Joint Interference Coordination and Load Balancing for OFDMA Multihop Cellular Networks

Yue Zhao, *Student Member, IEEE*, Xuming Fang, *Member, IEEE*, Rongsheng Huang, *Member, IEEE*, and Yuguang Fang, *Fellow, IEEE*

Abstract—Multihop cellular networks (MCNs) have drawn tremendous attention due to its high throughput and extensive coverage. However, there are still three issues not well addressed. With the existence of relay stations (RSs), how to efficiently allocate frequency resource to relay links becomes a challenging design issue. For mobile stations (MSs) near the cell edge, cochannel interference (CCI) become severe, which significantly affects the network performance. Furthermore, the unbalanced user distribution will result in traffic congestion and inability to guarantee quality of service (QoS). To address these problems, we propose a quantitative study on adaptive resource allocation schemes by jointly considering interference coordination (IC) and load balancing (LB) in MCNs. In this paper, we focus on the downlink of OFDMA-based MCNs with time division duplex (TDD) mode, and analyze the characteristics of resource allocation according to IEEE 802.16j/m specification. We also design a novel frequency reuse scheme to mitigate interference and maintain high spectral efficiency, and provide practical LB-based handover mechanisms which can evenly distribute the traffic and guarantee users' QoS. Our study shows that our scheme not only meets the requirement on coverage, but also improves the throughput while accommodating more users in MCNs.

Index Terms—Multihop cellular networks, resource scheduling, OFDMA/TDD, interference coordination, load balancing

1 Introduction

The future wireless cellular networks, such as 3GPP advanced long term evolution (LTE-Advanced) [1] and IEEE 802.16m systems [2], will adopt orthogonal frequency division multiple access (OFDMA) technology for multihop cellular networks (MCNs). OFDMA is regarded as the most promising physical layer technology for the fourth generation (4G) wireless networks. New relay strategies and technologies are proposed to provide services with extended coverage and higher data rate. Fixed relay stations (RSs) with fewer functionalities than base stations (BSs) can be deployed to overcome poor channel conditions while maintaining low infrastructure cost [3]. Nevertheless, MCNs have inherent drawbacks, for example, extra radio resource are required on relay links (BS-RS links) [4], [5]. Therefore, well-designed radio resource allocation schemes are crucial for MCNs to effectively exploit the benefit of RSs, while overcoming the disadvantages.

- Y. Zhao and X. Fang are with the Provincial Key Lab of Information Coding and Transmission, Southwest Jiaotong University, Sichuan, Chengdu 610031, China. E-mail: yuezhao@yahoo.com.cn, xmfang@swjtu.edu.cn.
- R. Huang is with the Olympus Communication Technology of America, Suite 450, 9605 Scranton Road, San Diego, CA 92121. E-mail: rhuang@olympus-cta.com.
- Y. Fang is with the Department of Electrical and Computer Engineering, University of Florida, 435 New Engineering Building, PO Box 116130, Gainesville, FL 32611-6130. E-mail: fang@ece.ufl.edu.

Manuscript received 28 Aug. 2011; revised 27 May 2012; accepted 4 Oct. 2012; published online 22 Oct. 2012.
For information on obtaining reprints of this article, please send e-mail to: tmc@computer.org, and reference IEEECS Log Number TMC-2011-08-0474.
Digital Object Identifier no. 10.1109/TMC.2012.224.

Since RSs always utilizes the same spectrum as MSs or BSs, cochannel interference (CCI) will be closely related to the radio resource allocation schemes in MCNs due to the intercell and intracell frequency reuse. OFDMA systems should employ frequency planning for better cell edge performance and the ease of interference management [6]. Traditional single-hop cellular networks (SCNs) typically employ the frequency reuse pattern with factor of 3 or 7 to reduce CCI, which results in low spectral efficiency. As we all know, high data rate is one of the desired features of the future cellular networks. It requires a highly efficient utilization of the available spectrum. Frequency reuse with factor of 1 is likely to be used in LTE-Advanced and IEEE 802.16m systems, aiming at improving the spectral efficiency [7], [8], [9], [10]. However, the CCI using this frequency planning causes severe performance degradation at cell boundaries.

Following the worldwide interoperability for microwave access (WiMAX) Forum [11], [12], the frequency reuse pattern can be denoted as $\mathrm{N} \times \mathrm{S} \times \mathrm{K}$, which means that the networks are divided into clusters of N cells (each cell in the cluster has a different frequency band), with S sectors and K different frequency bands per cell. According to these reuse patterns, all available spectrum is assigned to all sector-BSs in the reuse pattern of $1 \times 3 \times 1$, whereas each sector-BS uses only one third of the total frequency bands in the reuse pattern of $1 \times 3 \times 3$. These two frequency reuse patterns are shown in Fig. 1. The CCI level is higher in the former, whereas the spectral efficiency is lower in the latter. If $1 \times 3 \times 3$ is used in MCNs, the spectral efficiency will be much lower because extra frequency resource has to be

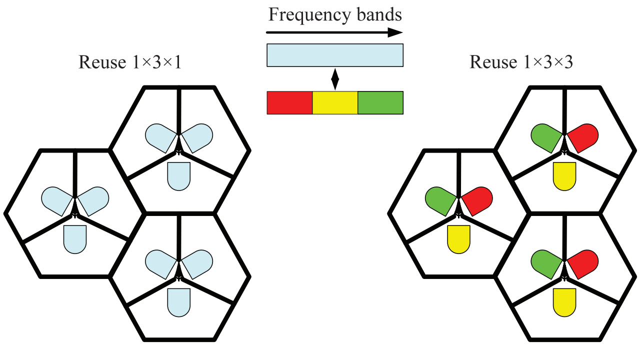


Fig. 1. Two basic frequency reuse patterns.

allocated to relay links. If $1 \times 3 \times 1$ is used in MCNs, the frequency reuse scheme is more important in a multicell scenario. Compared with BSs deployed at the cell center, RSs deployed at the cell edge cause serious interference because RSs are closer to the mobile stations (MSs) in the adjacent cells than those BSs.

In the existing literature, there are several works about reducing CCI in MCNs. In [13], several static resource allocation schemes with different partitions and reuse factors are discussed. The CCI in these schemes is analyzed in a multicell scenario. In [14], a relay-based orthogonal frequency planning strategy is proposed to improve cell edge performance. In [15], fractional frequency reuse (FFR) is extended to MCNs as a compromise solution to reduce CCI while maintaining the sector frequency reuse factor as 1. The main idea of FFR is to adopt frequency reuse $1 \times 3 \times 1$ at the cell center to maximize the network spectral efficiency while harnessing frequency reuse $1 \times 3 \times 3$ at the cell edge to alleviate CCI [11]. In [16], the minimum CCI has been achieved by adjusting the transmission (Tx) power at BSs and RSs under orthogonal frequency resource allocation. The essence of these works is to use partial frequency bands while maintaining frequency orthogonal at the cell edge and the remaining frequency bands at the cell center.

Moreover, there are several static frequency allocation schemes proposed in the aforementioned works, which fit for uniform traffic distribution only. In reality, users are not evenly distributed among cells. Too many users accessing one station (BS or RS) yields load imbalance in MCNs. Such an imbalance could severely affect the performance of hot spot areas, which may not meet the users' quality of service (QoS) requirements. This is another major reason for system performance degradation. To guarantee users' QoS, therefore, load balancing (LB) should be adopted along with IC for MCNs.

LB has been widely studied in SCNs and heterogeneous networks (HetNets). For SCNs, resource allocation schemes have to work in conjunction with the connection admission control (CAC) mechanisms, which determines, based on available resource and users' QoS, whether to admit an incoming connection to a particular cell or to reject it in the current cell, but to switch the user to an adjacent noncongested cell through a handover mechanism. Here, the corresponding handover mechanism is not executed due to position change of users, but due to the lack of resource in the original cell. As important methods in LB, the cell breathing and load-ware handover are proposed in [17] and [18]. The idea is that if a cell is heavily congested, the adjacent noncongested cell may expand the coverage and accommodate more users by raising transmission power. In [19], a scheme jointly considering IC and LB is designed to improve the weighted sum of data rates in multicell networks. Zhang and Rangarajan [19] show that the problem is NP-hard and then develop a local-improvement-based algorithm to solve it. These works suggest not only to use higher transmission power at the adjacent cell stations, but also report continually a large amount of information related to signal quality and traffic load in the surrounding cells, to the mobile switch center (MSC), to calculate the best connection to the BS. Apparently, this would increase the system overhead and management complexity. For HetNets, an integrated cellular and ad hoc relay (iCAR) system has been proposed in [20] and [21], in which some users can be switched to adjacent cells through ad hoc RSs and the spare resource are then acquired by incoming users. However, this type of LB only works with HetNets.

HetNets intend to change the traditional system architecture of cellular networks, while MCNs only attempt to improve the network performance of the traditional cellular networks through the use of RSs. It is noticeable that MCNs differ from HetNets in the following few characteristics: 1) RSs are important add-on communications facilities of cellular networks, which also share the same spectrum with BSs; 2) BSs and RSs are connected through wireless radio interfaces; and 3) the users associated with an RS need to access BS ultimately, which may ask for two-hop transmissions to deliver data.

With the deployment of RSs in MCNs, more handover opportunities arise, leading to better resource management and performance gain. This paper focuses on how to switch the connections from congested stations to noncongested stations and increase the available frequency resource for congested stations to achieve LB. In a cell, the traffic load information of RSs as well as link qualities between RSs and MSs are reported to BS by RSs. The BS is directly responsible for performing handover mechanisms in each sector. This method does not require to collect and process all kinds of information for a group of cells, which can reduce the complexity of the system implementation and guarantee QoS for users in hot spots.

The main contributions of this paper can be summarized as follows: We provide a quantitative study on an adaptive resource allocation scheme by jointly considering IC and LB in MCNs. We also present a novel frequency reuse scheme to mitigate interference and maintain high spectral efficiency, and propose practical LB-based handover mechanisms which can evenly distribute the traffic load and guarantee users' QoS. Extensive simulations demonstrate that our proposed schemes can provide higher throughput and accommodate more QoS-guaranteed users than what conventional SCNs can do.

The remainder of this paper is organized as follows: Section 2 describes the system model. Section 3 presents interference coordination (IC) and adaptive resource scheduling for MCNs. Section 4 gives two practical LB-based handover mechanisms according to different stations, and demonstrates that these mechanisms are able to improve

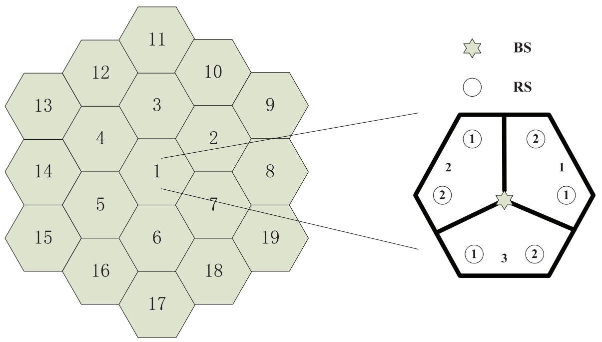


Fig. 2. Cell structure and numbering for cells/sectors/RSs in MCNs.

the system throughput and accommodate more users. Section 5 describes the simulation environment, and carries out performance analysis. Section 6 concludes this paper.

2 System Model

As shown in Fig. 2, we consider the MCNs consisting of 19 hexagonal cells with wrap-around model. In this scenario based on IEEE 802.16j/m specification, each cell is divided into three sectors with two above rooftop (ART) RSs deployed in each sector. The BS-RS distance is equal to 3/8 of the site-to-site distance and the angle between RS location and BS antenna boresight direction is 26 degree [7], [22]. For the sake of convenient expression below, in the jth sector of ith cell, let $N_{R(i,j,k)}$ be RS k, and $N_{B(i,j)}$ be the BS, where $i = 1, 2, \ldots, 19$, $j = 1, 2, 3$, $k = 1, 2$. In the first sector of the first cell, let N_M denote the MS, $N_{B(1,1)}$ and $N_{R(1,1,1)}$ be the candidate stations for N_M to access.

According to IEEE 802.16j/m specification [23], each time division duplex (TDD) frame consists of downlink and uplink subframes. Each subframe is subsequently divided into two time zones which are named as relay zone (RZ) and access zone (AZ), respectively. RZ is dedicated to the BS transmission toward both RSs and MSs, while AZ is dedicated to the reception of MSs from the BS or two RSs. Assuming each RS receives data for relaying in RZ at the current frame, it should be scheduled to transmit the data in AZ and empty its buffer at the next frame. In each subframe, the frequency domain consists of subchannels and the time domain consists of slots. A slot in a subchannel is the minimum frequency-time resource unit. Fig. 3 shows TDD relay frame structure for MCNs.

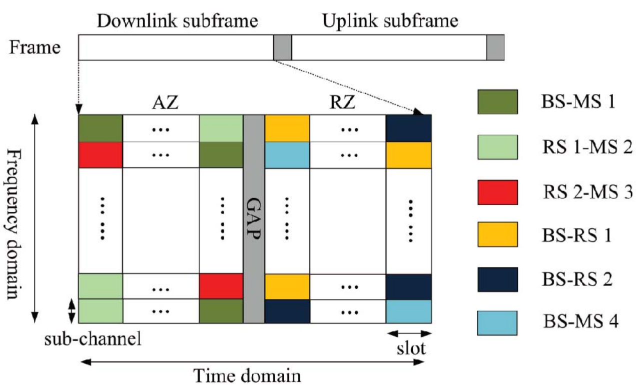


Fig. 3. TDD relay frame structure for MCNs.

TABLE 1
Antenna Models

Parameter			Value
BS	antenna pattern		$A(\theta) = -\min[12(\frac{\theta}{\theta_{3dB}})^2, A_m]$, θ_{3dB}=70°, A_m=30dB
	antenna gain (boresight)		17dBi
RS	BS-RS link	antenna pattern	$A(\theta) = -\min[12(\frac{\theta}{\theta_{3dB}})^2, A_m]$, θ_{3dB}=20°, A_m=23dB
		antenna gain (boresight)	20dBi
	RS-MS link	antenna pattern	Omni
		antenna gain	7dBi
MS	antenna pattern		Omni
	antenna gain		0dBi

The signal to interference plus noise ratio (SINR) on each link in each time zone is expressed as

$$\Gamma_{\mathrm{rz-br}} = \frac{H[N_{B(1,1)}, N_{R(1,1,1)}] \cdot P_B}{I_1 + N_0}, \tag{1}$$

$$\Gamma_{\mathrm{rz-bm}} = \frac{H[N_{B(1,1)}, N_M] \cdot P_B}{I_2 + N_0}, \tag{2}$$

$$\Gamma_{\mathrm{az-bm}} = \frac{H[N_{B(1,1)}, N_M] \cdot P_B}{I_3 + N_0}, \tag{3}$$

$$\Gamma_{\mathrm{az-rm}} = \frac{H[N_{R(1,1,1)}, N_M] \cdot P_R}{I_4 + N_0}, \tag{4}$$

where $\Gamma_{\mathrm{rz-br}}, \Gamma_{\mathrm{rz-bm}}, \Gamma_{\mathrm{az-bm}}, \Gamma_{\mathrm{az-rm}}$ are the SINR of RSs on BS-RS links in RZ, of MSs on BS-MS links in RZ, of MSs on BS-MS links in AZ, of MSs on RS-MS links in AZ, respectively. I_1, I_2, I_3, and I_4 are the total interference from the cochannel stations on the corresponding links.[1] P_B is the BS's Tx power and P_R is the RS's Tx power. In addition, N_0 is the additive white Gaussian noise power and $H(x, y)$ is the channel gain of link between x and y, which is defined as [7]

$$H(x, y) = [PL(x, y)]^{-1} \cdot 10^{\xi(x,y)/10} \cdot G_t(x, y) \cdot G_r(x, y), \tag{5}$$

where $PL(x, y)$ is the path loss from x to y. $\xi(x, y)$ is the shadow fading on the link between x and y. $G_t(x, y)$ is the transmitting antenna gain of x and $G_r(x, y)$ is the receiver antenna of gain of y. The values of (1)-(4), which are determined by the kinds of links between x and y, can be found or calculated from Tables 1 and 2 [7].

Adaptive modulation and coding (AMC) is utilized in the system. Slot efficiency means the data rate achieved by a slot within a TDD frame length. If M-ary ($M = 2^m$) modulation is chosen and the coding rate is $r, m \cdot r$ bits can be carried by one data subcarrier. A slot contains 48 data subcarriers, and then slot efficiency can be defined as

$$\eta(m, r, T_F) = \frac{48 \cdot m \cdot r}{T_F}, \tag{6}$$

where T_F is the TDD frame length.

1. See Section 3.1 for the detailed analysis and derivation of the interference on the corresponding links.

TABLE 2
Channel Models

Parameter		Value
Propagation Model	BS-RS link	Recommendation ITU-R M.1225 [24]
	BS-MS and RS-MS link	IEEE 802.16j EVM Type D [25]
Log-Normal Shadowing σ	BS-RS link	3.4dB
	BS-MS and RS-MS link	8dB

TABLE 3
Modulation and Coding Scheme

MCS Level	SINR(dB)	Slot Efficiency (Kbps/Slot)
QPSK(1/12)	-3.14	1.6
QPSK(1/6)	-0.73	3.2
QPSK(1/3)	2.09	6.4
QPSK(1/2)	4.75	9.6
QPSK(2/3)	7.86	12.8
16QAM(1/2)	9.94	19.2
16QAM(2/3)	13.45	25.6
64QAM(2/3)	18.6	38.4
64QAM(5/6)	24.58	48

Table 3 lists the required SINR level for the given modulation and coding scheme (MCS) [26]. The MCS level is determined by the SINR of a link. Γ denotes the SINR of a link, and $\eta(\Gamma)$ denotes slot efficiency corresponding to SINR level. The mapping relationships among MCS level, Γ, and $\eta(\Gamma)$ can be found in Table 3.

Let L_{rz-br}, L_{rz-bm}, L_{az-bm}, and L_{az-rm} represent the BS-RS links in RZ, BS-MS links in RZ, BS-MS links in AZ, and RS-MS links in AZ, respectively. The first kind of links is relay link and the last three kinds of links are access links, i.e., $L_R = \{L_{rz-br}\}$ and $L_A = \{L_{rz-bm}, L_{az-bm}, L_{az-rm}\}$. Each user can access the BS directly over a single hop on L_{rz-bm} and on L_{az-bm}, or otherwise, establish a two-hop path by using both L_{rz-br} and L_{az-rm}.

We suppose each user is allocated with the amount of data D. The corresponding slot efficiency of each link is represented by $\eta(\Gamma_{rz-bm})$, $\eta(\Gamma_{az-bm})$, $\eta(\Gamma_{rz-br})$, and $\eta(\Gamma_{az-rm})$. S_{rz-bm}, S_{az-bm}, S_{rz-br}, and S_{az-rm} represent the numbers of slots allocated on the corresponding links.

When the two-hop path is selected, the cost of selecting the RS as access station can be represented as

$$\begin{aligned} S_{brm} &= S_{rz-br} + S_{az-rm} \\ &= \left\lceil \frac{D}{\eta(\Gamma_{rz-br})} \right\rceil + \left\lceil \frac{D}{\eta(\Gamma_{az-rm})} \right\rceil, \end{aligned} \tag{7}$$

where S_{brm} represents the cost of accessing the BS indirectly via an RS, and $\lceil x \rceil$ represents the smallest integer larger than or equal to x.

Similarly, the cost of selecting to access the BS can be represented as

$$\begin{aligned} S_{bm} &= \min(S_{rz-bm}, S_{az-bm}) \\ &= \min\left(\left\lceil \frac{D}{\eta(\Gamma_{rz-bm})} \right\rceil, \left\lceil \frac{D}{\eta(\Gamma_{az-bm})} \right\rceil \right), \end{aligned} \tag{8}$$

where S_{bm} represents the cost of accessing the BS directly. Choice between L_{rz-bm} and L_{az-bm} to access BS depends on which one can offer the user a higher slot efficiency.

The station with the minimum cost will be selected as the access station because the system can use the smallest slot resource to transmit the same amount of data. If there is not enough resource to support the optimal selection, suboptimal path will be the substitute until the MS is served or rejected.

It is noticeable that wireless mesh networks (WMNs) (such as IEEE 802.11s WMNs) and MCNs are different in dealing with the problem of path selection or routing. The most concern for WMNs is to find an optimal way (e.g., with lowest delay) to support a fixed or mobile terminal to a gateway to access Internet [27], while for MCNs, it is to find an optimal station (BS or RS, with maximum throughput or the lowest energy consumption) to support an MS to access cellular networks. Additionally, for WMNs, the frequency spectrum is shared and randomly contended by all stations. The access scheme with the lowest overhead is optimal. However, for example, in this paper, a centrally controlled optimal resource allocation for OFDMA-based MCNs is our target.

To provide analytical performance evaluation, we make two assumptions for the remainder of this paper:

1. All users have a single type of data service and thus have the same QoS requirements.
2. All cells/sectors have the same channel conditions, traffic load, and distribution of users.

3 Interference Coordination and Resource Scheduling for MCNs

In this section, we present our proposed schemes to efficiently utilize the resource. The idea is to mitigate the interference by better coordination of transmissions (i.e., interference coordination) and design more efficient resource scheduling algorithms.

3.1 Frequency Reuse Schemes and Interference Analysis

We present our IC scheme in Fig. 4. Fig. 4a shows downlink subframe structure for MCNs; Figs. 4b and 4c show different frequency reuse schemes in different time zones.[2] The system bandwidth is W, divided equally into three different subbands: W_1 (red), W_2 (yellow), and W_3 (green). We differentiate subbands adopted by each station throughout cells with different colors. BSs (and RSs) allocate power equally to all subchannels, which has been often used for implementation simplicity as well as analytical tractability in downlink resource allocation problems [28], [29].

In RZ, the RS cannot transmit data to MSs in the receiver mode. In this case, it is unnecessary to allocate frequency resource to RSs. As shown in Fig. 4b, the total frequency band is allocated to the BS in each sector per cell with the $1 \times 3 \times 1$ reuse pattern. By taking advantage of the positions of RSs, MSs located at the cell/sector edge might access RSs

2. In Figs. 4b and 4c, we take a 3-cell topology model as an example. In fact, we adopt a 19-cell topology model for performance analysis and simulations. It is similar to Figs. 5b and 5c.

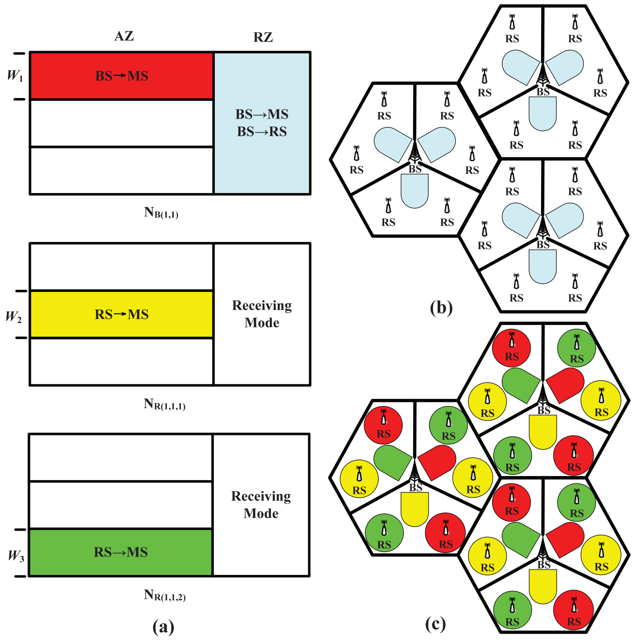


Fig. 4. IC scheme in MCNs: (a) downlink subframe structure (the first sector), (b) frequency reuse scheme in RZ, and (c) frequency reuse scheme in AZ.

in AZ for better link quality. Similarly, MSs distributed near the cell center are able to access the BS directly. With the increased intercell/sector interference distance, similar to FFR scheme, the system is less interfered by using $1 \times 3 \times 1$ in RZ than in SCNs. Therefore, it is efficient for MCNs to apply the reuse pattern of $1 \times 3 \times 1$ in RZ.

In AZ, the RS delivers data from the BS to the MS and frequency resource should be allocated to the RS. As shown in Fig. 4c, three different subbands are allocated to a BS and two RSs in each sector for mitigating intrasector interference among these stations. Taking the first sector, for example, when W_1 is allocated to $N_{B(1,1)}$, W_2 to $N_{R(1,1,1)}$, and W_3 to $N_{R(1,1,2)}$, the total frequency is allocated to every sector by means of $1 \times 3 \times 1$. The RSs located adjacent to each other share different frequency bands. Only the two RSs located at the opposite angles of the hexagonal cell can reuse the same frequency bands. Therefore, the CCI from the adjacent-sector/cell of RSs deployed at the cell edge is less severe. Consequently, this scheme can meet the requirements of both fully utilizing the available frequency resource in each cell and strictly planned orthogonal frequency bands in the edge area of cells/sectors.

Next, we analyze the interference on different links in different time zones for performance evaluation. In RZ, the interference for RSs/MSs comes from the 56 BSs of the 56 sectors in all 19 cells, except for the sector where the measured RSs/MSs are located. Then, the total interference on BS-RS links (I_1) and BS-MS links (I_2) can be expressed as

$$I_1 = \sum_{i,j,(i,j)\neq(1,1)} H[N_{B(i,j)}, N_{R(1,1,1)}] \cdot P_B, \tag{9}$$

$$I_2 = \sum_{i,j,(i,j)\neq(1,1)} H[N_{B(i,j)}, N_M] \cdot P_B. \tag{10}$$

In AZ, when the proposed reuse scheme is adopted, there are three kinds of interference suffered by MSs on BS-MS links. The first kind of interference comes from 18 BSs located in the first sector of the other 18 cells. The second kind of interference comes from RS1s located in the second sector of all cells. The third kind of interference comes from RS2s located in the third sector of all cells. Thus, the total interference on BS-MS links (I_3) can be estimated by

$$\begin{aligned} I_3 = &\sum_{i,i\neq 1} H[N_{B(i,1)}, N_M] \cdot P_B + \sum_{i} H[N_{R(i,2,1)}, N_M] \cdot P_R \\ &+ \sum_{i} H[N_{R(i,3,2)}, N_M] \cdot P_R. \end{aligned} \tag{11}$$

Similarly, the total interference suffered by MSs on RS-MS links (I_4) can be calculated by

$$\begin{aligned} I_4 = &\sum_{i} H[N_{B(i,3)}, N_M] \cdot P_B + \sum_{i,i\neq 1} H[N_{R(i,1,1)}, N_M] \cdot P_R \\ &+ \sum_{i} H[N_{R(i,2,2)}, N_M] \cdot P_R. \end{aligned} \tag{12}$$

3.2 Resource Scheduling for MCNs

Since the combination of IC and efficient resource scheduling can further improve system performance, we then extend the proportional fair (PF) algorithm for MCNs in this section.

Besides the PF algorithm, the other two classical scheduling algorithms of round robin (RR) [30] and maximum SINR (MaxSINR) [4] are often applied to cellular networks. In RR algorithm, slots are allocated to the users in the cell coverage in due order and thus seem to be absolutely fair. Nonetheless, it is not efficient since the difference of slot efficiency of users is not taken into consideration. In MaxSINR algorithm, slots are allocated to the users with the highest SINR at per scheduling instant, which can maximize the system throughput, but it is not fair since the users with low slot efficiency are not guaranteed to obtain slots.

The PF algorithm has been investigated in the literature of scheduling in SCNs [31], [32]. It provides an efficient throughput-fairness tradeoff. In MCNs, the BS is responsible for gathering link information and allocating the available resource to the corresponding links according to the PF algorithm.

For access links, we assume that N_l is the set of users on link l, and denote the slot assignment indicator by $X_{s,l}^n$. $X_{s,l}^n = 1$ when user n is allocated to slot s on link l, and 0 otherwise. We define R_l^n to be the data rate of user n on link l

$$R_l^n = \eta(\Gamma_l^n) \cdot S_l^n, \tag{13}$$

where $\eta(\Gamma_l^n)$ represents the slot efficiency of user n on link l, and S_l^n the number of slots allocated to user n on link l. S_l^n can be expressed as

$$S_l^n = \sum_{s\in S_l} X_{s,l}^n \quad \forall n \in N_l \quad \forall l \in L_A, \tag{14}$$

where S_l denotes the set of the assignable slots on link l. Since a slot can only be assigned to one user on each link, we should obey the following constraint:

$$\sum_{n \in N_l} X_{s,l}^n = 1 \quad \forall s \in S_l \quad \forall l \in L_A. \tag{15}$$

In one sector, one BS and two RSs can reuse the same slot for transmitting data to different users in AZ [33], [34]. Therefore, we can derive the constraint as follows:

$$\sum_{l \in L_A} \sum_{n \in N_l} X_{s,l}^n \leq 3 \quad \forall s \in S_l. \tag{16}$$

The PF metric $M_l^n(s)$ for user n on link l is given by

$$M_l^n(s) = \frac{\eta(\Gamma_l^n)}{\overline{R}_l^n(s-1)}, \tag{17}$$

where $\overline{R}_l^n(s)$ denotes the average data rate of user n from the start of the frame to the allocation of slot s in the frame. $\overline{R}_l^n(s)$ will be updated after user n is allocated to slot s

$$\overline{R}_l^n(s) = \left(1 - \frac{1}{T_c}\right)\overline{R}_l^n(s-1) + \frac{1}{T_c}\left[\eta(\Gamma_l^n)\right], \tag{18}$$

where T_c represents the latency scale of PF [35]. Slot s is assigned to user n^* with the highest PF metric on link l

$$n^* = \arg\max_{n \in N_l} \{M_l^n(s)\}. \tag{19}$$

For relay links, based on the allocation result of the second-hop links, slots should be assigned to first-hop link with proportion to the aggregate data rate of the second-hop link of each RS. Notice that the resource allocation to the first-hop link via each RS will end when the first-hop data rate is greater than or equal to the aggregate second-hop data rate. The other slots of RZ are assigned to BS-MS links according to (8). Considering the assignable slots of one frame are limited, the attainable balance of slot allocation determines the ratio of RZ and AZ in the time domain in each frame. The detailed algorithm is shown in Algorithm 1.

Algorithm 1. Resource Scheduling Algorithm.

Inputs: $n \in N_l$, $N_l \neq \emptyset$, and S_D
Outputs: S_{AZ}, S_{RZ}, and $X_{s,l}^n$
Initialization: $s = 0, S_l^n = 0$
while $s < S_D$ **do**
  $s = s + 1$
  if $l \in L_A$ **then**
    $n^* = \arg\max_n\{M_l^n(s)\}$
    $X_{s,l}^{n^*} = 1$
    $S_l^{n^*} = S_l^{n^*} + 1$
    $\overline{R}_l^{n^*}(s) \leftarrow \overline{R}_l^{n^*}(s-1)$
  else
    for RS k $= 1:2$ **do**
      $S_{rz-br}^k = \left\lceil \frac{\sum_{l \in \{L_{az-rm}\}} \sum_{n \in N_l} R_l^n}{\eta(\Gamma_{rz-br})} \right\rceil$
    end for
  end if
  $S_{RZ} = \sum_{k \in \{1,2\}} S_{rz-br}^k + \sum_{l \in \{L_{rz-bm}\}} \sum_{n \in N_l} S_l^n$
  $S_{AZ} = \sum_{l \in \{L_{az-rm}, L_{az-bm}\}} \sum_{n \in N_l} S_l^n$
end while

Note: S_{AZ}, S_{RZ}, and S_D represent the numbers of slots in AZ, in RZ, and in one downlink subframe, respectively.

4 Joint Interference Coordination and Load Balancing

Since traffic load distribution of each cell/sector affects the system performance significantly, we propose joint IC and LB (ICLB) for MCNs. The objective is to improve system throughput under the constraint of the basic requirement on coverage [7]. P_{cov}, the cell coverage probability, is defined as the percentage of area within the cell that has received SINR above the threshold of the most robust MCS, i.e., QPSK(1/12) modulation. Therefore, the coverage probability can be estimated as

$$P_{cov} = P\{\Gamma_l^n \geq Z_{th}[QPSK(1/12)]\} \quad \forall n \in N_l \quad \forall l \in L_A, \tag{20}$$

where $Z_{th}[QPSK(1/12)]$ is the SINR threshold for QPSK(1/12) modulation.

Let R_{min} denote the user's minimum traffic rate under QoS requirements, and $\widetilde{R}_l^n(T_W)$ the average data rate for user n over fixed-length time window T_W [31]. If $\widetilde{R}_l^n(T_W)$ is less than R_{min}, it means that the user's QoS cannot be satisfied and new users will be blocked on link l because there is no more available resource to guarantee the QoS. Under this circumstance, we say the cell/sector is overloaded. Therefore, we obtain the constraint for every connected user as $\widetilde{R}_l^n(T_W) \geq R_{min}$.

Since the two-hop transmission with the minimum downlink data rate is usually the RS-MS link, the system throughput can be computed just by accumulating data rate on access links in a frame, i.e.,

$$T = \sum_{l \in L_A} \sum_{n \in N_l} R_l^n. \tag{21}$$

Since PF is applied to MCNs, increasing throughput implies that more users' QoS requirements are met. Therefore, system throughput is improved and more reliable service is attained. For different station types, we present two LB mechanisms to improve the system throughput.

4.1 Load Balancing for Traffic Load at RSs

When an RS is overloaded, it does not have enough frequency resource for the users nearby. We aim at switching some users from the congested RS to some noncongested BSs in the same sector through the handover mechanism to keep traffic load in balance and reduce the blocking probability. Since the low Tx power at RS imposes limitation on the coverage area of RS, the users associated with one RS cannot establish connections with the other RS in the same sector. Neither can the congested RS obtain any more spectrum because it would violate the IC rules that the users at the cell edge are only allowed to use orthogonal frequency bands with the adjacent-sector RSs. The obvious benefit from the handover mechanism is to replace two-hop transmission with one-hop transmission, thereby saving more resource for the rest of the users associated with RSs.

The following is the detailed procedure we propose for the handover mechanism.

Step 1 (measurement and report). The BS periodically measures and computes $\widetilde{R}_l^n(T_W)$ for any user associated with RSs so that the BS has the knowledge of the RS's load status. When a new user arrives at the RS, the BS will determine whether the RS is overloaded and announce the information about its own load status together.

Step 2 (decision and execution). If the RS is determined to be overloaded, and the BS has available resource to accept more users, some users originally associated with the RS will change their serving station from the RS to the BS. However, the BS cannot always cover the place where a new arrival arises, and the optimal value of the throughput cannot be obtained by the handover of the new arrival either. Therefore, the BS needs to calculate the expected data rate if a user is switched to the BS, and choose user n' that can achieve the largest benefit by switching serving station which turns out to be

$$n' = \underset{n \in N_l}{argmin}(S_{bm} - S_{brm}). \tag{22}$$

If more than one user achieves the same largest benefit, one user is randomly selected. The data rates of user n' have the smallest difference between its connections with the BS and the RS. If the user is a valid candidate, the handover is carried out and the user receives data through direct link from the BS. If the RS is still overloaded, the next eligible user will be selected to connect to the BS until the RS is no longer overloaded or the BS is unable to admit any other users.[3] If some users associated with the RS change their connections to the BS, while there are new users to connect to the BS, this may result in collision. Thus, the BS should first choose the new users associated with the BS to approach the optimal throughput because they have higher slot efficiency and less system overhead.

Step 3 (notification). To prevent possible ping-pong effects [36], the users switched to the BS are not switched back to the RS even if the traffic load is reduced at the RS. Because the handover mechanism between a BS and an RS is implemented in one sector, it brings less delay and system overheads than LB among cells in the traditional sense. Thus, the handover should be performed on a short time scale.

The handover mechanism aims at redistributing the partial traffic load from an RS to a BS to reduce heavy traffic load of the RS, which is consistent with our objective to increase the number of users in MCNs.

4.2 Load Balancing for Traffic Load at BSs

When a BS is overloaded, it is not helpful to switch the users associated with the BS to an RS in the same sector. Because these kinds of users are close to the BS and far from RSs, once handover occurs, the data rates of these users will decrease. Moreover, the users only need the good one-hop links to directly access the BS, whereas they need the poor two-hop links to re-access the BS via RSs. Therefore, when the overall concerns are taken into consideration, it will occupy additional resource in the system.

To increase the available frequency resource to accommodate more users, BS-MS links need to reuse the same frequency as RS-MS links in one sector. Consequently, the spectral efficiency will be improved as the opportunities of simultaneous transmissions with the same frequency increases. However, it is inevitable to cause additional interference. Our proposed FFR scheme in MCNs can efficiently improve the spectral efficiency while mitigating the interference. The frame structure and cell configuration in enabling FFR are presented in Fig. 5. The AZ of a BS is divided into inner zone and outer zone. In this scheme, the frequency reuse factor is 1 in the inner zone and 3 in the outer zone. The essence of the handover mechanism is to switch some users associated with the BS from the outer zone to the inner zone.

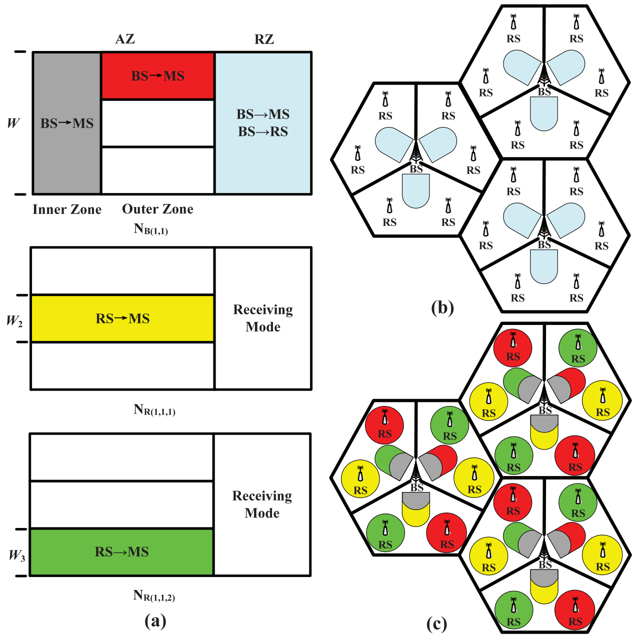


Fig. 5. FFR scheme in MCNs: (a) downlink subframe structure (the first sector), (b) frequency reuse scheme in RZ, and (c) frequency reuse scheme in AZ.

It must be noted that users may suffer from serious CCI, both intercell and intracell. Thus, the Tx power for users in the inner zone should be much lower than that in the outer zone to decrease the interference to the outside area. The BS assigns different Tx powers to users in the inner zone and the outer zone. Let the Tx power for users in the inner zone be represented by P_B^{in}, and the Tx power for the users in the outer zone be P_B^{out} which is equal to P_B. We define the power ratio of the Tx power for inner zone users to the Tx power for outer zone users as $r = P_B^{in}/P_B^{out}$[37]. The power ratio r varies from 0 to 1, which is equivalent to the frequency reuse factor changes from 3 to 1 gradually.

The handover mechanism of BS's traffic load is summarized in Algorithm 2, in which two key issues are taken into consideration. First, the criterion on deciding which users being switched from the outer zone to the inner zone is required. Second, we have to consider the sizes of the inner zone and the outer zone in each frame. The specific procedures are described as follows:

Step 1 (measurement and report). The BS is in charge of periodically measuring and computing $\widetilde{R}_l^n(T_W)$ of the users associated with the BS. When any new user accesses the BS, the BS will compare $\widetilde{R}_l^n(T_W)$ with R_{min} and determine whether the BS itself is overloaded or not.

3. This issue will be discussed in detail in Section 4.2.

Step 2 (decision and execution). If the BS is found overloaded when new users arrive, the handover mechanism will be executed to prevent heavy traffic load.

Algorithm 2. Handover Algorithm for Traffic Load at BSs.

1: Initialization: $l \in \{L_{az-bm}\}, r = 0, S_{in} = 0, S_{out} = 0, I[\,], O[\,]$

2: Handover Execution:

while $\exists n : \widetilde{R}_l^n(T_W) < R_{min}$ **do**

  $r = r + 0.1, P_B^{out} = P_B, P_B^{in} = r \cdot P_B^{out}$

  for $n \in N_l$ **do**

    if $\Gamma_{az-bm}^{in} \geq Z_{th}[64QAM(5/6)]$ **then**

      $i(1,n) = 1, o(1,n) = 0$

    else

      $i(1,n) = 0, o(1,n) = 1$

    end if

  end for

  for $s = 1 : S_{AZ}$ **do**

    if $\max_{n \in N_{in}}\{M_l^n(s)\} > \max_{n \in N_{out}}\{M_l^n(s)\}$ **then**

      $S_{in} = S_{in} + 1$, and update $M_l^n(s)$, $n \in N_{in}$

    else

      $S_{out} = S_{out} + 1$, and update $M_l^n(s)$, $n \in N_{out}$

    end if

    if $S_{in} + 3 \cdot S_{out} > S_{AZ}$ **then**

      break.

    end if

  end for

  if $T(r) < T(r - 0.1)$ and $P_{cov} < 95\%$ **then**

    $r = r - 0.1$

    break.

  end if

end while

3: Update the results to the BS, and end the algorithm.

Note: $I[\,]$ is the N_l column of vector of users in the inner zone, and $i(1,n)$ is the nth element of $I[\,]$.

$O[\,]$ is the N_l column of vector of users in the outer zone, and $o(1,n)$ is the nth element of $O[\,]$.

$T(r)$ is the throughput with power ratio at r [37].

The power ratio, r is initialized to 0. When the BS is overloaded, r is increased according to the actual load status. When the BS adopts Tx power P_B^{in}, the users associated with the BS with the highest MCS are switched to the inner zone. It demonstrates that these users can achieve the maximum of slot efficiency, no matter whether the Tx power is either P_B^{in} or P_B^{out}. Let Γ_{az-bm}^{in} represent the SINR on BS-MS links in AZ when Tx power of the BS is P_B^{in}. If Γ_{az-bm}^{in} conforms to (23), user n is allocated to the inner zone. Otherwise, user n is allocated to the outer zone.

$$\Gamma_{az-bm}^{in} \geq Z_{th}[64QAM(5/6)], \tag{23}$$

where $Z_{th}[64QAM(5/6)]$ is the SINR threshold of the 64QAM(5/6) modulation.

Besides, determining the proportions of the inner zone to the outer zone is also one important remaining problem. Let S_{in} and S_{out} be the numbers of slots allocated to the inner zone and the outer zone, respectively. As shown in Fig. 5, S_{out} contains slots of only one subband, whereas S_{in} contains slots of three subbands W_1, W_2, and W_3. Hence, some users are switched to other subbands to alleviate the shortage of slots in the outer zone. In AZ, S_{in} and S_{out} should conform to total slots constraint S_{AZ}, i.e., $S_{in} + 3S_{out} = S_{AZ}$, and users should access the BS according to the PF algorithm.

However, too many slots occupied by the inner zone users may consequently cause serious interference suffered by users associated with RSs, so the PF metric $M_l^n(s)$ is rewritten as

$$M_l^n(s) = \begin{cases} \delta \cdot \dfrac{\eta(\Gamma_l^n)}{\overline{R}_l^n(s-1)} & n \in N_{in}, \\ \dfrac{\eta(\Gamma_l^n)}{\overline{R}_l^n(s-1)} & n \in N_{out}, \end{cases}$$

where N_{in} and N_{out} represent the sets of the users in the inner zone and the outer zone, respectively. δ is a fairness factor with $\delta \in (0,1)$. The algorithm is revised to ensure the outer zone users' priority to obtain slots, while fairness is guaranteed to all users throughout the cells.

We observe P_B^{in} varies as r changes. In AZ, the SINR on a BS-MS link in the inner zone can be estimated as

$$\Gamma_{\mathrm{az-bm}}^{in} = \frac{H[N_{B(1,1)}, N_M] \cdot P_B^{in}}{I_3^{\varphi} + N_0}, \tag{24}$$

where φ represents the sub-band used on the BS-MS link in the inner zone, and it is estimated as

$$\varphi = \begin{cases} 1, & \text{when using sub-band } W_1 \text{ in the inner zone}, \\ 2, & \text{when using sub-band } W_2 \text{ in the inner zone}, \\ 3, & \text{when using sub-band } W_3 \text{ in the inner zone}. \end{cases}$$

In AZ, the total interference on BS-MS links (I_3^{φ}) with various φ can be expressed as

$$\begin{aligned} I_3^1 = &\sum_i H[N_{R(i,2,1)}, N_M] \cdot P_R + \sum_i H[N_{R(i,3,2)}, N_M] \cdot P_R \\ &+ \sum_{i,j,(i,j)\neq(1,1)} H[N_{B(i,j)}, N_M] \cdot P_B^{in}, \end{aligned} \tag{25}$$

$$\begin{aligned} I_3^2 = &\sum_i H[N_{R(i,1,1)}, N_M] \cdot P_R + \sum_i H[N_{R(i,2,2)}, N_M] \cdot P_R \\ &+ \sum_{i,j,(i,j)\neq(1,1)} H[N_{B(i,j)}, N_M] \cdot P_B^{in}, \end{aligned} \tag{26}$$

$$\begin{aligned} I_3^3 = &\sum_i H[N_{R(i,1,2)}, N_M] \cdot P_R + \sum_i H[N_{R(i,3,1)}, N_M] \cdot P_R \\ &+ \sum_{i,j,(i,j)\neq(1,1)} H[N_{B(i,j)}, N_M] \cdot P_B^{in}. \end{aligned} \tag{27}$$

We observe that, $\Gamma_{\mathrm{az-bm}}^{out}$, the SINR on the BS-MS link in the outer zone, is equal to $\Gamma_{\mathrm{az-bm}}$ of (3); $\Gamma_{\mathrm{rz-br}}$ and $\Gamma_{\mathrm{rz-bm}}$ can still be calculated by (1) and (2), respectively. In addition, the SINR on the RS-MS link is affected because BS-MS links reuse the same slots as RS-MS links

$$\Gamma_{\mathrm{az-rm}}^{\omega} = \frac{H[N_{R(1,1,1)}, N_M] \cdot P_R}{I_4^{\omega} + N_0}, \tag{28}$$

where ω is 1 if slots used by an RS1-MS link are reused by the BS-MS link in the inner zone, and ω is 2 if in the outer zone. With different ω, the total interference on the RS-MS links (I_4^{ω}) can be expressed as

$$I_4^1 = \sum_{i,i\neq 1} H[N_{R(i,1,1)}, N_M] \cdot P_R + \sum_{i} H[N_{R(i,2,2)}, N_M] \cdot P_R + \sum_{i,j} H[N_{B(i,j)}, N_M] \cdot P_B^{in}, \quad (29)$$

$$I_4^2 = \sum_{i,i\neq 1} H[N_{R(i,1,1)}, N_M] \cdot P_R + \sum_{i} H[N_{R(i,2,2)}, N_M] \cdot P_R + \sum_{i} H[N_{B(i,3)}, N_M] \cdot P_B^{out}. \quad (30)$$

Notice that the change of throughput is affected only by the handover of the users from the outer zone to the inner zone. The BS can calculate the throughput and determine whether the handover condition is satisfied or not.

Due to the constraints mentioned above, there are still two handover conditions that should be considered. First, the coverage probability is higher than the required coverage, i.e., 95 percent. Second, the handovers of some users to the inner zone improve the throughput. If either of the two conditions is not satisfied, the Tx power for users in the inner zone is reduced and the handover mechanism is terminated.

Step 3 (notification). If the handover condition is not satisfied, new users will be rejected because admitting new users will deteriorate the network performance. Otherwise, the BS will carry out the handover mechanism and accept new users. This type of handovers is a procedure to switch the frequency band currently being used rather than an actual handover.

Admittedly, our two handover mechanisms may cause additional signaling overhead. But the effect is far less than that of LB among cells in the traditional SCNs. In addition, signaling transmission is implemented on control subchannels, so it is independent of data transmissions on traffic subchannels. Therefore, the signaling overhead is supportable in OFDMA systems, and has no negative effects on the system throughput as long as the control subchannels are not congested.

5 Performance Evaluation

In this section, we evaluate our proposed schemes by using simulation study.

5.1 Simulation Model and Parameters

The detailed system-level simulation parameters are shown in Table 4. To provide more realistic simulation results, we have investigated the performance of our scheme in a dynamic setting. Users arrive according to a Poisson process with rate λ and depart from the cell after holding times which are exponentially distributed with mean of $1/\mu = 100$ sec. During the holding time, we assume that users adopt full buffer traffic model and do not change their positions. The traffic load of a cell, $\rho = \frac{\lambda}{\mu}$, i.e., the average number of users in the cell, can be changed by choosing different arrival rate λ.

For an illustrative purpose, only the measurement of the center cell are plotted to show performance results. We will investigate the performance in terms of different load distributions and different traffic load degrees.

TABLE 4
System-Level Simulation Parameters [7], [38]

Parameter		Value
Number of cells		19
Number of sectors per cell		3
Channel bandwidth		10 MHz
Number of points in full FFT		1024
Number of sub-channels		30
Number of data subcarriers		720
Carrier frequency		2.5GHz
Frame length		5ms
Number of slots per traffic sub-channel		TDD (11 slots for downlink) [40]
Number of RSs		2 RSs per sector
BS-BS distance		1.5Km
BS-RS distance		3/8 of BS-BS distance
Angle between RS location and BS antenna boresight direction(φ)		26°
BS	Tx power	46dBm per antenna
	antenna height	32m
	cable loss	3dB
RS	Tx power	36dBm per antenna
	antenna height	32m
	noise figure	5dB
	cable loss	2dB
MS	antenna height	1.5m
	noise figure	7dB
Schedule model		PF
Minimum traffic rate		38.4kbps

5.2 Results and Discussions

There are four key performance metrics analyzed, i.e., the average cell spectral efficiency, the sector throughput, the cell coverage probability, and the QoS violation probability. The average cell spectral efficiency is defined as the ratio of the aggregate cell throughput to the channel bandwidth of one cell [7]. QoS violation probability represents the percentage of users whose $\widetilde{R}_l^n(T_W)$ is lower than R_{min} [31]. Besides IC and ICLB schemes, we also study the universal frequency reuse scheme for the cluster size $N = 1$ (N/A) [40], the novel frequency reuse scheme (NFRS) [16] and the orthogonal resource allocation algorithm (ORAA) [14]. We adopt the same system model and parameters when verifying these schemes.

At first, we assume users are under uniform distribution per cell, and set the total traffic load of a cell at 150. We make performance comparisons between SCNs and MCNs which adopt different scheduling algorithms. Fig. 6 shows the average cell spectral efficiency in terms of various scheduling algorithms, considering six frequency reuse patterns, i.e., frequency reuse $1 \times 3 \times 3$ and $1 \times 3 \times 1$ for SCNs, and N/A, NFRS, ORAA, and IC for MCNs. No matter what frequency reuse pattern is adopted, the channel bandwidth is W per cell. The results show that MCNs always outperform SCNs due to the fact that RSs can provide higher data rate for users in cell edge and access more users in remote areas. For SCNs, the pattern of $1 \times 3 \times 1$ outperforms the pattern of $1 \times 3 \times 3$, so it shows that the compact frequency reuse pattern of the three sectors

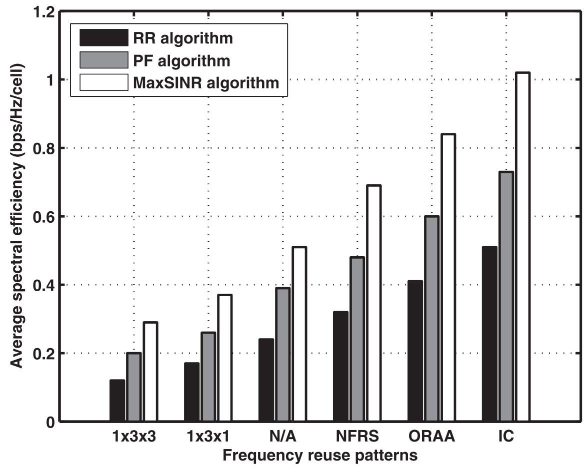


Fig. 6. Performance comparison of cell spectral efficiency.

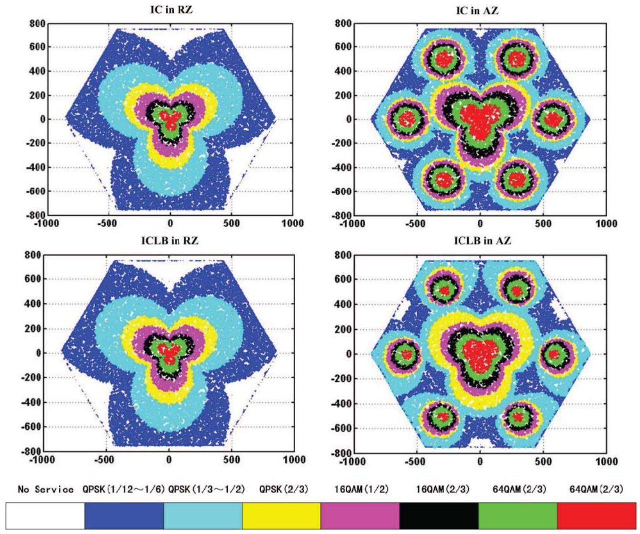


Fig. 8. SINR distribution in RZ and AZ.

reusing the same resource in one cell can enhance the cell spectral efficiency. With regard to scheduling algorithms, MaxSINR algorithm has better spectral efficiency than RR algorithm and PF algorithm. However, there is the unfairness nature of MaxSINR algorithm. RR algorithm without considering the slot efficiency of users leads to the worst performance. We can observe that the PF algorithm offers significantly improved performance compared with the RR algorithm and takes fairness into account which MaxSINR does not. With superiority mentioned above, the PF is the most competitive scheduling algorithm for MCNs. Therefore, we adopt it as the scheduling algorithm for subsequent performance simulations.

The following schemes are evaluated to determine the sector throughput in terms of different sector traffic loads. The results of sector throughput are shown in Fig. 7. For N/A, the sector throughput gets the minimum value because of its low spectral efficiency. Also, there is serious CCI at the cell/sector edge, which deteriorates the sector throughput even further. For IC, the total frequency is allocated to the BS in the RZ, which can greatly improve spectral efficiency. That is why IC offers better performance than NFRS and ORAA. When ICLB is used for the overloaded situation ($\rho \geq 55$), the sector throughput will keep increasing until the handover conditions cannot be satisfied. When ρ is higher than 70, the sector throughput and power ratio tend to be stable in spite of the increase in the traffic load. Accordingly, we find that the optimum power ratio is 0.3. Specifically, when traffic load of a sector is 70, the sector throughput with IC is 2.52 Mbps and the sector throughput with ICLB is 2.81 Mbps. Compared to ORAA, IC, and ICLB improve the sector throughput by 17.4 and 26.1 percent, respectively.

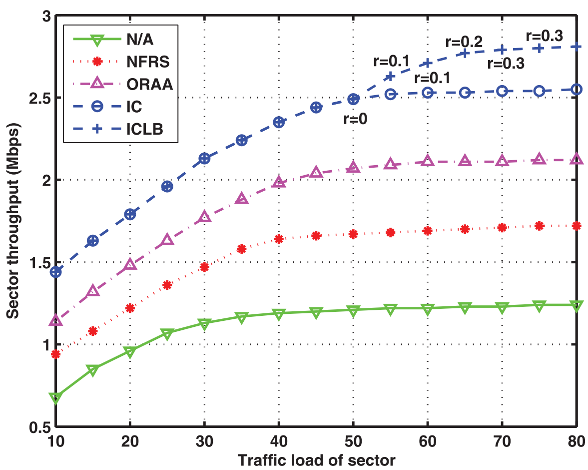


Fig. 7. Sector throughput with five different schemes.

Compared with IC, the coverage probability is inevitably decreasing while the sector throughput improves with ICLB. It is caused by the intracell interference between BS and RSs. Fig. 8 represents the SINR distribution of IC and ICLB in AZ and RZ when the power ratio is 0.3. The white area represents the area without service where the quality of signal received by the MS is below the SINR threshold for QPSK (1/12). The rest of colors represent areas with various degrees of MCS levels, respectively. Both IC and ICLB have the same SINR distribution and coverage probability in the RZ. However, IC is superior to ICLB in both SINR distribution and coverage probability in AZ. The coverage probability with IC is 99.18 percent and with ICLB is 95.57 percent. According to the results of comprehensive comparisons in terms of coverage probability and sector throughput, we conclude that ICLB obtains better sector throughput and accommodates larger number of users than IC, though it has lower SINR distribution. According to the results of comprehensive comparisons in terms of coverage probability and sector throughput, ICLB attains better throughput under the condition that its coverage probability is higher than 95 percent.

Next, we evaluate the performance in different schemes under the heterogeneous user distribution. The whole cell area can be partitioned into two nonoverlapping concentric areas, i.e., the center area and the edge area. The center area is within 450 meters of BS, and the rest area is the edge area. We set the total traffic load in a sector at 50, and change the user distribution by varying traffic loads in the center area and the edge area, respectively. Fig. 9 shows the sector throughput under various user distribution. It

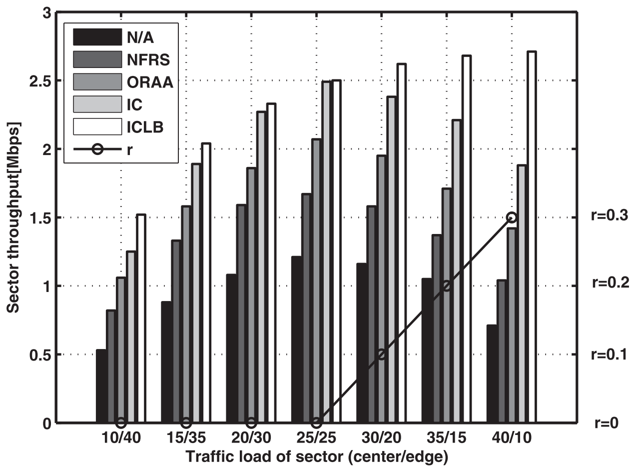


Fig. 9. Sector throughput with different user distribution.

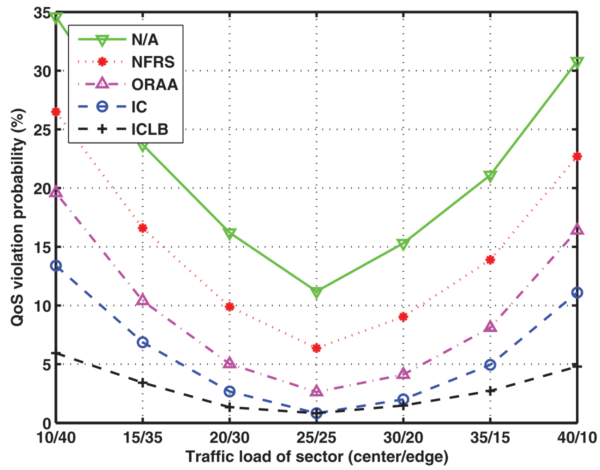


Fig. 10. QoS violation probability with five different schemes.

demonstrates that ICLB outperforms the other schemes because ICLB allows dynamic resource allocation according to the load status. When traffic load in the sector is 10/40, most of users are concentrated in the edge area and associated with RSs. ICLB can switch some users from RSs to BS and accommodate more users in the network. ICLB obtains performance gain 34.2 percent higher than ORAA and 22.6 percent higher than IC. When traffic load in the sector is 40/10, most of users are concentrated in the center area and associated with the BS. The sector throughput with ICLB is 2.67 Mbps, much better than 1.92 Mbps with IC. It benefits from increasing the amount of available resource at the BS and improving spectral efficiency. The gain depends on the value of r as we explained in the previous section. The performance gain decreases as r increases. When r is out of this optimal range, there may be degradation in the sector throughput or violation of the coverage requirement (e.g., coverage probability is higher than 95 percent).

QoS violated users are classified into two groups: the users who are unable to access to the network because of their low received SINR and the users who are unable to get enough resource to guarantee their requirements for QoS. As illustrated in Fig. 10, we calculate the percentage of users whose $\widetilde{R}_l^n(T_W)$ is lower than R_{min}. When the traffic load in the sector is 25/25, QoS violation probability is comparatively low because of the uniform user distribution. In this case, IC is equivalent to ICLB, which means there is no overload problem when adopting IC. When users are unevenly distributed, QoS violation probability increases with the aggravation of the degree of inhomogeneity. IC and ICLB have lower QoS violation probability than other schemes, mainly because IC not only guarantees good CCI control, but also improves spectral efficiency. ICLB, on the basis of IC, can accommodate more QoS-guaranteed users by utilizing handover mechanism. When the traffic load in the sector is 10/40, ICLB reduces IC and ORAA by 8.5 percent and 16.6 percent, respectively, in the QoS violation probability. When the traffic load in the sector is 40/10, ICLB reduces IC and ORAA by 7.2 and 14.2 percent, respectively, in the QoS violation probability.

6 Conclusion

In this paper, we have carried out a quantitative study on an adaptive resource allocation scheme based on interference coordination and load balancing for multihop cellular networks. We also propose a novel frequency reuse scheme to mitigate interference and maintain high spectral efficiency, and present practical LB-based handover mechanisms which can evenly distribute the traffic load and guarantee users' quality of service. Extensive simulations demonstrate that our scheme not only meets the requirement on coverage probability, but also improves the sector throughput and accommodates more users. To the best of our knowledge, this is the first work to provide dynamic resource allocation by jointly considering interference coordination and load balancing for MCNs. We expect that our method will play a significant role in network planning and resource allocation in the future MCNs.

Acknowledgments

The authors would like to thank Chi Zhang, Miao Pan, Hao Yue, and Zhengguang Zhao for their kind help and valuable suggestions. They would also like to thank the anonymous reviewers for their insightful comments which have significantly improved the quality of the paper. The work of Y. Zhao and X. Fang was partially supported by the 973 Program under grant 2012CB316100, the NSFC under grant 61071108 and 61032002, the the National Key Special Project under grant 2011ZX03004-003, and the 863 Program under grant 2011AA110504. The work of R. Huang and Y. Fang was partially supported by the US National Science Foundation under grants ECCS-1129062 and CNS-0916391. The work of Y. Fang was also partially supported by the National Natural Science Foundation of China under grant no. 61003300.

References

[1] S.W. Peters, A.Y. Panah, K.T. Truong, and R.W. Heath Jr., "Relay Architectures for 3GPP LTE-Advanced," *EURASIP J. Wireless Comm. and Networking*, article 618787, May 2009.

[2] *IEEE P802.16m/D6, Draft Amendment to IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems*, IEEE, May 2010.

[3] *IEEE P802.16j/D7, Draft Amendment to IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems: Multihop Relay Specification,* IEEE, Oct. 2008.

[4] M. Salem, A. Adinoyi, H. Yanikomeroglu, and D. Falconer, "Opportunities and Challenges in OFDMA-Based Cellular Relay Networks: A Radio Resource Management Perspective," *IEEE Trans. Vehicular Technology,* vol. 59, no. 5, pp. 2496-2510, Jan. 2010.

[5] Y. Zhao, X. Fang, and Z. Zhao, "Interference Coordination in Compact Frequency Reuse for Multihop Cellular Networks," *IEICE Trans. Fundamentals of Electronics, Comm. and Computer Sciences,* vol. E93-A, no. 11, pp. 2312-2319, Nov. 2010.

[6] Third Generation Partnership Project, "Technical Specification Group Radio Access Network; Physical Layer Aspects for Evolved Universal Terrestrial Radio Access (UTRA) (Release 7)," 3GPP Technical Report 25.814 v7.1.0, Sept. 2006.

[7] *IEEE 802.16m-08/004r3, IEEE 802.16m Evaluation Methodology Document (EMD),* IEEE, Oct. 2008.

[8] G. Boudreau, J. Panicker, N. Guo, R. Chang, N. Wang, and S. Vrzic, "Interference Coordination and Cancellation for 4G Networks," *IEEE Comm. Magazine,* vol. 47, no. 4, pp. 74-81, Apr. 2009.

[9] Third Generation Partnership Project, "Technical Specification Group Radio Access Network; Requirements for Further Advancements for E-UTRA (LTE-Advanced) (Release 8)," 3GPP Technical Report 36.913 V8.0.0, June 2008.

[10] Y. Wang, S. Kumar, L. Garcia, K.I. Pedersen, I.Z. Kovcs, S. Frattasi, N. Marchetti, and P.E. Mogensen, "Fixed Frequency Reuse for LTE-Advanced Systems in Local Area Scenarios," *Proc. IEEE 69th Vehicular Technology Conf.,* pp. 1-5, Apr. 2009.

[11] "Mobile WiMAX - Part I: A Technical Overview and Performance Evaluation," *WiMAX Forum,* Aug. 2006.

[12] "WiMAX System Evaluation Methodology," *WiMAX Forum,* Version 2.1, July 2008.

[13] L. Wang, Y. Ji, F. Liu, and J. Li, "Performance Improvement through Relay-Channel Partitioning and Reuse in OFDMA Multihop Cellular Networks," *Proc. IEEE Int'l Wireless Comm. and Mobile Computing Conf.,* pp. 177-182, Aug. 2008.

[14] W. Lee, M. Nguyen, and J. Jeong, "An Orthogonal Resource Allocation Algorithm to Improve the Performance of OFDMA-Based Cellular Wireless Systems Using Relays," *Proc. IEEE Fifth Consumer Comm. and Networking Conf.,* pp. 917-921, May 2008.

[15] Y. Zhao, X. Fang, X. Hu, Z. Zhao, and Y. Long, "Fractional Frequency Reuse Schemes and Performance Evaluation for OFDMA Multi-Hop Cellular Networks," *Proc. IEEE Fifth Testbeds and Research Infrastructures for the Development of Networks and Communities and Workshops,* pp. 1-5, Apr. 2009.

[16] M. Liang, F. Liu, Z. Chen, Y. Wang, and D. Yang, "A Novel Frequency Reuse Scheme for OFDMA Based Relay Enhanced Cellular Networks," *Proc. IEEE 69th Vehicular Technology Conf.,* pp. 1-5, Apr. 2009.

[17] A. Sang, X. Wang, M. Madihian, and R.D. Gitlin, "Coordinated Load Balancing, Handoff/Cell-Site Selection, and Scheduling in Multi-Cell Packet Data Systems," *Proc. 10th Ann. Int'l Conf. Mobile Computing and Networking,* pp. 302-314, Sept. 2004.

[18] D. Amzallag, R. Bar-Yehuda, D. Raz, and G. Scalosub, "Cell Selection in 4G Cellular Networks," *Proc. IEEE 27th Conf. Computer Comm.,* pp. 700-708, Apr. 2008.

[19] H. Zhang and S. Rangarajan, "Joint Load Balancing, Scheduling, and Interference Mitigation in Multi-Cell and Multi-Carrier Wireless Data Systems," *Proc. IEEE Seventh Int'l Symp. Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks,* pp. 1-10, June 2009.

[20] H. Wu, C. Qiao, S. De, and O. Tonguz, "Integrated Cellular and Ad Hoc Relaying Systems: iCAR," *IEEE J. Selected Areas in Comm.,* vol. 19, no. 10, pp. 2105-2115, Oct. 2001.

[21] E. Yanmaz and O.K. Tonguz, "Dynamic Load Balancing and Sharing Performance of Integrated Wireless Networks," *IEEE J. Selected Areas in Comm.,* vol. 22, no. 5, pp. 862-872, June 2004.

[22] L. Wang, W. Su, J. Huang, A. Chen, and C. Chang, "Optimal Relay Location in Multi-Hop Cellular Systems," *Proc. IEEE Wireless Comm. and Networking Conf.,* pp. 1306-1310, Mar. 2008.

[23] *IEEE 802.16m-08/003r7, IEEE 802.16m System Description Document (SDD),* IEEE, Feb. 2009.

[24] "Guidelines for Evaluation of Radio Transmission Technologies for IMT-2000," Technical Recommendation ITU-R M.1225, 1997.

[25] "Multi-Hop Relay System Evaluation Methodology (Channel Model and Performance Metric)," IEEE 802.16j-06/013r3, Feb. 2007.

[26] J.M. Ku, P.K. Kim, S.J. Lee, S. Shin, and C.G. Kang, "On the Performance of Broadband Mobile Internet Access System," *Proc. IEEE First Int'l Symp. Wireless Pervasive Computing,* pp. 1-6, Jan. 2006.

[27] G.R. Hiertz, D. Denteneer, S. Max, R. Taori, J. Cardona, L. Berlemann, and B. Walke, "IEEE 802.11S: The WLAN MESH Standard," *IEEE Wireless Comm. Magazine,* vol. 17, no. 1, pp. 104-111, Feb. 2010.

[28] G. Li and H. Liu, "Downlink Radio Resource Allocation for Multi-Cell OFDMA System," *IEEE Trans. Wireless Comm.,* vol. 5, no. 12, pp. 3451-3459, Dec. 2006.

[29] Z. Zhang, Y. He, and E.K.P. Chong, "Opportunistic Downlink Scheduling for Multiuser OFDM Systems," *Proc. IEEE Wireless Comm. and Networking Conf.,* pp. 1206-1212, Mar. 2005.

[30] L. Huang, M. Rong, L. Wang, Y. Xue, and E. Schulz, "Resource Scheduling for OFDMA/TDD Based Relay Enhanced Cellular Networks," *Proc. IEEE Wireless Comm. and Networking Conf.,* pp. 1544-1548, Mar. 2007.

[31] K. Son, S. Chong, and G. de Veciana, "Dynamic Association for Load Balancing and Interference Avoidance in Multi-Cell Networks," *IEEE Trans. Wireless Comm.,* vol. 8, no. 7, pp. 3566-3576, July 2009.

[32] T. Nguyen and Y. Han, "A Proportional Fairness Algorithm with QoS Provision in Downlink OFDMA Systems," *IEEE Comm. Letters,* vol. 10, no. 11, pp. 760-762, Nov. 2006.

[33] W. Park and S. Bahk, "Resource Management Policies for Fixed Relays in Cellular Networks," *Proc. IEEE Global Telecomm. Conf.,* pp. 1-5, Nov. 2006.

[34] J. Jeon, K. Son, and S. Chong, "Spatial Resource Reuse in the Multi-Hop Cellular Networks: Difficulties and Benefits," *Proc. IEEE Global Telecomm. Conf.,* pp. 1-6, Nov. 2008.

[35] A. Jalali, R. Padovani, and R. Pankaj, "Data Throughput of CDMA-HDR a High Efficiency-High Data Rate Personal Communication Wireless System," *Proc. IEEE 51st Vehicular Technology Conf.,* pp. 1854-1858, May 2000.

[36] G.P. Pollhi, "Trends in Handover Design," *IEEE Comm. Magazine,* vol. 34, no. 3, pp. 82-90, Mar. 1996.

[37] D.S. Kim, J.-W. Jung, H.-G. Park, and Y.-S. Kim, "Cell Center/Edge Transmission Power Ratio Optimization to Mitigate Downlink Inter-Cell Interference in OFDM Systems," *J. Electromagnetic Waves and Applications,* vol. 22, nos. 11/12, pp. 1451-1460, Nov. 2008.

[38] H. Zeng and C. Zhu, "System-Level Modeling and Performance Evaluation of Multi-Hop 802.16j Systems," *Proc. IEEE Wireless Comm. and Mobile Computing Conf.,* pp. 354-359, Aug. 2008.

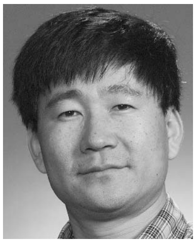
[39] Z. Zhao, X. Fang, Y. Long, X. Hu, Y. Zhao, Y. Liu, Y. Chen, H. Qu, and L. Xu, "Cost Based Local Forwarding Transmission Schemes for Two-Hop Cellular Networks," *Proc. IEEE 71st Vehicular Technology Conf.,* pp. 1-5, May 2010.

[40] H. Hu, H. Yanikomeroglu, D.D. Falconer, and S. Periyalwar, "Range Extension Without Capacity Penalty in Cellular Networks with Digital Fixed Relays," *Proc. IEEE Global Telecomm. Conf.,* pp. 3053-3057, Nov. 2004.

Yue Zhao received the BS degree in communication engineering in 2006 from the North China Institute of Science and Technology, Langfang, China, and the PhD degree in information and communication systems in 2012 from Southwest Jiaotong University, Chengdu, China. From September 2010 to September 2011, he was a visiting student in the Department of Electrical and Computer Engineering, University of Florida. He is currently the chair of IEEE VT Chengdu Student Chapter and a student member of IEICE. He served as the publication chair of PerMedia in 2010. His research interests include radio resource allocation, modeling and performance evaluation of next generation cellular networks. He is a student member of the IEEE.

Xuming Fang received the BE degree in electrical engineering, the ME degree in computer engineering, and the PhD degree in communication engineering from Southwest Jiaotong University, Chengdu, China, in 1984, 1989, and 1999, respectively. He was a faculty member in the Department of Electrical Engineering, Tongji University, Shanghai, China, in September 1984. He then joined the School of Information Science and Technology, Southwest Jiaotong University, Chengdu, where he has been a professor since February 2001 and the chair of the Department of Communication Engineering. He held visiting positions with the Technical University at Berlin, Germany, in 1998 and 1999, and with the University of Texas at Dallas, Richardson, in 2000 and 2001. His research interests include wireless broadband wireless networks, multihop networks, and broadband wireless access for high speed railway. He has to his credit around 200 high-quality research papers in journals and conference publications. He has authored or coauthored five books or textbooks. He is a member of the IEEE.

Rongsheng Huang received the BS and MS degrees in electrical engineering from Xi'an Jiaotong University, China, in 1996 and 1999, respectively, and the PhD degree from the Department of Electrical and Computer Engineering at University of Florida in 2011. From 1999 to 2001, he worked at Huawei Technologies Co., Ltd., as an R&D engineer on GPRS and 3G projects. From 2002 to 2005, he worked at UTStarcom Research Center, Shenzhen, China, as a senior engineer and team leader on a 3G project. He now works at Olympus Communication Technology of America as a member of the technical staff. His research interests include the area of media access control, protocol, and architecture for wireless networks. He is a member of the IEEE.



Yuguang Fang received the PhD degree in systems engineering from Case Western Reserve University in January 1994 and the PhD degree in electrical engineering from Boston University in May 1997. He was an assistant professor in the Department of Electrical and Computer Engineering at the New Jersey Institute of Technology from July 1998 to May 2000. He then joined the Department of Electrical and Computer Engineering at the University of Florida in May 2000 as an assistant professor, got an early promotion to an associate professor with tenure in August 2003, and to a full professor in August 2005. He held a University of Florida Research Foundation (UFRF) Professorship from 2006 to 2009, a Changjiang Scholar Chair Professorship with Xidian University, Xi'an, China, from 2008 to 2011, and a guest chair professorship with Tsinghua University, China, from 2009 to 2012. He has published more than 300 papers in refereed professional journals and conferences. He received the US National Science Foundation Faculty Early Career Award in 2001 and the US Office of Naval Research Young Investigator Award in 2002. He was the recipient of the Best Paper Award at IEEE GlobeCom in 2011 and the IEEE International Conference on Network Protocols (ICNP) in 2006, and the recipient of the IEEE TCGN Best Paper Award from the IEEE High-Speed Networks Symposium, IEEE GlobeCom 2002. He also received a 2010-2011 UF Doctoral Dissertation Advisor/Mentoring Award, 2011 Florida Blue Key/UF Homecoming Distinguished Faculty Award, and the 2009 UF College of Engineering Faculty Mentoring Award. He is also active in professional activities. He is currently serving as the editor-in-chief of the *IEEE Transactions on Vehicular Technology* (April 2013-present) and serves/served as the editor-in-chief of *IEEE Wireless Communications* (2009-2012) and on several editorial boards of technical journals including the *IEEE Transactions on Mobile Computing* (2003-2008, 2011-present), *IEEE Network* (2012-present), *IEEE Transactions on Communications* (2000-2011), *IEEE Transactions on Wireless Communications* (2002-2009), *IEEE Journal on Selected Areas in Communications* (1999-2001), *IEEE Wireless Communications Magazine* (2003-2009), and *ACM Wireless Networks* (2001-present). He served on the steering committee for the *IEEE Transactions on Mobile Computing* (2008-2010). He has actively participated in professional conference organizations such as serving as the technical program cochair for IEEE INFOCOM 2014, steering committee cochair for QShine (2004-2008), technical program vice-chair for IEEE INFOCOM 2005, technical program area chair for IEEE INFOCOM (2009-2013), technical program symposium cochair for IEEE GlobeCom 2004, and a member of the technical program committee for IEEE INFOCOM (1998, 2000, 2003-2008). He is a fellow of the IEEE and a member of the ACM.