QoS-Aware Full-Duplex Concurrent Scheduling for Millimeter Wave Wireless Backhaul Networks

WEIGUANG DING[1,2], YONG NIU[1,2], (Member, IEEE), HAO WU[1,2], (Member, IEEE),
YONG LI[3], (Senior Member, IEEE), AND ZHANGDUI ZHONG[1,2], (Senior Member, IEEE)

[1]State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China
[2]Beijing Engineering Research Center of High-Speed Railway Broadband Mobile Communications, Beijing Jiaotong University, Beijing 100044, China
[3]State Key Laboratory on Microwave and Digital Communications, Tsinghua National Laboratory for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

Corresponding author: Yong Niu (niuy11@163.com)

ABSTRACT The development of self-interference (SI) cancelation technology makes full-duplex (FD) communication possible. Considering the quality of service (QoS) of flows in small cells densely deployed scenario with limited time slot resources, this paper introduces the FD communication into the concurrent scheduling problem of millimeter-wave wireless backhaul network. We propose a QoS-aware FD concurrent scheduling algorithm to maximize the number of flows with their QoS requirements satisfied. Based on the contention graph, the algorithm makes full use of the FD condition. Both residual SI and multi-user interference are considered. Besides, it also fully considers the QoS requirements of flows and ensures the flows can be transmitted at high rates. Extensive simulations at 60 GHz demonstrate that with high SI cancelation level and appropriate contention threshold, the proposed FD algorithm can achieve superior performance in terms of the number of flows with their QoS requirements satisfied and the system throughput compared with other state-of-the-art schemes.

INDEX TERMS Millimeter-wave communications, full duplex, QoS, concurrent scheduling, backhaul networks.

I. INTRODUCTION

In the fifth generation (5G) mobile cellular network, due to the densification of small cells, the massive backhaul traffic becomes a significant problem [1], [2]. Compared with the fiber based backhaul network, the wireless backhaul network in millimeter-wave (mmWave) bands also has huge bandwidth, and can provide a more cost-effective and flexible solution to this problem [3]. In the mmWave wireless backhaul network, directional antennas and beamforming techniques are often used to compensate for the high path loss [4], [5]. The directional communication can reduce the interference between different flows, and thus concurrent transmissions (i.e. spatial reuse) of flows become possible. Concurrent transmissions can significantly increase the system throughput [6]. However, the concurrent transmissions of multiple flows result in higher mutual interference, which will conversely degrade the system performance. Therefore, how to efficiently schedule the flows transmitted concurrently is worth to study and thus has attracted considerable interest from researchers [2], [7]–[10].

Most existing concurrent scheduling schemes [2], [7]–[10] in mmWave bands hold the assumption of half-duplex (HD). Recently, with the development of self interference (SI) cancelation technology [11]–[15], it becomes possible to enable the full-duplex (FD) communication in mmWave wireless backhaul networks [16]. Here, the SI means the transmitted signal received by the local receiver at the same base station (BS) [15], which is shown in Figure 1. It seriously affects the performance of FD system [17]. By transmitting and receiving information simultaneously at the same BS over the same frequency [11], [17], the FD communication may theoretically double the spectral efficiency [18], which brings an important opportunity for the concurrent scheduling problem in mmWave wireless backhaul networks.
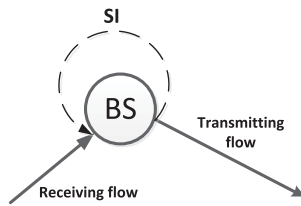
FIGURE 1. Self interference received at the same BS.

However, the SI can't be completely eliminated in practice. There is still residual self interference (RSI) in the system. Therefore, for the FD backhaul system, the interference we need to consider is more complex than that in HD system: not only multi-user interference (MUI), but also RSI. This is a big challenge for the concurrent scheduling problem in mmWave backhaul networks.

Moreover, in the future 5G mmWave backhaul network, many applications are bandwidth-intensive (e.g. uncompressed video streaming), and should be provided with multi-Gbps throughput [8]. The data flows of these applications all have their own minimum throughput requirements. In the following paper, the minimum throughput requirements will be referred to as the quality of service (QoS) requirements. To guarantee the required quality of service, the QoS requirements of flows need to be satisfied [2]. Although in [16], Feng *et al.* introduced the FD communication into the scheduling scheme for 5G mmWave backhaul networks, the scheduling solution was designed for the case with sufficient time slot (TS) resources. The QoS requirements were not specially considered in [16]. Therefore, for the case where the TS resources are limited compared with the intensive traffic demands of users [2], [8], how to satisfy the QoS requirements of flows as many as possible is still a challenge.

The above opportunities and challenges motivate us to investigate a QoS-aware FD concurrent scheduling scheme for the mmWave wireless backhaul network with limited TS resources. The contributions of this paper can be summarized as follows.

- We innovatively introduce the FD technology into the concurrent scheduling problem of mmWave wireless backhaul networks with limited number of TSs. Both RSI and MUI are simultaneously taken into account so that the advantages of the FD technology and the concurrent transmission can be brought into full play.
- The QoS requirements of flows in the case where the TS resources are limited are specially considered. We formulate a nonlinear integer programming (NLIP) problem aiming at maximizing the number of flows with their QoS requirements satisfied. Then, a QoS-aware FD scheduling algorithm is proposed, which can keep the flow rate high and satisfy the QoS requirements of flows as many as possible.
- We evaluate the proposed algorithm in the 60GHz mmWave wireless backhaul network with limited TS resources. The extensive simulations demonstrate

that compared with other state-of-the-art algorithms, the proposed QoS-aware FD algorithm can significantly improve the number of flows with their QoS requirements satisfied and the total system throughput. Furthermore, we also analyze the impact of SI cancelation level and contention threshold on the performance improvement.

The structure of this paper is organized as follows. Section II introduces the related work. Section III introduces the system overview and assumption. In Section IV, the optimal concurrent scheduling problem in FD mmWave wireless backhaul networks with limited TSs is formulated into an NLIP. In Section V, a QoS-aware FD concurrent scheduling algorithm is proposed. In Section VI, we conduct extensive simulations, and in Section VII we conclude this paper.

II. RELATED WORK

Compared with the serial TDMA scheme, the concurrent transmission scheduling can significantly increase the system throughput, and thus has been extensively studied [2], [7]–[10]. Cai *et al.* [7] proposed a scheduling algorithm based on exclusive region to support concurrent transmissions. To maximize the number of flows scheduled in the network so that the QoS requirement of each flow is satisfied, Qiao *et al.* [8] proposed a flip-based scheduling algorithm. In [2], Zhu *et al.* proposed a Maximum QoS aware Independent Set (MQIS) based scheduling algorithm for mmWave backhaul networks to maximize the number of flows with their QoS requirements satisfied. In MQIS, the concurrent transmission and the QoS aware priority are exploited to achieve more successfully scheduled flows and higher network throughput. In [9], based on Stackelberg game, Li *et al* proposed a distributed transmission power control solution for the concurrent transmission scheduling between interference D2D links to further enhance the network throughput. Niu *et al.* [10] proposed an energy efficient scheduling scheme for the mmWave backhaul network, which exploits concurrent transmissions to achieve higher energy efficiency. However, all the above scheduling algorithms assume the devices are HD.

Recently, the development of SI cancelation technology has made FD communication possible. Jain *et al.* [11] proposed the signal inversion and adaptive cancelation. Combining signal inversion cancelation with digital cancelation can reduce SI by up to 73dB. Everett *et al.* [12] showed the BS could exploit directional diversity by using directional antennas to achieve additional passive suppression of the SI. Besides, Miura and Bandai [13] proposed a novel node architecture introducing directional antennas into FD wireless technology. Rajagopal *et al.* [14] proved enabling backhaul transmission on one panel while simultaneously receiving backhaul on an adjacent panel is attainable for next generation backhaul designs. In [15], Xiao *et al.* showed the configuration with separate Tx/Rx antenna arrays appeared more flexible in SI suppression, and proposed the beamforming cancelation in FD mmWave communication.

Considering the potential of the FD communication in increasing network performance, Feng et al. [16] proposed a design framework for 5G mmWave backhaul, which combined FD transmissions and hybrid beamforming with routing and scheduling schemes. However, the scheduling solution in [16] was for the system with sufficient TS resources and aimed at accomplishing all of the transmissions with the minimum time. Thus, there was no special consideration for the QoS requirements of flows in limited time. Therefore, for mmWave backhaul networks with limited TS resources, a more QoS-favorable FD scheduling algorithm is needed.

III. SYSTEM OVERVIEW AND ASSUMPTION

In this paper, we consider a typical FD mmWave wireless backhaul network in the small cells densely deployed scenario. As shown in Figure 2, the network includes N BSs. The BSs are connected through backhaul links in the mmWave band. When there are some traffic demands from one BS to another, we say there is a flow between them. As shown in Figure 3, each BS operates in FD mode and is equipped with two steerable directional antennas: one for transmitting
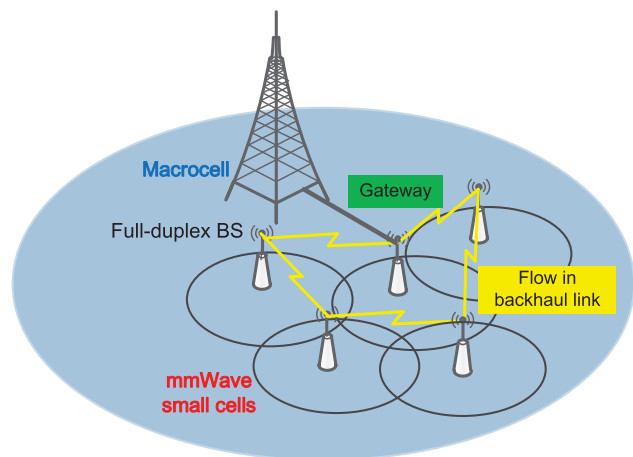

FIGURE 2. A full-duplex mmWave wireless backhaul network in the small cells densely deployed scenario.
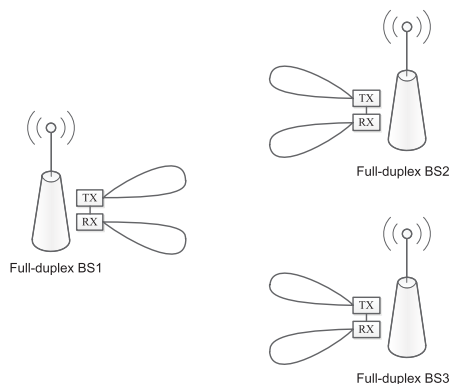

FIGURE 3. Full-duplex base stations.

and another for receiving. Therefore, a BS can at most simultaneously support two flows. It can simultaneously serve as the transmitter of one flow and the receiver of another, but it can't simultaneously serve as the transmitters or receivers of both two flows. There are one or more BSs connected to the backbone network via the macrocell, which is (are) called gateway(s) [10]. A backhaul network controller (BNC) resides on one of the gateways, which can synchronize the network, receive the QoS requirements of flows and obtain the locations of BSs [19].

A. THE RECEIVED POWER

Since non-line-of-sight (NLOS) transmissions suffer from high attenuation, we use the line-of-sight (LOS) path loss model for mmWave as described in [2]. For flow f, the received signal power at its receiver r_f from its transmitter t_f can be expressed as

$$P_r\left(t_f, r_f\right) = kP_tG_t(t_f, r_f)G_r(t_f, r_f)d_{t_f r_f}^{-n}. \quad (1)$$

k is a factor that is proportional to $\left(\frac{\lambda}{4\pi}\right)^2$, where λ denotes the wave length; P_t denotes the transmission power of the transmitter; $G_t\left(t_f, r_f\right)$ denotes the transmitter antenna gain in the direction of from t_f to r_f, and $G_r\left(t_f, r_f\right)$ denotes the receiver antenna gain in the direction of from t_f to r_f; $d_{t_f r_f}$ denotes the distance between t_f and r_f, and n is the path loss exponent [8].

According to the FD assumption mentioned above, the two flows scheduled simultaneously either have no common node or one's transmitter is the receiver of the other. Therefore, the interference between different flows can be divided into two cases: 1) the interference between two flows without any common node, namely, MUI; 2) the RSI after SI cancelation. The MUI caused by the transmitter t_l of flow l to the receiver r_f of flow f is defined as

$$P_r\left(t_l, r_f\right) = \rho kP_tG_t(t_l, r_f)G_r(t_l, r_f)d_{t_l r_f}^{-n}, \quad (2)$$

where ρ is the MUI factor between different flows, which is related to the cross correlation of signals from different flows [2]. According to [11], after SI cancelation, the effect of RSI can be modeled in terms of the SNR loss. Therefore, we can use $\beta_n N_0 W$ to denote the RSI, where the non-negative parameter β_n represents the SI cancelation level of the nth BS. The smaller β_n, the higher the level of SI cancelation. Due to various factors, we assume the parameters for different BSs are different. N_0 is the onesided power spectral density of white Gaussian noise; W is the channel bandwidth.

B. DATA RATE

With the reduction of multipath effect, mmWave channel can be approximated as Gaussian channel. With the interference from other flows, the data rate of flow f can be estimated according to the Shannon's channel capacity [10].

C. ANTENNA MODEL

In this paper, we adopt the realistic antenna model in [10]. The gain of a directional antenna in units of dB can be
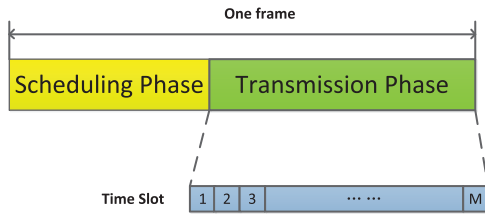
FIGURE 4. The structure of one frame.

expressed as

$$G(\theta) = \begin{cases} G_0 - 3.01 \times \left(\frac{2\theta}{\theta_{-3\text{dB}}}\right)^2, & 0° \le \theta \le \theta_{ml}/2 \\ G_{sl}. & \theta_{ml}/2 < \theta \le 180° \end{cases} \quad (3)$$

θ denotes an angle within the range $[0°, 180°]$. The maximum antenna gain G_0 can be calculated as $G_0 = 10\log(1.6162/\sin(\theta_{-3\text{dB}}/2))^2$. $\theta_{-3\text{dB}}$ is the angle of the half-power beamwidth. The main lobe width θ_{ml} in units of degrees can be calculated as $\theta_{ml} = 2.6 \times \theta_{-3\text{dB}}$. The sidelobe gain $G_{sl} = -0.4111 \times \ln(\theta_{-3\text{dB}}) - 10.579$ [10].

IV. PROBLEM FORMULATION

In this paper, we consider a QoS-aware FD concurrent scheduling problem when the time is limited. System time is divided into a series of non-overlapping frames. As shown in Figure 4, each frame consists of a scheduling phase, where a transmission schedule **S** is computed by the BNC, and a transmission phase, where the BSs start concurrent transmissions following the schedule [6]. The transmission phase is further divided into M equal TSs. It's assumed that there are F flows in the network and each flow f has its QoS requirement q_f. For each flow f, we define a binary variable a_f^i to indicate whether flow f is scheduled in the ith TS. If so, $a_f^i = 1$; otherwise, $a_f^i = 0$. Since there may be different flows to be transmitted in different TSs, we denote the actual transmission rate of flow f in the ith TS by R_f^i. According to the Shannon's channel capacity [10], R_f^i can be calculated as

$$R_f^i = \eta W \log_2(1 + \frac{a_f^i P_r(t_f, r_f)}{N_0 W + \sum_h a_h^i \beta_{t_h} N_0 W + \sum_l a_l^i P_r(t_l, r_f)}). \quad (4)$$

where η is the factor that describes the efficiency of the transceiver design, which is in the range of $(0, 1)$. W is the bandwidth, and N_0 is the one-sided power spectra density of white Gaussian noise. h denotes the flow whose transmitter is the same as the receiver of flow f. β_{t_h} is the parameter of SI cancelation level at BS t_h. l denotes the flow without any common node with f.

Then we can define the actual throughput of flow f based on the schedule **S** as

$$T_f = \frac{\sum_{i=1}^{M} R_f^i \triangle t}{T_s + M \triangle t}, \quad (5)$$

where T_s is the time of scheduling phase and $\triangle t$ is the time of one TS. When the actual throughput T_f of flow f is greater than its QoS requirement q_f, we say flow f has satisfied its QoS requirement, and call the flow a completed flow. A binary variable I_f is used to indicate whether flow f is completed. $I_f = 1$ indicates f is completed, while $I_f = 0$ indicates f is not completed.

As we investigate a QoS-aware scheduling for a backhaul network with limited time, given the QoS requirements of flows, with the limited number of TSs in the transmission phase, the optimal schedule should accommodate as many flows as possible [2]. In other words, we should aim at maximizing the number of flows that satisfy their QoS requirements (i.e. the number of flows that are completed). Therefore, the objective function can be formulated as

$$\max \sum_{f=1}^{F} I_f, \quad (6)$$

and the first constraint is

$$I_f = \begin{cases} 1, & T_f \ge q_f; \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

Next, we analyze the other constraints. Firstly, we use variable f_n to denote the flow whose transmitter or receiver is the nth BS B_n, such as the transmitting flow and the receiving flow in Figure 1; thus $a_{f_n}^i$ indicates whether flow f_n is scheduled in the ith TS, that is, whether f_n does use B_n in the ith TS. According to our FD assumption described in section III, because each BS is just equipped with two steerable directional antennas, the number of flows that simultaneously use the same BS B_n can't exceed two; this constraint can be expressed as

$$\sum_{f_n} a_{f_n}^i \le 2, \quad \forall i, n. \quad (8)$$

Then we use f_n^1 and f_n^2 stand for the two flows that simultaneously use B_n; we also use $T(B_n)$ and $R(B_n)$ stand for the wireless links with B_n as the transmitter and the receiver, respectively. As assumed in section III, for the two antennas of a FD BS, one of them is a transmitting antenna and the other is a receiving antenna. Therefore, when two flows simultaneously use the same BS, the BS can only serve as the transmitter for one flow and as the receiver for the other, which can be expressed as:

$$\begin{aligned} &f_n^1 \in T(B_n) \& f_n^2 \in R(B_n) \\ &or\ f_n^1 \in R(B_n) \& f_n^2 \in T(B_n),\ \text{if} \sum_{f_n} a_{f_n}^i = 2. \end{aligned} \quad (9)$$

In summary, the problem of optimal scheduling (**P1**) can be formulated as follows:

$$\max \sum_{f=1}^{F} I_f$$

s.t. Constraints (7) - (9)

This is a nonlinear integer programming (NLIP) problem and is NP-hard [2]. The optimization problem is similar to that in [2]. Reference [2] is for the HD scenario while ours is for the FD scenario. Compared with [2], the number of constraints for our optimization problem is more, and the problem is obviously more complex than that in [2]. Reference [2] is NP-hard, and thus our optimization problem is also NP-hard. In each TS, every flow is either scheduled or unscheduled. Therefore, when the number of TSs is M and the number of flows is F, the computational complexity using exhaustive search algorithm is 2^{MF}, which is exponential. In the small cells densely deployed scenario, the number of flows may be large, and thus it will be time-consuming if we use exhaustive algorithm to solve **P1**. The computational time is unacceptable for practical mmWave small cells where the duration of one TS is only a few microseconds [20]. Consequently, a heuristic algorithm with low complexity is desired to solve it in practice.

V. QOS-AWARE FULL-DUPLEX SCHEDULING ALGORITHM
In this section, we propose a QoS-aware full-duplex concurrent scheduling algorithm for problem **P1**. Borrowing the idea of contention graph from [10], the algorithm makes full use of the FD condition and satisfies the QoS requirements of flows as many as possible. Next, we first describe how to construct the contention graph and then describe the proposed algorithm in detail.

A. THE CONSTRUCTION OF CONTENTION GRAPH
In FD mmWave wireless backhaul networks, not all pairs of flows can be concurrently scheduled. In contention graph [10], when the two flows can't be concurrently scheduled, we define there is a contention between them. In this paper, based on the assumption and analysis mentioned above, we define the flows that can't be concurrently scheduled into the following two cases.

Firstly, according to the FD assumption described in section III, for the two antennas of a FD BS, one of them is a transmitting antenna and the other is a receiving antenna. Therefore, the two flows that simultaneously use the same BS as their transmitters (or receivers) can't be concurrently scheduled. This case is shown in Figure 5. Figure 5 (a) shows that two flows simultaneously use the same BS as their transmitters. Similarly, Figure 5 (b) shows that two flows simultaneously use the same BS as their receivers. Accordingly, based on the analysis for this case, the flows
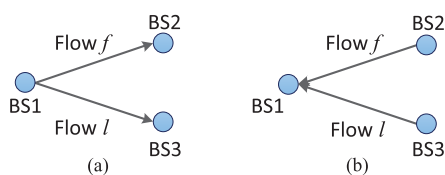

FIGURE 5. The flows that can't be concurrently scheduled due to the full-duplex assumption.
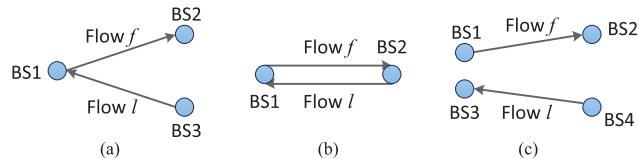

FIGURE 6. The flows that are allowed concurrently scheduled under the full-duplex assumption.

that can be concurrently scheduled are divided into following three cases. 1) As shown in Figure 6 (a), the transmitter of flow f is the receiver of flow l, but the receiver of flow f is not the transmitter of flow l. 2) As shown in Figure 6 (b), the transmitter of flow f is the receiver of flow l, and the receiver of flow f is the transmitter of flow l. 3) As shown in Figure 6 (c), the transmitter of flow f is not the receiver of flow l, and the receiver of flow f is not the transmitter of flow l, either.

Secondly, considering the QoS requirements of flows, to guarantee the flow rate and the system throughput, the two flows whose relative interference (RI) [2] between each other is large can't be concurrently scheduled. When the RI between two flows is large, the rates of the flows become low. The low rates result in inefficient resource utilization. In other words, the TS resources are allocated to the flows, but the QoS requirements of them are hard to be satisfied, and thus they can't support the specific applications [8]. For the three cases in Figure 6, we now define their RI, respectively. 1) For the case in Figure 6 (a), the interference from flow f to flow l is RSI. Therefore, the RI from flow f to flow l can be defined as

$$RI_{f,l} = \frac{N_0 W + \beta_{t_f} N_0 W}{P_r(t_l, r_l)}, \qquad (10)$$

where $P_r(t_l, r_l)$ is calculated as (1). The interference from flow l to flow f is MUI; so the RI from flow l to flow f is defined as

$$RI_{l,f} = \frac{N_0 W + P_r(t_l, r_f)}{P_r(t_f, r_f)}, \qquad (11)$$

where $P_r(t_l, r_f)$ is calculated as (2) and $P_r(t_f, r_f)$ is calculated as (1). 2) For the case in Figure 6 (b), both the interference from flow f to l and that from flow l to f is RSI. Therefore, the RI between the two flows are both similar to (10). 3) For the case in Figure 6 (c), both the interference from flow f to l and that from flow l to f is MUI. Therefore, the RI between the two flows are both similar to (11).

Next, let's construct the contention graph. In the contention graph, each vertex represents a flow. If two flows can't be concurrently scheduled (i.e., there is a contention between them), an edge is inserted between the two corresponding vertices. For example, as shown in Figure 7, there is a contention between flow 1 and flow 2. In contrast, there is no contention between flow 1 and flow 3. Specifically, for the two pairs of flows in Figure 5, there is an edge between the two corresponding vertices, respectively. In addition, for
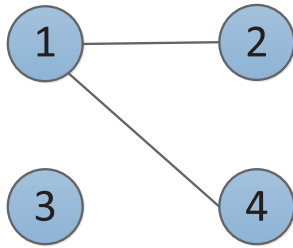
FIGURE 7. The contention graph with contention between flow 1 and flow 2 and contention between flow 1 and flow 4.

the three pairs of flows in Figure 6, we should examine whether the RI between the flows is too large. When the RI between two flows is larger than a contention threshold σ, we say there is a contention between them. In other words, if $\max(RI_{f,l}, RI_{l,f}) > \sigma$, an edge is inserted into the two corresponding vertices.

B. THE QoS-AWARE FULL-DUPLEX SCHEDULING ALGORITHM

Based on the contention graph, we now concretely describe the proposed algorithm. The pseudo code for it is shown in Algorithm 1.

Algorithm 1 The Full-Duplex Scheduling Algorithm

---

1 BNC obtains Loc, β_n and q_f;
2 calculate ξ_f for each flow;
3 remove $\mathbb{D} = \{f | \xi_f > M\}$;
4 sort the remaining F flows in non-decreasing order
   according to ξ_f and get the pre-scheduling set $\mathbb{P}$;
5 generate **G** for all flows in $\mathbb{P}$;
6 **Initialization:** $\mathbf{S}_{F \times M} = \mathbf{0}$ and *change* = 1;
7 **for** *slot i* $(1 \le i \le M)$ **do**
8   **if** *change* = 1 **then**
9   **for** *flow f* $(1 \le f \le F)$ **do**
10   **if** $\mathbf{S}_i(f) = 0$ *and f has no contention with*
            the flow(s) that is(are) ongoing **then**
11   **if** *set* $\mathbf{S}_i(f) = 1$ *could increase the*
               system throughput **then**
12                $\mathbf{S}_i(f) = 1$;
13   *change* = 0; $\mathbf{S}_{i+1} = \mathbf{S}_i$;
14   **if** *any* $T_f > q_f$ **then**
15   *change* = 1;
16       $\mathbf{S}_i(f) = -1$;

To begin with, line 1 is some preparation work. The BNC obtains the BS location (Loc), the SI cancelation level (β_n) at each BS and the QoS requirement (q_f) of each flow. Next, in line 2, we calculate the number of TSs that each flow spends to complete its QoS requirement when there is no interference from others. The number of TSs is calculated as

$$\xi_f = \frac{q_f * (T_s + M \triangle t)}{R_f * \triangle t}. \tag{12}$$

R_f is the rate of flow f without interference from others, which can be calculated as

$$R_f = \eta W log_2(1 + \frac{P_r(t_f, r_f)}{N_0 W}). \tag{13}$$

Since the scheduling problem we investigate is in limited time, i.e., in M TSs, the flow whose ξ_f is greater than M should be removed. In the actual scheduling, there exists interference from other flows; so the flow rates will be further reduced and the spent number of TS will be further increased. The judgment for these flows becomes meaningless to our optimization goal. Removing the flows (represented by set $\mathbb{D}$) can not only reduce the complexity of subsequent scheduling, but also save more TSs to schedule more worthwhile flows, that is, the flows that can be completed in M TSs. Pseudo code for this step is shown in line 3.

Then, as shown in lines 4-5, we sort the remaining F flows in non-decreasing order according to ξ_f and call the set of the sorted flows "pre-scheduling set" $\mathbb{P}$. Next, we construct the contention graph **G** for all the flows in $\mathbb{P}$. A $F \times M$ matrix $\mathbf{S} = [\mathbf{S}_1, \mathbf{S}_2, \ldots \mathbf{S}_i, \ldots \mathbf{S}_M]$ denotes the scheduling decision in M slots. Vector $\mathbf{S}_i$ indicates which flows are scheduled in the ith TS. If flow f is scheduled in TS i, $\mathbf{S}_i(f) = 1$; if not, $\mathbf{S}_i(f) = 0$; if f is completed, $\mathbf{S}_i(f) = -1$. Flag variable *change* indicates whether there is(are) some flow(s) newly completed in every TS. If any, *change* = 1; if not, *change* = 0. For the first slot, *change* is initialized to 1. The initialization steps are shown in line 6.

Then we make the scheduling decision slot by slot. In lines 9-12, to complete more flows in the limited TSs, we first determine the flow with the smallest ξ. In other words, we determine the flows in $\mathbb{P}$ one by one from the beginning. If flow f has never been scheduled and has no contention with the flow(s) that is(are) ongoing, then the profit of scheduling the flow is evaluated: if scheduling it can increase the total system throughput, we schedule it; otherwise, skip the flow and determine the next. These rules help to guarantee the flow rate and the system throughput; so it's more QoS-aware.

In every TS, as shown in lines 14-16, it is necessary to check whether some flow(s) has(have) completed its(their) QoS requirement. If so, the corresponding $\mathbf{S}_i(f)$ is(are) set to -1, which means the flow(s) will never be scheduled later. When one flow is completed, allocating resources to it is of little significance to further improve its QoS. Therefore, we should stop scheduling it and save the TSs to serve more other flows. At the same time, *change* is set to 1.

In fact, as shown in line 8, only when it's the 1th TS or some flow(s) is(are) newly completed, that is, when *change* = 1, we need to make new scheduling decision. If *change* = 0, the scheduling vector is the same as the previous TS, which is shown in line 13. In this way, the scheduling complexity is greatly reduced. The algorithm is repeated until M TSs are over, and we finally obtain the scheduling vector for each TS. Obviously, in the worst case, the variable *change* is 1 in every TS, that is, there is(are) some flow(s) to be newly completed

in every TS. Therefore, for the M TSs and F flows, the worst computational complexity of Algorithm 1 is $O(MF)$.

VI. PERFORMANCE EVALUATION

A. SIMULATION SETUP

In the simulations, we evaluate the performance of the proposed algorithm in a 60GHz mmWave wireless backhaul network that 10 BSs are uniformly distributed in a $100m \times 100m$ square area. Every BS has the same transmission power P_t. The transmitters and receivers of flows are randomly selected, and the QoS requirements of flows are uniformly distributed between 1Gbps and 3Gbps. The SI cancelation parameters β for different BSs are uniformly distributed in a certain range. To be more realistic, other parameters are shown in Table 1.

TABLE 1. Simulation parameters.

Parameter	Symbol	Value
Transmission power	P_t	1000mW
Path loss exponent	n	2
MUI factor	ρ	1
Transceiver efficiency factor	η	0.5
System bandwidth	W	1200MHz
Background noise	N_0	-134dbm/MHz
Slot time	$\triangle t$	18us
Scheduling phase time	T_s	850us
Number of slots in transmission phase	M	2000
Half-power beamwidth	θ_{-3dB}	30°

Because we focus on the QoS of flows, according to our optimization goal, we use the number of completed flows and the system throughput as evaluation metrics. When one flow achieves its QoS requirement, it is called a completed flow. System throughput represents the throughput of all flows in the network per slot.

To show the advantages of the proposed QoS-aware FD concurrent scheduling algorithm (Proposed-FD) in the network system with limited TS resources, we compare it with the following four schemes.

1) *TDMA:* In TDMA, the flows are transmitted serially. We use TDMA as the baseline for evaluating performance without concurrent transmissions.

2) *MQIS:* MQIS [2] is a HD concurrent scheduling algorithm based on the maximum QoS-aware independent set. It first schedules the flow with the smallest degree in contention graph. It doesn't remove the flow(s) spent too much slots, nor does it evaluate the profits when adding a new flow. To the best of our knowledge, MQIS achieves the best performance in the network system with limited TS resources in terms of the number of completed flows and system throughput among the existing scheduling algorithms. Therefore, we use it as the baseline for evaluating performance without FD communication.

3) *Proposed-HD:* It uses the same scheduling algorithm with Proposed-FD, but it only allows the HD communication. We also use it as a baseline without the FD communication.

4) *FDP:* Full-Duplex (FDP) scheme [16] is for the system where the TS resources are sufficient. It aims at accomplishing all of the transmissions with the minimum time. In every

phase, higher priority is given to the flow that occupies the most TSs. If another flow is qualified to be transmitted together in the current phase, i.e., the number of flows simultaneously using the same BS doesn't exceed the number of RF chains and the SINR is larger than a certain threshold, the corresponding flow is also scheduled. Only when all the flows scheduled together in one phase are completed, can we start the next phase and make a new scheduling decision. We use it as a baseline for FD communication.

Each simulation performs 100 times to get a more reliable average result.

B. SIMULATION RESULTS

1) UNDER DIFFERENT NUMBERS OF FLOWS

In this case, the contention threshold σ is set to 0.001, and the SI cancelation level parameter β is uniformly distributed between $2 - 4$. The simulation results are shown in Figure 8. We can find that the Proposed-FD algorithm always shows superior performance compared with other algorithms, and the more the number of flows, the more obvious the advantages of the Proposed-FD. Compared with the HD algorithms, the Proposed-FD allows simultaneous transmission and reception at the same BS. In fact, the Proposed-HD algorithm also performs better than MQIS.
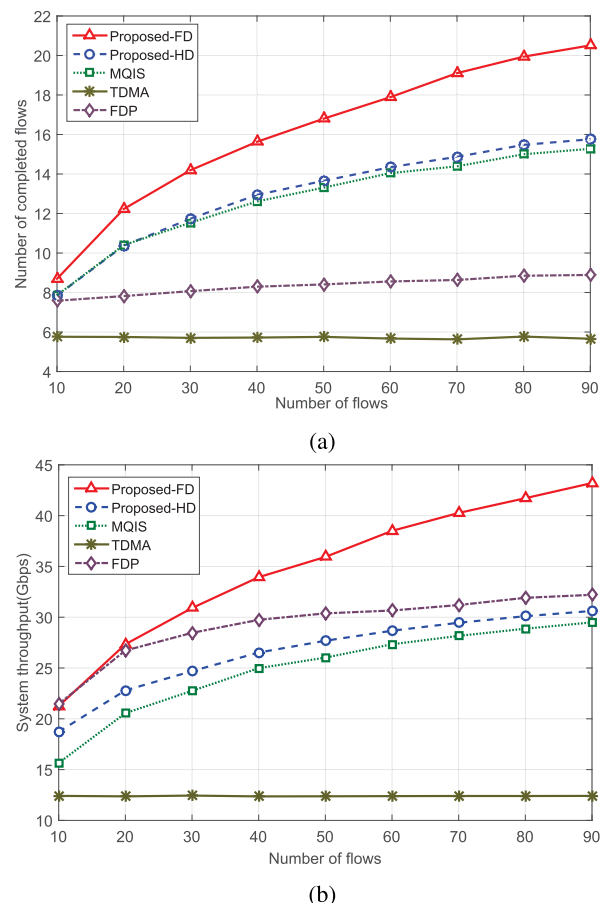

(a)

(b)

FIGURE 8. The number of completed flows and system throughput under different number of flows. (a) number of completed flows. (b) system throughput.
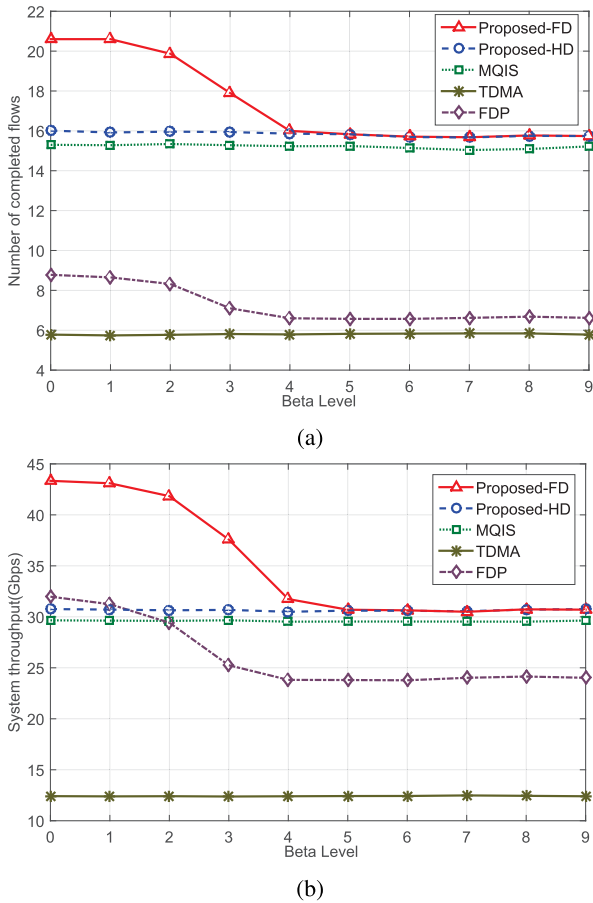
FIGURE 9. The number of completed flows and system throughput under different SI cacellation levels. (a) number of completed flows. (b) system throughput.

FIGURE 10. The number of completed flows and system throughput under different contention thresholds. (a) number of completed flows. (b) system throughput.

This is because when deciding whether or not to schedule a flow, we consider whether adding the flow can improve the system throughput. This makes each flow be scheduled at a higher rate, and thus the QoS requirements of flows can be achieved more quickly. As for the FDP algorithm, the number of completed flows for it is not large enough. This is mainly because the issue they investigate and the optimization goal are different from ours. It is for the network where the TS resources are sufficient and aims at accomplishing all the transmissions with the minimum time. Therefore, it is not suitable for the investigated problem in this paper that maximizing the completed flows in limited time. Moreover, with the increase of the total number of flows, the number of completed flows for FDP doesn't increase significantly. This is because in FDP, only when all the flows scheduled together in one phase are completed, can we start the next phase and make a new scheduling decision. As a result, when some flows are completed quickly, due to the lack of scheduling of new flows, a large amount of TS resources are wasted. Thus, in limited time, the number of completed flows has almost no change. However, the system throughput of FDP is higher than other HD algorithms. This is because FDP
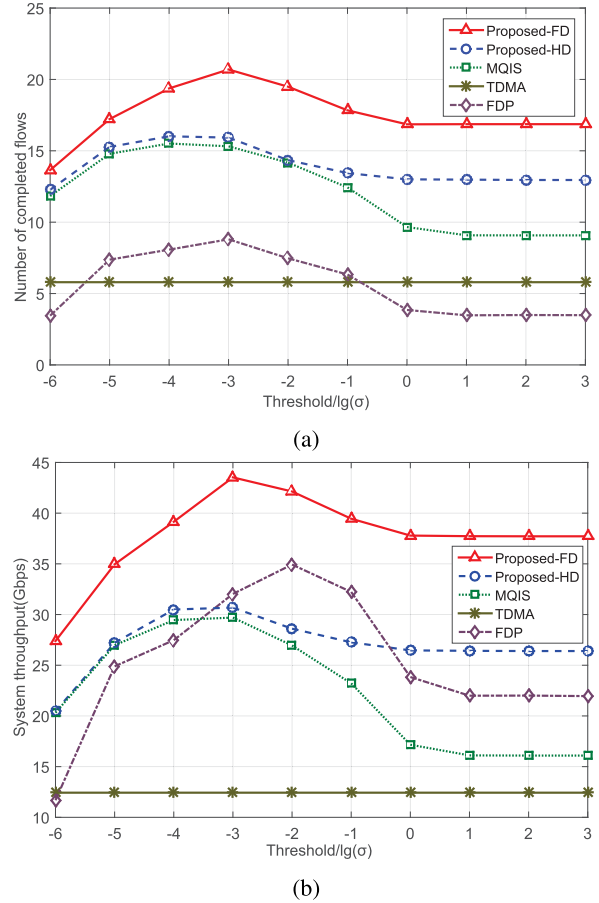
prefers the flows that occupy more TSs. These flows usually have higher QoS requirements (i.e. the minimum throughput requirement), so even the number of completed flows is small, the system throughput is still high. In particular, when the number of flows is 90, the Proposed-FD improves the number of completed flows by 30.1% compared with Proposed-HD and improves the system throughput by 34.1% compared with FDP.

2) UNDER DIFFERENT SI CANCELATION LEVELS

For the two FD algorithms (the Proposed-FD and FDP), the SI cancelation level β has an obvious impact on the performance. Thus, we simulate the performance under different magnitudes of β, as shown in Figure 9. The abscissa x is the magnitude of β. For example, when x = 2, β is uniformly distributed in $2 \times 10^2 - 4 \times 10^2$. In this case, the total number of flows is 90, and $\sigma = 0.001$. We can find that the performance of the Proposed-FD is better when β is smaller, that is, when the SI cancelation level is higher. As β becomes larger, the performance of the Proposed-FD gradually deteriorates. In particular, when β reaches 10^4 magnitude, the Proposed-FD has the same performance as the Proposed-HD. This tells us that not in any case can the FD communication improve the

system performance, and better SI cancelation techniques are needed. The trend of the performance for FDP is similar to the Proposed-FD. However, due to the applicable scenario and optimization goal are different from ours, the performance of FDP is relatively poor.

3) UNDER DIFFERENT CONTENTION THRESHOLDS

To study the impact of contention threshold σ on the performance, we simulate the two metrics under different σ, as shown in Figure 10. The abscissa x represents the magnitude of σ. For example, x $= -3$ means $\sigma = 10^{(-3)}$. In this case, the number of flows is 90, and β is uniformly distributed between $2-4$. We can observe that as σ increases, in addition to TDMA, the performance of the other four solutions first increases, then degrades and finally almost keep unchanged. This is because when σ is small, it is not conducive to concurrent transmissions. When σ is greater than a certain threshold (e.g., $10^{(-3)}$ for Proposed-FD), there is severe interference between concurrent flows, which leads to the rate reduction and is harmful to satisfy the QoS requirements. Therefore, to achieve the best performance, we should choose the appropriate threshold. Under the simulation conditions in this paper, we choose $\sigma = 10^{(-3)}$. Specifically, when $\sigma = 10^{(-3)}$, the Proposed-FD improves the number of completed flows by 29.9% compared with Proposed-HD and improves the system throughput by 35.9% compared with FDP. Although FDP doesn't use the contention graph, we convert its SINR threshold into the contention threshold, so its performance also varies with σ.

VII. CONCLUSION

In this paper, we propose a QoS-aware full-duplex concurrent scheduling algorithm for mmWave wireless backhaul networks. Considering the FD characteristics and the QoS requirements of flows in the system with limited TS resources, the proposed algorithm exploit the contention graph to find the concurrently scheduled flows and maximize the number of completed flows. Extensive simulations show that the proposed FD scheduling algorithm can significantly increase the number of completed flows and the system throughput compared with other scheduling schemes. In addition, the effects of SI cancelation level and contention threshold on the performance are also simulated to guide a better scheduling.

In the future work, we will also consider the blockage problem in mmWave communications into problem, and propose a robust scheme for the mmWave full-duplex backhaul network.

REFERENCES

[1] R. Baldemair *et al.*, "Ultra-dense networks in millimeter-wave frequencies," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 202–208, Jan. 2015.
[2] Y. Zhu, Y. Niu, J. Li, D. O. Wu, Y. Li, and D. Jin, "QoS-aware scheduling for small cell millimeter wave mesh backhaul," in *Proc. IEEE ICC*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.
[3] M. N. Islam, A. Sampath, A. Maharshi, O. Koymen, and N. B. Mandayam, "Wireless backhaul node placement for small cell networks," in *Proc. 48th Annu. Conf. Inf. Sci. Syst. (CISS)*, Princeton, NJ, USA, Mar. 2014, pp. 1–6.
[4] Y. M. Tsang, A. S. Y. Poon, and S. Addepalli, "Coding the beams: Improving beamforming training in mmWave communication system," in *Proc. IEEE Global Telecommun. Conf.*, Houston, TX, USA, Dec. 2011, pp. 1–6.
[5] Z. Xiao, T. He, P. Xia, and X.-G. Xia, "Hierarchical codebook design for beamforming training in millimeter-wave communication," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3380–3392, May 2016.
[6] I. K. Son, S. Mao, M. X. Gong, and Y. Li, "On frame-based scheduling for directional mmWave WPANs," in *Proc. IEEE INFOCOM*, Orlando, FL, USA, Mar. 2012, pp. 2149–2157.
[7] L. X. Cai, L. Cai, X. Shen, and J. W. Mark, "REX: A randomized exclusive region based scheduling scheme for mmWave WPANs with directional antenna," *IEEE Trans. Wireless Commun.*, vol. 9, no. 1, pp. 113–121, Jan. 2010.
[8] J. Qiao, L. X. Cai, X. Shen, and J. W. Mark, "STDMA-based scheduling algorithm for concurrent transmissions in directional millimeter wave networks," in *Proc. IEEE ICC*, Ottawa, ON, Canada, Jun. 2012, pp. 5221–5225.
[9] Y. Li, Z. Zhang, W. Wang, and H. Wang, "Concurrent transmission based Stackelberg game for D2D communications in mmWave networks," in *Proc. IEEE ICC*, Paris, France, May 2017, pp. 1–6.
[10] Y. Niu *et al.*, "Energy-efficient scheduling for mmWave backhauling of small cells in heterogeneous cellular networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2674–2687, Mar. 2017.
[11] M. Jain *et al.*, "Practical, real-time, full duplex wireless," in *Proc. MobiCom*, Las Vegas, NV, USA, Sep. 2011, pp. 301–312.
[12] E. Everett, M. Duarte, C. Dick, and A. Sabharwal, "Empowering full-duplex wireless communication by exploiting directional diversity," in *Proc. Conf. Rec. 45th Conf. Signals, Syst. Comput. (ASILOMAR)*, Pacific Grove, CA, USA, Nov. 2011, pp. 2002–2006.
[13] K. Miura and M. Bandai, "Node architecture and MAC protocol for full duplex wireless and directional antennas," in *Proc. IEEE PIMRC*, Sydney, NSW, Australia, Sep. 2012, pp. 369–374.
[14] S. Rajagopal, R. Taori, and S. Abu-Surra, "Self-interference mitigation for in-band mmWave wireless backhaul," in *Proc. CCNC*, Las Vegas, NV, USA, Jan. 2014, pp. 551–556.
[15] Z. Xiao, P. Xia, and X.-G. Xia, "Full-duplex millimeter-wave communication," *IEEE Wireless Commun.*, vol. 24, no. 6, pp. 136–143, Dec. 2017.
[16] W. Feng, Y. Li, D. Jin, L. Su, and S. Chen, "Millimetre-wave backhaul for 5G networks: Challenges and solutions," *Sensors*, vol. 16, no. 6, p. 892, Jun. 2016.
[17] S. Li *et al.*, "Full-duplex amplify-and-forward relaying: Power and location optimization," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 8458–8468, Sep. 2017.
[18] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1637–1652, Sep. 2014.
[19] Y. Niu, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmWave) for 5G: Opportunities and challenges," *Wireless Netw.*, vol. 21, no. 8, pp. 2657–2676, Nov. 2015.
[20] Y. Niu, C. Gao, Y. Li, L. Su, D. Jin, and A. V. Vasilakos, "Exploiting device-to-device communications in joint scheduling of access and backhaul for mmWave small cells," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 10, pp. 2052–2069, Oct. 2015.
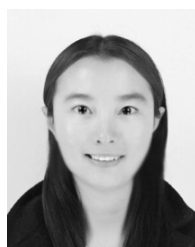
WEIGUANG DING received the B.E. degree from Beijing Jiaotong University, China, in 2016, where she is currently pursuing the M.S. degree. Her research interests include millimeter wave communications, medium access control, and VANETs.
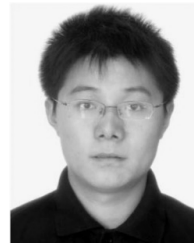
YONG NIU (S'13–M'16) received the B.E. degree in electrical engineering from Beijing Jiaotong University, China, in 2011, and the Ph.D. degree in electronic engineering from Tsinghua University, Beijing, China, in 2016.

From 2014 to 2015, he was a Visiting Scholar at the University of Florida, Gainesville, FL, USA. He is currently an Associate Professor with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University. His research interests are in the areas of networking and communications, including millimeter wave communications, device-to-device communication, medium access control, and software-defined networks. He was a recipient of the Ph.D. National Scholarship of China in 2015, the Outstanding Ph.D. Graduates and Outstanding Doctoral Thesis of Tsinghua University in 2016, and the Outstanding Ph.D. Graduates of Beijing in 2016. He has served as a Technical Program Committee member for CHINACOM 2015 and IWCMC 2017, and also a Session Chair for IWCMC 2017 and CHINACOM 2017. He also received the Outstanding Doctorate Dissertation Award from the Chinese Institute of Electronics in 2017.

HAO WU (M'10) received the Ph.D. degree in information and communication engineering from the Harbin Institute of Technology in 2000. She is currently a Full Professor with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, China. She has authored or co-authored over 100 papers in international journals and conferences. Her research interests include intelligent transportation systems, security and QoS issues in wireless networks (VANETs, MANETs, and WSNs), wireless communications, and Internet of Things. She is a reviewer of IEEE major conferences and journals in wireless networks and security.

YONG LI (S'08–M'11–SM'16) received the B.S. degree in electronics and information engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2007, and the Ph.D. degree in electronic engineering from Tsinghua University, Beijing, China, in 2012. He was a Visiting Research Associate with the Telekom Innovation Laboratories and the HK University of Science and Technology, in 2012 and 2013, respectively. From 2013 to 2014, he visited the University of Miami, Coral Gables, FL, USA, as a Visiting Scientist. He is currently a Faculty Member of electronic engineering, Tsinghua University. His research interests are in the areas of networking and communications, including mobile opportunistic networks, device-to-device communication, software-defined networks, network virtualization, and future Internet. He was a recipient of the Outstanding Postdoctoral Researcher, Outstanding Ph.D. Graduates, and Outstanding Doctoral thesis of Tsinghua University. His research was granted by the Young Scientist Fund of Natural Science Foundation of China, Postdoctoral Special Fund of China, and industry companies of Hitachi and ZET. He has authored or co-authored over 100 research papers and has 10 granted and pending Chinese and International patents. His Google Scholar Citation is about 440 with H-index of 11, also with more than 120 total citations without self-citations in Web-of Science. He has served as the Technical Program Committee Chair for WWW workshop of Simplex 2013, and the TPC of several international workshops and conferences. He is also a Guest-Editor for ACM/Springer *Mobile Networks and Applications*, Special Issue on Software-Defined and Virtualized Future Wireless Networks. He is currently the Associate Editor of the EURASIP *Journal on Wireless Communications and Networking*.

ZHANGDUI ZHONG (M'15–SM'17) received the B.S. and M.S. degrees from Beijing Jiaotong University (BJTU), China, in 1983 and 1988, respectively. He is currently a Professor with BJTU and a Chief Scientist with the State Key Laboratory of Rail Traffic Control and Safety, BJTU. He is also a Director of the Innovative Research Team, Ministry of Education, and a Chief Scientist with the Ministry of Railways in China. His research interests include wireless communications for railways, control theory and techniques for railways, and global system for mobile communications- railway. He was a recipient of the Mao Yisheng Scientific Award of China, the Zhan Tianyou Railway Honorary Award of China, and the Top Ten Science/Technology Achievements Award of Chinese Universities.

· · ·