicit hardware module is attached for harvesting energy from external sources (e.g., sunlight, wind, and vibration) [6, 7] and, in the second, the power generated from radio frequency (RF) signals can be used to transfer energy [8–10]. Given that the latter is only minimally affected by external environmental factors, we believe that it is an interesting research direction to explore. Since they do not require a large-sized energy harvesting unit, applying energy transfer techniques to data communications can effectively reduce the size of the the hardware used in low-power wireless networking. Based on these benefits, in this work we study the possibilities of information and energy transfer using RF signals for powering low-power embedded computing platforms.

Given its attractiveness, a number of previous works have tried tackling interesting issues in various aspects of this research field. For example, in [11, 12], the authors investigated the theoretical performance limits for simultaneous wireless information and power transfer (SWIPT). The works in [13, 14] proposed time switching and dynamic power splitting for enabling efficient SWIPT. Furthermore, Nasir et al. took these findings to a networking perspective and introduced the relaying protocol in [15]. Here, the authors proposed a network where an energy constrained relay node harvests energy from the RF signals of a source node and uses this

harvested energy for relaying information to the next hop. In [16], Shen et al. proposed transmitter designs for sum-rate maximization with energy harvesting constraints on a multiple-input single-output (MISO) interference channel. In addition, an energy efficient resource allocation algorithm for SWIPT was investigated [17] and multiuser scheduling schemes for improving user fairness were studied in wireless networks with energy harvesting [18].

Despite these efforts from the research community, in this work we identify one important assumption that most of these previous works made. Namely, these works commonly took the assumption that wireless channels would be continuously stable and the communication quality would be perfect in all cases. However, research from the wireless networking systems community showed that this observation is far from being true. These RF signals can be severely impacted by external factors such as human movement, environmental changes, and even the time of day. Therefore, we believe that taking such real-world channel factors into consideration as we model the information and energy transfer behaviors of RF signals is important. Specifically, in this work we try to understand the performance of wireless links in reality and present a novel information and energy transfer model that reflects the nonperfect nature and inevitably imperfect channel quality estimations of real-world wireless environments. In addition, we propose an adaptive power allocation and splitting (APAS) scheme with considerations for imperfect wireless channel estimations in energy harvesting based SONs. Our evaluations show that APAS outperforms preexisting schemes and performs close to the optimal.

We summarize the contributions of this work threefold.

(i) We present empirical results on the RF characteristics in indoor environments to showcase the dynamics of real-world wireless channels. Our findings lead to a conclusion that perfect channel estimations are difficult to achieve due to various unexpected external factors.

(ii) We introduce a novel signal transfer model with channel estimation errors in consideration for analyzing the simultaneous transfer of information and energy using RF signals. Our model reflects realistic channel environments and, therefore, provides accurate performance bounds.

(iii) Using convex optimization techniques, we propose a resource allocation strategy that finds suboptimal transmission power allocation and power splitting ratios iteratively for a given training interval. Through simulations, we show that the proposed scheme provides near-optimal performance as well as considerable performance improvement compared to conventional schemes.

The rest of this paper is structured as follows. We start off the paper with an empirical study of real-world channel environments in Section 2. The findings from our preliminary studies become the basis of our system model presented in Section 3 and the proposed adaptive power allocation and splitting (APAS) scheme in Section 4. Using Section 5,
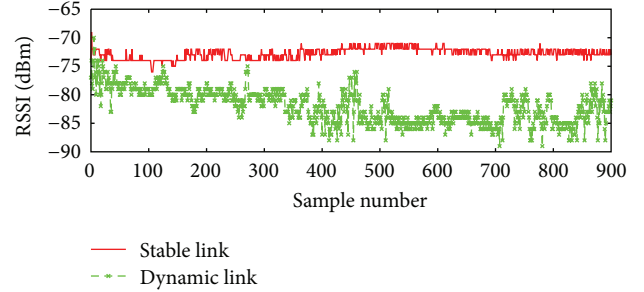

Figure 1: Empirical RSSI traces for stable and dynamic indoor environments.

we evaluate the performance of our proposed scheme and summarize our work in Section 6.

2. Real Channel Environments

Before we explain the details of our proposed scheme, we start by presenting empirical study results that show how dynamic a practical wireless channel environment can be. Specifically, we configure a transmitter node and a receiver node to be positioned ~5 meters apart. The transmitter node sends periodic packets with a packet transmission interval of 250 msec at 0 dBm and we test the wireless link in two different scenarios: one with no surrounding human movement activities (e.g., night-time) and another with continuous movement (e.g., day-time) between the two nodes. Since nodes were installed in a hallway environment with consistent human activities, the dynamic links in Figure 1 are sure to have high channel quality variability. In this experimental setting, we present the received signal strength indicator (RSSI) value observed at the receiver for incoming packets. Specifically, in Figure 1 we plot the RSSI over time for both the stable and dynamic links. Notice from Figure 1 that, with natural human-generated link dynamics (e.g., typical human movement behaviors), the signal strengths of incoming packets severely fluctuate.

These empirical results, though tested for a single environment, provide experimental evidence on the dynamics of real wireless channels. Using Figure 1, we try to show that schemes designed for real-world should not assume a stable wireless channel. This leads to an observation that perfect channel estimation can be difficult to achieve; thus, channel estimation analysis for information and energy transfer cannot be perfect in most cases. The remainder of this work builds up on such empirical findings to propose a model for analyzing the RF-based information and energy transfer under imperfect channel estimations.

3. System Model

Based on our observations of real-world channel environments, we take into consideration imperfect channel estimations in an orthogonal frequency division multiplexing (OFDM) based wireless point-to-point link consisting of a single transmitter (Tx) and receiver (Rx) pair, as shown in

Figure 2. Tx and Rx nodes each are equipped with a single-antenna, and the frequency band is divided into independent N subchannels. We assume that the subchannels follow a discrete time block-fading model in which the channel state is invariant for a transmission block interval T [19, 20]. In addition, each real subchannel, h_n, is assumed to be an independent identically distributed complex random variable, such as $h_n \sim \mathscr{CN}(0, \sigma_h^2)$ for $1 \le n \le N$. In a practical system, the channel estimation process obtains and exploits the channel state information (CSI). The transmission block T is divided into two periods, where, in the first, we define T_τ for channel training and, in the second, we define the duration $T - T_\tau$ for data transmission. The minimum mean square error (MMSE) criterion is assumed to be used for estimating the channel status, and each estimated subchannel, $\widehat{h}_n$, follows a Gaussian distribution, such as $\widehat{h}_n \sim \mathscr{CN}(0, \sigma_{\widehat{h}}^2)$ for $1 \le n \le N$.

Under such settings, we note that the RF signals (at the Rx node) can be used in two ways, either for information decoding (ID) or for energy harvesting (EH), while these modes cannot take place simultaneously. Here, the received signal is split in two portions in each subchannel; the portion of ρ_n among $T - T_\tau$ is reserved for ID, while $1 - \rho_n$ is for EH before performing active analog or digital signal processing. (We assume that the Rx node is equipped with a perfect passive power splitting unit [17].) In addition, as RF signals are split in two streams, two types of noise should be considered: antenna noise, n_A, and signal processing noise, n_S. n_A is generated at the Rx antenna while n_S is introduced when the received signal is divided into ID and EH. Nevertheless, we neglect n_A since it is much smaller than n_S [14]. Furthermore, we assume $n_S \sim \mathscr{CN}(0, 1)$. Then, the achievable sum rate using the estimated subchannels can be represented by [20]

$$
\begin{aligned}
R &= \sum_{n=1}^{N} r_n \\
&= \sum_{n=1}^{N} \frac{T - T_\tau}{T} \log_2 \left(1 + \frac{(1 + p_\tau T_\tau) \rho_n p_n |\widehat{h}_n|^2}{1 + \rho_n p_n + p_\tau T_\tau} \right),
\end{aligned}
\tag{1}
$$

where p_τ and p_n denote the power for estimating channel conditions and power allocated in subchannel n for data transmission, respectively. When performing EH at the Rx, the harvested energy is represented as

$$
\sum_{n=1}^{N} g_n = \sum_{n=1}^{N} \frac{T - T_\tau}{T} \eta |\widehat{h}_n|^2 (1 - \rho_n) p_n,
\tag{2}
$$

where η is the energy conversion efficiency achieved by converting the received RF signals into harvestable energy at the Rx. We also assume that $\sum_{n=1}^{N} \eta |h_n|^2 \le 1$, according to the laws of thermodynamics.

We now look into the training interval, the power allocation, and the power splitting ratio required to maximize the sum rate while guaranteeing the minimum harvested energy $g_{\min}$. Consider

$$
\begin{aligned}
\max_{T_\tau, \vec{p}, \vec{\rho}} \quad & \sum_{n=1}^{N} r_n \\
\text{s.t.} \quad & \text{C1: } \sum_{n=1}^{N} g_n \ge g_{\min}, \\
& \text{C2: } \sum_{n=1}^{N} p_n \le p_{\max}, \\
& \text{C3: } p_n \ge 0, \quad \forall n, \\
& \text{C4: } 0 \le \rho_n \le 1, \quad \forall n, \\
& \text{C5: } 0 \le T_\tau \le T.
\end{aligned}
\tag{3}
$$

Here, $\vec{p} = \{p_1, \ldots, p_N\}$ and $\vec{\rho} = \{\rho_1, \ldots, \rho_N\}$. Furthermore, the five constraints can be explained as follows. Constraint C1 ensures that the amount of energy harvested should be larger than the minimum amount of required energy $g_{\min}$. Constraint C2 limits the available transmission power of the Tx to $p_{\max}$. Constraint C3 is a nonnegative constraint on the transmission power. Constraints C4 and C5 are the ranges of ρ_n and T_τ, respectively.

4. Adaptive Power Allocation and Splitting

We now propose an adaptive power allocation and splitting (APAS) algorithm by solving the optimization problem in (3). We note that it is difficult to find a closed form solution for T_τ^* from optimization techniques, given that the objective function of (3) is not in concave form with respect to T_τ. However, considering the fact that T_τ lies within the interval $(0, T)$, T_τ^* can be derived using a one-dimensional exhaustive search. For example, T_τ^* can be found from the probability density function (PDF) of channel distribution. If T_τ^* is determined at once, it can be used for estimating all unknown channels that will be generated. Furthermore, using the concavity of $\vec{p}$ and $\vec{\rho}$, we can find their suboptimal values iteratively, which reflect channel estimation errors. (In a biconvex problem, where the problem in (3) is convex with respect to $\vec{p}$ for a fixed $\vec{\rho}$ or vice versa, the convergence of a partial optimum solution can be guaranteed by using the block coordinate descent (BCD) algorithm [21].)

For a given T_τ, Tx estimates CSI on subchannels and determines the allocated power and the power splitting ratio. The Lagrangian function of (3) can be expressed by

$$
\begin{aligned}
\Lambda(\vec{p}, \vec{\rho}, \lambda, \mu) = &\sum_{n=1}^{N} r_n + \lambda \left(\sum_{n=1}^{N} g_n - g_{\min} \right) \\
&+ \mu \left(p_{\max} - \sum_{n=1}^{N} p_n \right).
\end{aligned}
\tag{4}
$$

Here, λ and μ are nonnegative Lagrangian multipliers, which correspond to constraints C1 and C2, respectively. To find a solution to this, we decouple the original problem into parallel N subproblems for each subchannel. By discarding the
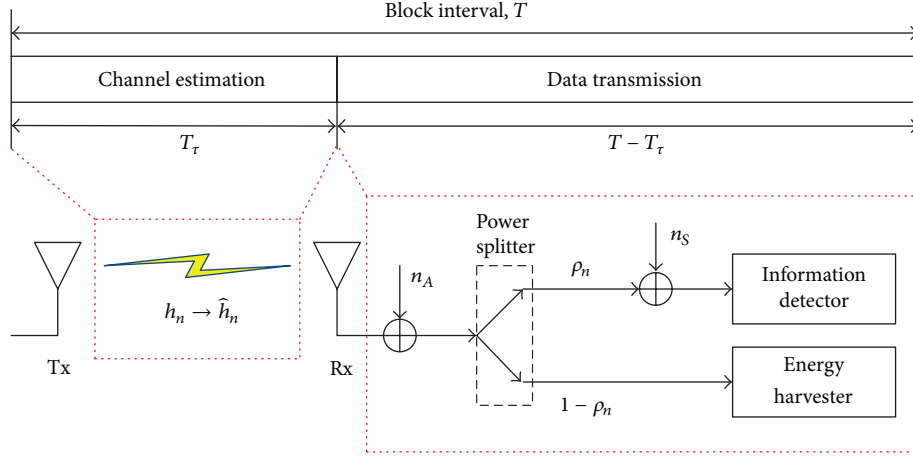
FIGURE 2: System model for wireless information and energy transfer.

constant terms, the Lagrangian function (4) for a particular subchannel n can be represented as

$$\Lambda\left(p_n, \rho_n, \lambda, \mu\right) = r_n + \lambda g_n - \mu p_n. \quad (5)$$

By taking the derivative of (5) with respect to p_n, we can obtain a power allocation strategy as in the following:

$$p_n^* = \left[\frac{H_e}{1 + H_e \left|\hat{h}_n\right|^2} \right.$$

$$\cdot \sqrt{\frac{4\rho_n \left|\hat{h}_n\right|^2 \left(1 + H_e \left|\hat{h}_n\right|^2\right) + (\ln 2) H_e^2 \left|\hat{h}_n\right|^4 \zeta_n}{4 (\ln 2) \rho_n^2 \zeta_n}}$$

$$\left. - \frac{H_e \left(2 + H_e \left|\hat{h}_n\right|^2\right)}{2\rho_n \left(1 + H_e \left|\hat{h}_n\right|^2\right)} \right]^+. \quad (6)$$

Here, $[x]^+ = \max(0, x)$, $H_e = 1 + p_\tau T_\tau$, and $\zeta_n = (T/(T - T_\tau))\mu - \lambda\eta|\hat{h}_n|^2(1 - \rho_n)$. For the obtained p_n^* from (6), we also take the derivative of (5) with respect to ρ_n, to obtain the power splitting strategy as in the following:

$$\rho_n^* = \left[\frac{H_e}{1 + H_e \left|\hat{h}_n\right|^2} \right.$$

$$\cdot \sqrt{\frac{4 \left|\hat{h}_n\right|^2 \left(1 + H_e \left|\hat{h}_n\right|^2\right) + (\ln 2) H_e^2 \lambda\eta \left|\hat{h}_n\right|^6}{4 (\ln 2) p_n^2 \lambda\eta \left|\hat{h}_n\right|^2}}$$

$$\left. - \frac{H_e \left(2 + H_e \left|\hat{h}_n\right|^2\right)}{2p_n \left(1 + H_e \left|\hat{h}_n\right|^2\right)} \right]_0^1. \quad (7)$$

Here, $[x]_0^1 = \min(\max(0, x), 1)$.

Then, p_n^* and ρ_n^* can be interpreted in terms of channel estimation, channel condition, and harvested energy as follows. We especially note that an increment in T_τ ensures the exact estimation of channel conditions but decreases the data transmission time. An extremely large T_τ does not allow for a dedicated data transmission time; as a result, p_n^* and ρ_n^* become zero. In addition, p_n^* and ρ_n^* are proportional to $|\hat{h}_n|^2$, so they show large values on subchannels with good channel conditions. p_n^* is proportional to $\lambda\eta|\hat{h}_n|^2$, which is related to the amount of harvested energy, but ρ_n^* is inversely proportional to $\lambda\eta|\hat{h}_n|^2$. An amount of p_n^* increases on a good subchannel, where large amounts of energy harvesting are possible, while ρ_n^* decreases on that subchannel to meet the constraint C1 for harvested energy tightly. In short, p_n^* and ρ_n^* are adjusted reciprocally to the maximize sum rate while guaranteeing $g_{\min}$.

Based on the obtained $\overrightarrow{p^*}$ and $\overrightarrow{\rho^*}$, Lagrangian multipliers can be updated by a well-known bisection algorithm or a gradient algorithm. We detail the overall procedure of the proposed algorithm in Algorithm 1.

5. Simulation Results and Discussion

For evaluations, we configure a simulation environment as in the following. Specifically, we set $N = 32$, $T = 1000$, $\eta = 0.9$, $p_\tau = p_{\max} = 43$ dBm, and $g_{\min} = 20$ dBm. We assume that subchannel h_n experiences Rayleigh fading, so h_n is generated as a random variable distributed exponentially with $\sigma_h^2 = 10^{-4}$. In addition, it is independent from other subchannels h_j for $j \neq n$. We compare the performance of APAS with three algorithms: OPAS, EPAS, and E²PAS.

(i) Optimal power allocation and splitting (OPAS): with perfect CSI at the Tx (CSIT), $\vec{p}$ and $\vec{\rho}$ are determined optimally.

(ii) Equal Power Allocation and Adaptive Power Splitting (EPAS): power is allocated equally to all subchannels, and $\vec{\rho}$ is determined adaptively to meet $g_{\min}$ with respect to (7).
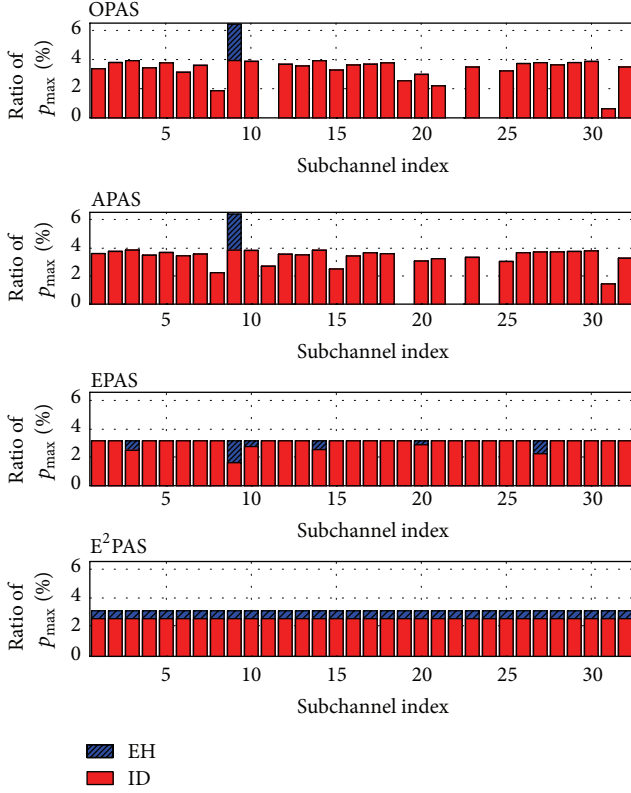
FIGURE 3: Power allocation and splitting of OPAS, APAS, EPAS, and E^2PAS.

```
(1)   Initialize $\vec{p}$, $\vec{\rho}$, and Lagrangian multipliers
(2)   for $T_\tau = 1 : T$
(3)       Estimate $|\hat{h}_n|^2$ for $\forall n$ based on the PDF of channel
(4)       Evaluate $R$
(5)   end for
(6)   Return $T_\tau^* = \max_{T_\tau} R$
(7)   repeat
(8)       Find $\overrightarrow{p^*}$ according to (6)
(9)       Find $\overrightarrow{\rho^*}$ according to (7)
(10)      Update Lagrangian multipliers, $\lambda$ and $\mu$
(11)  until $\overrightarrow{p^*}$ and $\overrightarrow{\rho^*}$ converge
```

ALGORITHM 1: Adaptive power allocation and splitting.

(iii) Equal Power Allocation and Equal Power Splitting (E^2PAS): power is allocated equally to all subchannels, and $\vec{\rho}$ is determined equally for all subchannels to meet g_{min}.

Figure 3 shows the power allocation and splitting of OPAS, APAS, EPAS, and E^2PAS, respectively. In OPAS and APAS, a large amount of power is allocated to the subchannel with the highest channel gain, and a portion of power is split to harvest energy on that subchannel. It is best to use the power allocated to the best subchannel for guaranteeing g_{min} since energy can be harvested with higher efficiency.
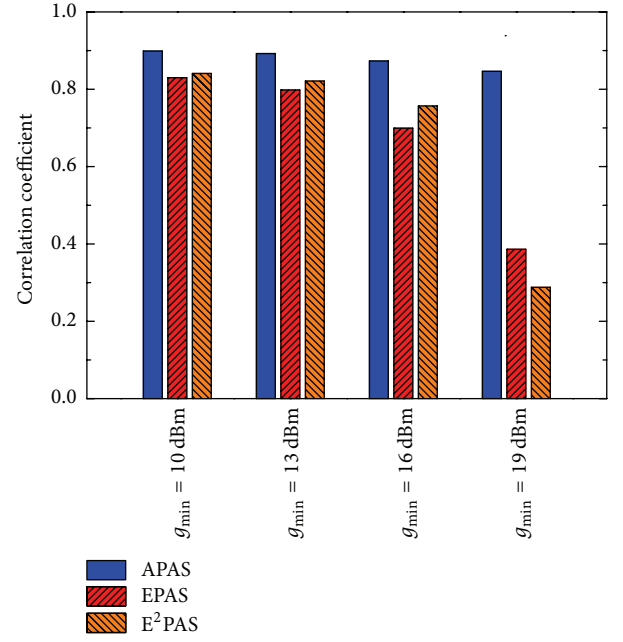

FIGURE 4: Correlation coefficient of APAS, EPAS, and E^2PAS.

The differences of resource allocation between OPAS and APAS come from the errors in channel estimation, but their resource allocation forms show similar tendency. This indicates that APAS can achieve a performance close to the optimal bound. On the other hand, in EPAS and E^2PAS, power is allocated equally on all subchannels and energy is harvested on several subchannels; therefore, a subset of the power can be used inefficiently. In particular, despite its implementation simplicity, performance can be degraded severely in E^2PAS due to the fact that resource allocation is performed regardless of the channel conditions.

Figure 4 shows the correlation coefficient of APAS, EPAS, and E^2PAS with varying g_{min}. This result targets for showing how the power allocation and splitting of each scheme is similar to OPAS. In OPAS, more power is allocated to the best subchannel to ensure g_{min} with a high efficiency as g_{min} increases. As a result, we can notice here that the correlation coefficients of EPAS and E^2PAS decrease seriously. On the other hand, the correlation coefficient of APAS stays high, ~0.9, even with increasing g_{min}. Therefore, this result suggests that APAS can adapt the power allocation and splitting strategy to a level similar to the optimal solution under various conditions.

Figure 5 plots the data rate R versus the training interval T_τ, which shows the effects of T_τ on R. As T_τ increases, it is possible to estimate the channel conditions more accurately. As a result, R increases gradually to a peak point. However, additional increase in T_τ beyond its optimal value causes reduction in the dedicated transmission time, thereby decreasing R. This suggests that there is an optimal value of T_τ for maximizing R from the tradeoff relationship between the accuracy of channel estimations and the duration of data transmission time. The optimal training interval T_τ^* increases with a large g_{min}, which indicates that an exact channel
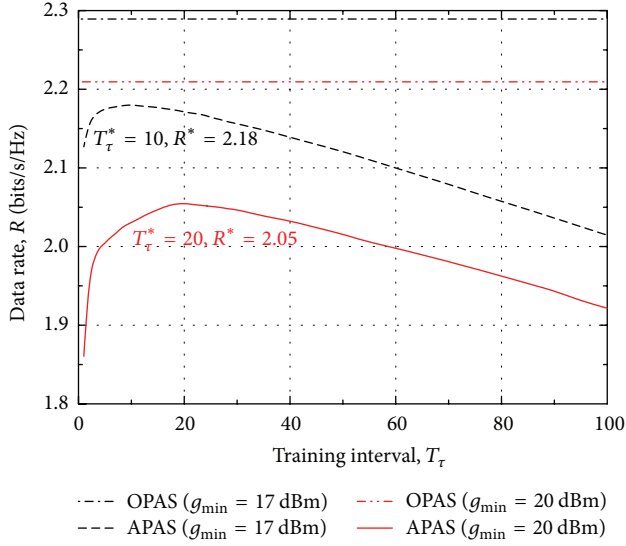
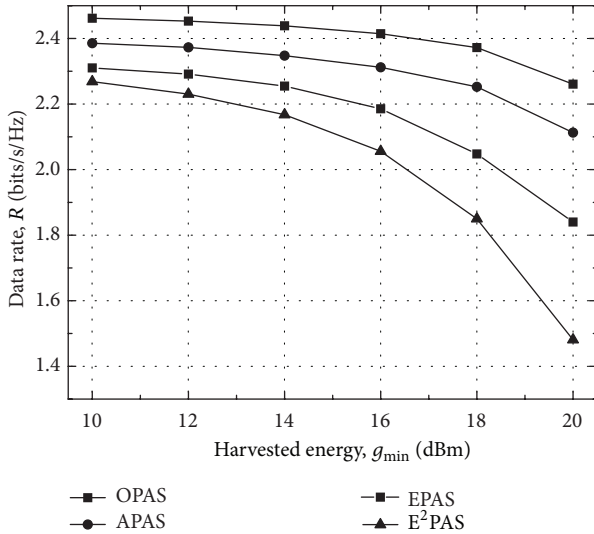FIGURE 5: Data rate R versus training interval T_τ.

FIGURE 6: Data rate R versus harvested energy g_{min} when $T_\tau = 20$.

Therefore, APAS outperforms EPAS and E^2PAS significantly at larger g_{min} values. For example, APAS outperforms EPAS and E^2PAS in terms of R by 20% and 40% at $g_{min} = 20$ dBm, respectively. On the other hand, the performance difference between OPAS and APAS remains relatively constant despite increasing g_{min}.

6. Conclusion

RF-based information and energy transferring techniques hold the potential to dramatically change the design of wireless systems and their networking architectures. Nevertheless, the research community is still in the early stages of validating their effectiveness. In this work, we target to tackle one of the strongest assumptions that most of the works in information and energy transfer made, which is the assumption that perfect channel estimation is possible. We show through an empirical study that the variability of wireless channels makes perfect estimations of the wireless environment close to impossible. For this, we propose APAS, which takes into consideration imperfect channel estimation results for evaluating the effectiveness of information and energy transfer on wireless devices for energy harvesting based SONs. In APAS, the power allocation and splitting ratio is determined adaptively with considerations for the estimated channel quality. In addition, our results indicate that APAS achieves near-optimal performances under various conditions. We hope that this work can act as a catalyst in enabling future research that tries to adopt RF-based information and energy transfer in realistic channel environments.

Competing Interests

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

Acknowledgments

References

[1] H. Hu, J. Zhang, X. Zheng, Y. Yang, and P. Wu, "Self-configuration and self-optimization for LTE networks," *IEEE Communications Magazine*, vol. 48, no. 2, pp. 94–100, 2010.

[2] 3GPP, "Self-organizing networks (SON): concepts and requirements," 3GPP TS 32.500, 2008, V8.0.0.

[3] G. Y. Li, Z. Xu, C. Xiong et al., "Energy-efficient wireless communications: tutorial, survey, and open issues," *IEEE Wireless Communications*, vol. 18, no. 6, pp. 28–35, 2011.

[4] C. Han, T. Harrold, S. Armour et al., "Green radio: radio techniques to enable energy-efficient wireless networks," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 46–54, 2011.

[5] H. Bogucka and A. Conti, "Degrees of freedom for energy savings in practical adaptive wireless systems," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 38–45, 2011.

[6] H. Kulah and K. Najafi, "Energy scavenging from low-frequency vibrations by using frequency up-conversion for wireless sensor

estimation is desired to satisfy a large requirement for the harvested energy. There is no significant difference, in particular less than 10%, between the maximum R achieved at T_τ^* and the optimal bound of R.

Figure 6 shows the data rate R versus the harvested energy g_{min} when $T_\tau = 20$. As g_{min} increases, a large portion of power should be used for harvesting energy. In consequence, R decreases gradually for all algorithms. APAS has the gain of adaptive power allocation compared with EPAS, while it has the gain of adaptive power allocation and splitting when compared to E^2PAS. Therefore, we can confirm the gain of each adaptive strategy by comparing APAS with EPAS and E^2PAS, respectively. In EPAS and E^2PAS, the rates at which R decreases with increasing g_{min} are noticeably steeper than those of OPAS and APAS. This is mainly due to the fact that, in EPAS and E^2PAS, the power cannot be used efficiently.

applications," *IEEE Sensors Journal*, vol. 8, no. 3, pp. 261–268, 2008.

[7] S. Sudevalayam and P. Kulkarni, "Energy harvesting sensor nodes: survey and implications," *IEEE Communications Surveys and Tutorials*, vol. 13, no. 3, pp. 443–461, 2011.

[8] T. Le, K. Mayaram, and T. Fiez, "Efficient far-field radio frequency energy harvesting for passively powered sensor networks," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 5, pp. 1287–1302, 2008.

[9] R. J. M. Vullers, R. van Schaijk, I. Doms, C. Van Hoof, and R. Mertens, "Micropower energy harvesting," *Solid-State Electronics*, vol. 53, no. 7, pp. 684–693, 2009.

[10] M. Piñuela, P. D. Mitcheson, and S. Lucyszyn, "Ambient RF energy harvesting in urban and semi-urban environments," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 7, pp. 2715–2726, 2013.

[11] L. R. Varshney, "Transporting information and energy simultaneously," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '08)*, pp. 1612–1616, IEEE, Toronto, Canada, July 2008.

[12] P. Grover and A. Sahai, "Shannon meets tesla: wireless information and power transfer," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '10)*, pp. 2363–2367, IEEE, Austin, Tex, USA, June 2010.

[13] L. Liu, R. Zhang, and K.-C. Chua, "Wireless information transfer with opportunistic energy harvesting," *IEEE Transactions on Wireless Communications*, vol. 12, no. 1, pp. 288–300, 2013.

[14] L. Liu, R. Zhang, and K.-C. Chua, "Wireless information and power transfer: a dynamic power splitting approach," *IEEE Transactions on Communications*, vol. 61, no. 9, pp. 3990–4001, 2013.

[15] A. A. Nasir, X. Zhou, S. Durrani, and R. A. Kennedy, "Relaying protocols for wireless energy harvesting and information processing," *IEEE Transactions on Wireless Communications*, vol. 12, no. 7, pp. 3622–3636, 2013.

[16] C. Shen, W.-C. Li, and T.-H. Chang, "Wireless information and energy transfer in multi-antenna interference channel," *IEEE Transactions on Signal Processing*, vol. 62, no. 23, pp. 6249–6264, 2014.

[17] D. W. K. Ng, E. S. Lo, and R. Schober, "Wireless information and power transfer: energy efficiency optimization in OFDMA systems," *IEEE Transactions on Wireless Communications*, vol. 12, no. 12, pp. 6352–6370, 2013.

[18] R. Morsi, D. S. Michalopoulos, and R. Schober, "Multiuser scheduling schemes for simultaneous wireless information and power transfer over fading channels," *IEEE Transactions on Wireless Communications*, vol. 14, no. 4, pp. 1967–1982, 2015.

[19] R. A. Berry and R. G. Gallager, "Communication over fading channels with delay constraints," *IEEE Transactions on Information Theory*, vol. 48, no. 5, pp. 1135–1149, 2002.

[20] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, 2003.

[21] J. Gorski, F. Pfeuffer, and K. Klamroth, "Biconvex sets and optimization with biconvex functions: a survey and extensions," *Mathematical Methods of Operations Research*, vol. 66, no. 3, pp. 373–407, 2007.