IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON MOBILE EDGE COMPUTING AND MOBILE CLOUD COMPUTING: ADDRESSING HETEROGENEITY AND ENERGY ISSUES OF COMPUTE AND NETWORK RESOURCES

Received September 19, 2019, accepted October 13, 2019, date of publication November 7, 2019,
date of current version December 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2952174

Multi-User MAC Protocol for WLANs in MmWave Massive MIMO Systems With Mobile Edge Computing

YIFENG ZHAO[1], XUETING XU[1], YUHAN SU[1], LIANFEN HUANG[1], XIAOJIANG DU[2], AND NADRA GUIZANI[3]

[1]Department of Communications Engineering, Xiamen University, Xiamen 361005, China
[2]Department of Computer and Information Sciences, Temple University, Philadelphia, PA 19122, USA
[3]Department of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA

Corresponding author: Yifeng Zhao (zhaoyf@xmu.edu.cn)

This work was supported in part by the 2019 National Natural Science Foundation of China under Grant 61871339.

ABSTRACT Merging mobile edge computing (MEC) service with the wireless local area networks (WLANs) provides enormous benefits such as intensive computation capabilities. Access to high-speed WLANs has become an urgent challenge due to the fast rate of application development and user adoption. Providing increased wireless resources is one solution, but millimeter (mmWave) multi-user massive Multiple-Input Multiple-Output (MIMO) technology has also attracted attention due to a series of features such as high throughput, strong anti-interference ability, and small fading. Hence, exploring the potential of mmWave MIMO systems for WLAN could play a significant role in improving network performance, especially throughput. In this paper, we introduce mmWave and massive MIMO into traditional WLAN with MEC, which enables a large portion of spatial resources to be allocated among users, and mitigates inter-user interference. Moreover, we propose a novel media access control (MAC) protocol, named as LSMWN-MAC, to adapt to a series of unique characteristics of mmWave and massive MIMO, which adopts space division multiple access (SDMA) as one of the resource methods, taking full advantage of spatial resources. In addition, we modify the protocol and propose the MLSMWN-MAC protocol to allocate resources dynamically, a solution that is particularly suitable for scenes with 30-50 users. Simulation results show that compared to 802.11ad, the proposed two protocols can increase the saturated throughput by 4.7Gb/s and 5.4Gb/s respectively, which are close to the throughput upper limit of the mmWave MIMO system. Meanwhile, we observe that when the number of users exceeds the threshold, the performance of the LSMWN-MAC protocol using a static allocation method is superior to the MLSMWN-MAC protocol using a dynamic method. In other cases, the dynamic protocol has better performances.

INDEX TERMS MEC, MAC protocol, massive MIMO, millimeter wave, resource pre-allocation.

I. INTRODUCTION

Booming high-speed applications (such as 4K video, virtual reality (VR), remote surgery, and remote-controlled slave robots) have extremely strict requirements on the high-data rate of a network, which places an enormous burden on a current wireless local area network(WLAN). In addition, the high-speed transmission characteristic of the fifth generation (5G) network raises the demand for high-speed Internet, forcing the WLAN to increase computing capabilities. Because in most cases people are indoors, high-speed WLAN is needed to meet people's growing demand of Internet.

Recently, high-speed applications and a large number of devices will continue to generate massive amounts of data, which will bring two main challenges.

1. A large amount of raw data and computing tasks need to be processed, but the computing capacity of each device is limited.

2. A huge volume of data needs to be transmitted over the network with very low latency to meet the requirements of real-time tasks, while both the wireless and the wired transmission resources are inadequate in the networks.

The associate editor coordinating the review of this manuscript and approving it for publication was Junaid Shuja.

Most storage and computing of user's operations are performed by mobile devices now. In order to process large amounts of raw data quickly, we can offloading tasks from devices to the cloud, which also can reduce the resources of both devices and networks used for processes. Mobile edge computing (MEC), a key enabler of 5G future mobile communication, provides intensive computation capabilities through computing and storing resources at the edge of the networks [1]. With MEC, the access points (APs) of WLAN, integrated with computing and storage functions, can access distributed storage and acquire computing resources quickly and conveniently to satisfy user requests [2].

However, Although MEC can solve some problems in computing, for high-speed applications with mass data, the rate of uploading or downloading to the cloud may be too slow, and the communication delay may cause the entire processing time longer than that of the network without MEC [3], [4], which will impact the system performance. In this case, these high-speed applications can still not be supported well. Therefore, it is also critical for the next generation WLAN systems to provide faster uploading and downloading and higher throughput in dense environments to solve the mismatch between computing and transmission.

Transferring data in the unused, non-traditional spectrum is one of the most effective solutions to increase throughput in the WLAN. Millimeter wave communication has several advantages over existing wireless communication. First and foremost, millimeter wave (mmWave) systems can provide considerable bandwidth of nearly 2 GHz without carrier aggregation, much larger than the 20 MHz bandwidth in current 4G (fourth generation) wireless communication [5]. In addition, the mmWave transmitter and receiver support multi-beam characteristics in 5G due to small element size, so the formed beam can be narrower, which promotes the development of other applications [6].

However, although mmWave systems have great advantages in terms of bandwidth, they still suffer more severe path loss than existing communication systems. An effective solution is to combine massive Multiple-Input Multiple-Output (MIMO) technology with millimeter wave systems. For massive MIMO systems, because a multi-antenna base station can target multiple users simultaneously and construct different beams toward multiple target users with reduced interference between individual beams, we can perform beamforming and allocate beams for target receivers in specific directions, which can concentrate transmission power and receive regions on narrow beams. Meanwhile, due to the short wavelength characteristics of millimeter waves which allow large-scale antenna arrays to be packaged in small physical sizes, systems combining the two technologies can adequately compensate for high path loss caused by high frequencies and improve spectral efficiency significantly [7]. Combining massive MIMO systems with mmWave can make full use of spatial resources and offer enlarged link capacity and spectral efficiency collective [8]–[10]; this could be a panacea for most current WLAN problems.

Traditional massive MIMO systems can use multiple antennas to obtain diversity gain or multiplexing gain, to improve the reliability of data transmission or increase system capacity [11]. However, not only the base station, but also the user, needs to be equipped with expensive multiple antennas. Furthermore, when the user is at the edge of the cell, or the propagation scatter is not sufficient, the point-to-point MIMO performance will be greatly affected. A technology that can solve the disadvantages above is multi-user MIMO (MU-MIMO) [12], one of the most promising technologies for coping with increasing network traffic. It can utilize multiple users as spatially distributed transmission resources, and support transmission of multiple spatial streams simultaneously, both of which can significantly improve network throughput [13]. All users share the multiplexing gain without being equipped with multiple antennas in massive MU-MIMO systems. Therefore, to improve the wireless network performance of IEEE 802.11ad and achieve both beamforming and multiplexing gain, IEEE 802.11ay defines a new mechanism for enabling MIMO. One of its novel features is support for downlink multi-user MIMO (DL MU-MIMO) transmission [14]. In the downlink, the access point (AP) performs beamforming by precoding and transmits data to multiple users simultaneously, so multipath data can be transmitted in parallel. In order to enable users to transmit data to the AP simultaneously, the base station applies multi-user detection technology to extract signals of different users in the uplink.

However, the traffic is still severely asymmetric between uplink and downlink; i.e., the downlink throughput is limited. When the number of users is small, the uplink and downlink rates of each user are similar and relatively high. As the number of users increases, access contention and collisions will exceed the thresholds that the network can withstand, which causes the sharp drop in transmission speed, especially the downloading speed [15]. This decline will affect the user's normal access and communication between the cloud and the devices. Similarly, there are some shortcomings in terms of latency. For WLANs, the round trip air-link delay can be shortened to 1ms approximately [16]. This can support many new applications, such as high-accuracy telesurgery, smart home devices, and autopilot vehicles. Although 802.11 systems have implemented schemes to offer sub-millisecond latencies, it is still challenging to achieve the same effect in massive MIMO systems. In addition, due to a series of unique characteristics of millimeter waves, the propagation characteristics of the physical layer have changed [17]. Consequently, new media access control (MAC) protocol designs are needed to support very low latency and higher data rates [18], [19].

MAC-layer design needs to consider how to allocate communication resources such as time, frequency, and space to users efficiently and fairly [20]. User scheduling has been one of the most critical aspects of designing the MAC-layer. The essence of MAC-layer scheduling is to determine the efficient utilization of radio resources, including a user selection

scheme for sharing public radio resources, and a transport format for each selected user. Due to the expectation of high-speed communication between the cloud and the devices, it is a challenge to design a MAC-layer protocol based on mmWave massive MIMO systems to get better performance of scheduling and resources allocation.

In order to solve these problems, we propose a large-scale MIMO based wireless network MAC (LSMWN-MAC) protocol, a multi-user MAC protocol of wireless network based on mmWave massive MIMO, which is suitable for the WLAN that accommodates MEC services, especially when the number of users is large. We introduce spatial resources into the MAC layer and use space division multiple access (SDMA) combined with frequency division multiple access (FDMA) to transmit data, which enables users to have both frequency and spatial resources. Because mmWave systems are rich in spatial resources and can be used to increase the spatial capacity by supporting the hotspot region where many users (or devices) exist, developing an SDMA scheme for mmWave massive MIMO WLAN is promising.

In this protocol, OFDMA systems and multi-user MIMO systems (or hybrid systems of the two) are used, and hybrid SDMA/FDMA modes are adopted to transmit data frame. In addition, resource allocation in the protocol is optimized by using channel fading selectivity in time, frequency, and space to improve the efficiency use of system resources. The main contributions of this paper are:

- We propose using SDMA combined with FDMA in MAC layer to transmit data, so multiple downlink data can be transmitted in parallel, to make full use of spatial resources for mmWave systems in order to increase the number of downlink users and improve downlink throughput. And, we present a MAC protocol for mmWave massive MIMO WLAN, also called LSMWN-MAC, which can solve the bottleneck problem of AP downlink throughput in WLAN. Moreover, the proposed protocol can reduce the interaction time between the AP and user through merging TDMA with FDMA.
- In the design of the interaction mode, we borrow the idea of resource blocks in LTE, divide the frequency band and time into N subchannels and multiple time slots respectively, and make the protocol a multi-channel multi-user MAC protocol, reducing interaction time and improving the proportion of data frame.
- We propose another protocol, the MLSMWN-MAC protocol, which can adapt to channel changes dynamically, and then we compare the two protocol. The simulations show that the dynamical protocol can significantly improve the saturated throughput when the channel condition is poor, and dynamical protocol performance becomes worse than the static one when the number of users increases to the threshold value.

The remainder of this paper is organized as follows. We review related works in Section II. Section III introduces the system model. In section IV, we propose the LSMWN-MAC protocol that is suitable for mmWave massive MIMO environment and model the protocol. Additionally, we propose a dynamic protocol, MLSMWN-MAC in section V. In section VI, we simulate and analyze the two protocols separately and compare their performance. The main conclusions are summarized in Section VII.

The notations used in this paper are as follows. Symbols for matrices (upper-case) and vectors (lower-case) are in boldface. According to the convention, a, a, A denote a scalar, a vector and a matrix, respectively. $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$ and $(\cdot)^{-1}$ denote the transpose, the conjugate, the conjugate transpose (Hermitian) and the inverse.

II. RELATED WORK

Recently, several papers containing different aspects of MEC have been published. Many previous studies design an offload scheme or a resource management strategy, such as methods to offloading workloads from user devices to edge systems [21], [22] or to allocate computing resources to each user [1], [23]. However, there is currently little literature related to the transmission capabilities between devices and mobile edges. To achieve efficient data processing and transmission, [3] proposed a multilayer data flow processing system to fully utilize the computation capacities of the network. Assuming a basic three-node MEC system, the author of [24] developed an efficient four-slot transmission protocol for joint computing and communication, which could improve the energy efficiency of delay constraint calculations. Most authors of papers set the communication delay between the device and the mobile network edge to zero, and only considered the allocation and optimization of computing resources. In fact, for high-speed applications with mass data, the communication ability between the two may not be able to match the ideal computing speed of MEC. Therefore, research on transmission capabilities is of great significance.

For this purpose, we introduced mmWave technology combined with Massive MIMO into WLAN. There are some papers about that technology [25]–[31]; they mainly focus on the physical layer performance, including resource management [25], [26], beamforming [27], [28], channel estimation [29], modulation model [30], interference elimination [31], and so on.

Owing to the unique characteristics of millimeter waves, the propagation characteristics of the physical layer have changed. In order to meet the expectation of directional transmission, low latency, and higher data rates, new MAC protocols for mmWave systems are needed. Due to the high path loss of mmWave systems, the author of [32] proposed a TrackMAC protocol to track highly directional beams that had been allocated to users according to a specific beamforming scheme, which would reduce interference and increase resource sharing gain. Based on the quantification of the resource utilization and physical layer overhead, [17] compares analog, full-digital and hybrid architectures of MAC layer design for mmWave systems, from which the author concluded that a full-digital design with low resolution has

lower overhead than others. The novel Directional MAC protocol for Basic Stations (DMBS) in [33] extends the results of [17], and takes full advantage of spatial reusability.

However, most of the aforementioned research focuses on the physical structure of the MAC layer, and presents little on its frame structure. Hence, this work is based on the framework related to [33], which modifies the frame structure of MAC protocol for mmWave massive MIMO systems, adopts its specific character and considers spatial resource for performance improvement.

In addition, mmWave massive MIMO systems allow more concurrent links compared to traditional wireless communication systems due to the narrower beam [34]. Consequently, it is also significant to introduce a method to avoid collision and mitigate interference among users in MAC-layer due to a mass of users sharing the same wireless physical link. Some researchers focus on how to schedule users to mitigate interference and enhance capacity in the MAC layer. Considering the unique focusing effects of time reversal (TR) and massive MIMO systems, [35] proposes a novel MAC-layer scheduler with robust performance on imperfect channel-state information. Parallelizing scheduling within one RBG, the scheme proposed in [36] eliminates redundant post-scheduling operations by introducing power-based weights to amend link adaptation deviation in each RB scheduling, which is suitable for uplink MAC-layer scheduling in massive MIMO-OFDM systems in 5G.

Nevertheless, the scheduling schemes mentioned in the aforementioned works are all based on a static situation. Hence, in this paper, we draw on the idea of the resource block in LTE and design a dynamic scheduling method, which can adjust the distribution of physical links between users dynamically.

III. SYSTEM MODEL

MmWave massive MIMO systems utilize the selectivity of channel fading in time, frequency, and space. We first analyze the mechanism of mmWave massive MIMO technology to improve capacity of WLAN with MEC.

We introduce mmWave massive MIMO into WLAN with MEC, and equip the access point (AP) with a large number of antennas to serve more users at the same time. As shown in Figure 1, the AP is equipped with many antennas, and each user is only equipped with a single antenna. Users can get high QoS without increasing the complexities of hardware.

The channel capacity of a tradition point-to-point MIMO system is

$$C = \log_2 \det(\mathbf{I}_{n_r} + \frac{\rho}{n_t}\mathbf{G}\mathbf{G}^H), \tag{1}$$

where $\mathbf{I}_{n_r}$ represents the identity matrix of $n_r \times n_t$, n_r and n_t represent the number of receiving and transmitting antennas respectively. $\mathbf{G}$ represents the channel matrix of $n_r \times n_t$, where "$(\cdot)^H$" represents conjugate transpose, and scalar ρ represents the signal noise ratio (SNR). The result of singular

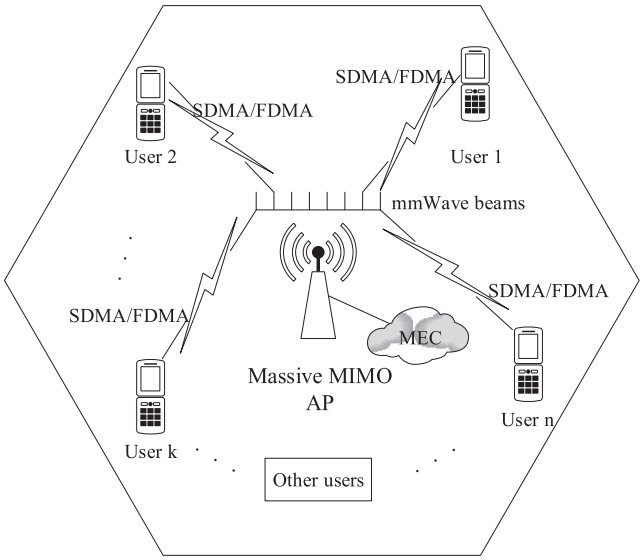


FIGURE 1. The user-intensive hotspot scenario.

value decomposition (SVD) of $\mathbf{G}$ is

$$\mathbf{G} = \Phi \mathbf{D}_v \Psi^H, \tag{2}$$

where Φ and Ψ are unitary matrix of $n_r \times n_r$ and $n_t \times n_t$ respectively, and $\mathbf{D}_v$ is diagonal matrix of $n_r \times n_t$. The diagonal elements of $\mathbf{D}_v$ are singular values and can be expressed as $\{v_1, v_2, \cdots\cdots v_{\min(n_t,n_r)}\}$. The channel capacity can be obtained as

$$C = \sum_{l=1}^{\min(n_t,n_r)} \log_2(1 + \frac{\rho v_l^2}{n_t}). \tag{3}$$

Now we discuss the upper and lower bounds of channel capacity. Let

$$\sum_{l=1}^{\min(n_t,n_r)} v_l^2 = Tr(\mathbf{G}\mathbf{G}^H), \tag{4}$$

where $Tr()$ represents the trace of matrix. The range of channel capacity is

$$\log_2(1 + \frac{\rho \cdot Tr(\mathbf{G}\mathbf{G}^H)}{n_t}) \le C \le \min(n_t, n_r) \cdot \log_2(1 + \frac{\rho \cdot Tr(\mathbf{G}\mathbf{G}^H)}{n_t \cdot \min(n_t, n_r)}). \tag{5}$$

Assuming that the channel matrix has been normalized, then (4) can approximate as

$$Tr(\mathbf{G}\mathbf{G}^H) \approx n_t n_r. \tag{6}$$

Then, (5) can be rewritten as

$$\log_2(1+\rho n_r) \le C \le \min(n_t, n_r) \cdot \log_2(1+\frac{\rho \max(n_t, n_r)}{n_t}). \tag{7}$$

As can be seen from (7), the lower bound of channel capacity occurs when there is only one singular value is nonzero,

and all others are zero. At this time, the rank of the channel matrix is 1, which occurs when the transmission route is line-of-sight or antennas are too correlated to distinguish multiple paths. In addition, the upper bound of channel capacity occurs when all singular values are nonzero and equal. At this time, the matrix is full rank, which happens when the coefficients of the channel matrix are independent and identically distributed (i.i.d) random variables.

Then, we discuss the change of channel capacity when the number of transmitting or receiving antennas increases.

When the number of transmitting antennas n_t increases, the number of receiving antennas n_r does not change and rows vectors of the channel matrix are uncorrelated to each other, we will get

$$(\frac{\mathbf{G}\mathbf{G}^H}{n_t})_{n_t>>n_r} \approx \mathbf{I}_{n_r}. \tag{8}$$

At this time, the channel capacity can approximate as

$$\begin{aligned} C_{n_t>>n_r} &\approx \log_2 \det(\mathbf{I}_{n_r} + \rho \cdot \mathbf{I}_{n_r}) \\ &= n_r \cdot \log_2(1+\rho). \end{aligned} \tag{9}$$

We see that the channel capacity of (9) reaches the upper bound of (7), which means that the number of transmitting antennas in the mmWave massive MIMO system is much larger than the number of receiving antennas, and row vectors of $\mathbf{G}$ are uncorrelated to each other, because the downlink users are separated from each other geographically. Therefore, the channel capacity can reach the upper bound when the mmWave massive MIMO system satisfies the precondition of (9).

In the mmWave MIMO system, the optimal transmission environment is one in which the singular values of the channel matrix tend to be large and stable values when the channel matrix coefficients are independent identically distributed. At this time, the channel capacity can reach the maximum, that is, the upper bound, of (7). Because the channel matrix's singular values of the mmWave massive MIMO system tend to be large stable values, the transmission environment tends to be optimal, which can make channel capacity reach the upper bound.

In summary, the mmWave massive MIMO system can greatly improve physical performance at the expense of increasing antenna cost. This is therefore a promising way to introduce the massive MIMO technology into WLAN, in order to improve the MAC-layer channel capacity.

IV. PROTOCOL STRUCTURE

In this section, we introduce the massive MIMO into WLAN and propose an LSMWN-MAC (Large-scale MIMO Based Wireless Network MAC) protocol suitable for a massive MIMO environment. The protocol is a multi-user MIMO downlink protocol, which uses the resource pre-allocation. It adopts the combination of both TDMA and FDMA techniques to reduce the interaction time between the AP and the user and combination of both SDMA and FDMA techniques to improve the network throughput. Finally, we model the protocol.

A. POSTULATED CONDITIONS

1) There are N channels available, and all channels have the same bandwidth. Packets transmitted on different channels do not affect each other, because no channels overlap. The AP and all nodes know how many channels are available in advance.
2) The AP is equipped with a large number of antennas (several hundred or more) in the center of a cell. There are M users in the cell, and each user equipped with an antenna is separate from each other geographically.
3) Because of the synchronization of nodes in the network, all nodes start the corresponding operations strictly at the specified time.
4) The AP can estimate the channel state accurately for precoding the transmitted data packet.
5) The channel state does not change within a time frame in the downlink.
6) The number of users carried by the AP on each downlink is U, that is, there are U destination nodes. Because the AP is equipped with a large number of antennas, the number of users capable of transmitting data simultaneously through beamforming is no longer limited by the number of antennas, but by the length of time during which the channel state remains unchanged.

B. PROTOCOL STRUCTURE

In this subsection, we design a downlink multi-user MAC protocol based on massive MIMO. The protocol is based on the principle that multiple degree of freedom of multi-antenna enables APs to serve multiple users simultaneously.

The protocol time frame structure in the downlink is shown in Figure 2. We divide the time frame in the transmission process into four phases.

Phase 1: The AP determines the maximum number of users allowed to access the system at the same time according to the number of effective receiving antennas. Users obtain transmission opportunities through competition. After the nodes with transmission opportunities are determined, the AP transmits the Multi-user Ready to Send (MU-RTS) frame, which contains the addresses of U (U<M) downlink destination nodes, to all users in channel 1. This process is similar to the transmission process of Ready to Send (RTS) frames in the RTS/CTS protocol. In the RTS/CTS protocol, all users sense channel state before sending a frame. When the channel state is idle, users send RTS frames to compete for transmission opportunities. In the MU-MIMO scenario of the LSMWN-MAC protocol, the U nodes of the M nodes obtain transmission opportunities. The AP sends MU-RTS frames containing the destination addresses of these nodes.

Phase 2: M nodes receive the MU-RTS frame in channel 1 and obtain the addresses of the U destination nodes from it to determine whether they are obtaining the transmission opportunity. The node that gets the transmission opportunity

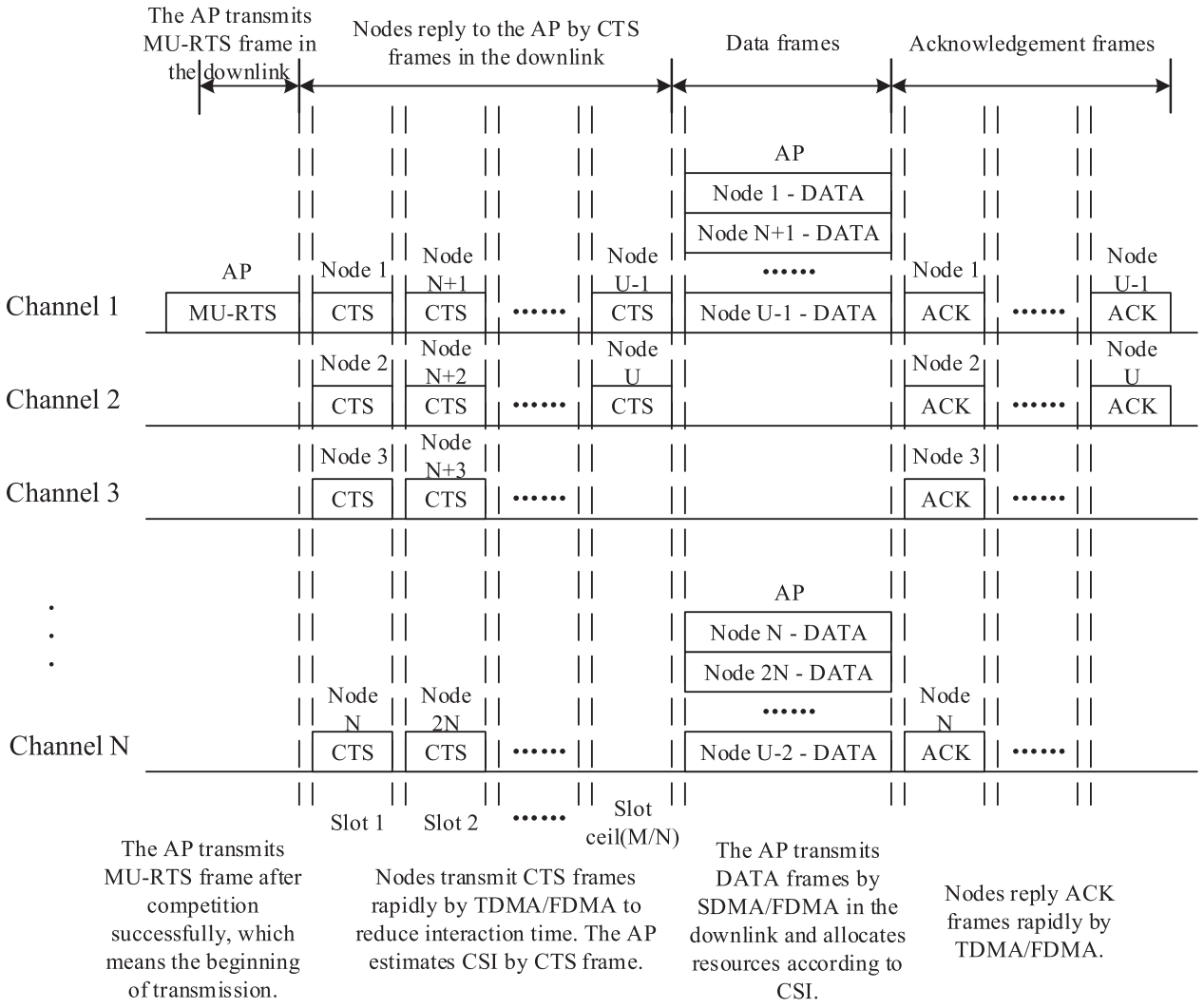


FIGURE 2. Time frame structure of LSMWN-MAC protocol in the downlink.

replies to the Clear to Send (CTS) frame in order, using the TDMA/FDMA. The AP estimates the channel state information (CSI) of the corresponding node on the corresponding channel through the CTS frame for subsequent resource allocation and precoding, which is similar to the channel estimation method of the TDD mode. The AP can use the uplink channel estimation information to perform downlink precoding in coherence time according to channel reciprocity. The overhead of using the AP for channel estimation is pretty low.

Because the AP needs to estimate the CSI through the CTS frame replied by the user, the interaction between the AP and the user is an essential part of the multi-user MIMO MAC protocol. The application scenario of the protocol designed in this paper is a hotspot area with many users, and it uses massive MIMO downlink to carry many users at the same time. With the traditional method, interaction time will be too long to decrease the proportion of data frame transmission time.

In the design of the interaction mode, this paper draws on the idea of resource blocks in LTE and divides the frequency band and time into N subchannels and multiple time slots separately, making the protocol a multi-channel multi-user MAC protocol. We allocate each small resource block to the users to reply to the AP by CTS frames, thereby reducing interaction time. The specific allocation manner is as follows: The user decides which resource block transmits the CTS frame according to the order of the address list of the destination node in the MU-RTS frame. Assuming that there are N channels with sufficient intervals and U destination nodes, the reply CTS can be completed by $P = \left[\frac{U}{N}\right]$ ($[\cdot]$ means round down) time slots. For example, if 40 users need to transmit in the downlink and 10 subcarriers are available, then $40/10 = 4$ time slots will allow the 40 sites to reply to the CTS frame. The first to the tenth site in the address list select the first to the tenth subcarriers of the first slot in turn to reply to the CTS frame. Similarly, the eleventh to the twentieth site select the first to the tenth subcarriers of the second

slot in turn. By analogy, the user U selects the subcarrier Q ($Q = U\%N$ and "%" means remainder) of the time slot P. The user sends a CTS frame at the beginning of each time slot, and the AP receives the 40 CTS frames correctly, thus estimating the channel and preparing for downlink transmission.

The number of required time slots is calculated according to the number of available channels and users carried in the downlink by the AP, which can be adjusted dynamically. Consequently, the duration of this phase can be dynamically changed.

Phase 3: The AP estimates the CSI according to the received CTS frame, and allocates resources accordingly. The estimated CSI of the AP is the user's channel matrix $\mathbf{G}$. Let $\mathbf{G}_{i,j}$ represent the channel matrix of the ith user in channel j. The AP transmits block by SDMA/FDMA in the downlink. The spatial and spectrum resources are available to the user. Therefore, the AP needs to conduct beamforming and allocate subcarrier for the user.

The resources of the OFDMA/SDMA system include spectrum resources and space resources, and the resource allocation problem involves subcarrier allocation, user spatial grouping, and power allocation. We divide the resource allocation problem of the system into two steps. The first step is to select users, that is, to allocate a group of users on each subcarrier. Therefore, we need to first group user spatially. The second step is to allocate bit and power for each user on each subcarrier.

1) user spatial grouping
   In a multi-user MIMO system, the geographical locations of different users are independent of each other, and the channel fading is also spatially irrelevant. Therefore, with a suitable resource allocation algorithm, we can obtain multi-user spatial diversity gain. In practice, the number of transmitting antennas at the base station or the AP is generally limited, but the number of users is large, resulting in a few downlink users. Therefore, we need to select some users for transmission. The goal of resource allocation is to select a group of users according to different optimization goals, and to optimize system performance through some resource allocation methods.
   The basic idea of user spatial grouping is to put users with strong spatial orthogonality into a group. If there is a difference among user priorities, we first select the user with the highest priority, and select the remaining users who are orthogonal to it into a group. Then we select the highest priority users among the remaining users. We repeat the previous operation until the grouping is complete.
2) bit and power allocation
   When the SNR is high, the channel capacity obtained by the water-filling power algorithm is similar to that obtained by the average power allocation algorithm. Therefore, in order to reduce the complexity, we assume that the power is evenly allocated. Complex joint resource optimization in the spatial, time, and frequency domains is translated into the problem of selecting the optimal spatial user group on each subcarrier.

According to the resource allocation method of multi-user in the spatial frequency domain, the users are first spatially grouped; then, the user group is placed on the most suitable subcarrier. In the LSMWN-MAC protocol, we divide the bandwidth and time into N subchannels and multiple time slots significantly, and allocate resource blocks to the users to send CTS frames. The user decides the resource block it uses based on the order of the destination node address list in the MU-RTS frame. Users who send CTS frames on a same channel are classified into a same spatial user group. For example, user 1, user $N+1, \ldots$, user $(U-Q)$ replying in channel m become a space group. The AP obtains corresponding channel matrix $\mathbf{G}_{1,m}$, $\mathbf{G}_{N+1,m}, \ldots$, $\mathbf{G}_{(U-Q),m}$ according to the CTS frames replied by users, thereby constituting the channel state matrix $\mathbf{G} = [\mathbf{G}_{1,m}\ \mathbf{G}_{N+1,m}\ \ldots\ \mathbf{G}_{(U-Q),m}]$ of channel m. In other words, the resources are already allocated in advance according to the order of destination address list (the channel used by the user and the space group are determined in advance). In this protocol, the AP does not need to allocate the resource; it is only necessary to precode according to $\mathbf{G}$ of the corresponding channel, and send data frames to the spatial user group in the channel simultaneously by beamforming.

However, the premise of all current resource allocation algorithms is that the AP knows all $\mathbf{G}$s of every user on each sub-channel, and the optimization of resource allocation is performed on this basis. However, because the user sends a CTS frame on a certain sub-channel in the protocol, the AP can only estimate $\mathbf{G}$ of the user on its sub-channel and does not know $\mathbf{G}$ of the user on other sub-channels, therefore hindering global optimization of resource allocation.

As shown in Figure 2, the AP conducts beamforming on N channels at the same time and transmits data frames to U users by SDMA/OFDMA simultaneously.

Phase 4: The user who received the data frame replies to the Acknowledgement (ACK) frame in the same way as the method described in phase 2.

In summary, this section designs a multi-channel multi-user downlink MAC protocol based on mmWave massive MIMO.

C. PROTOCOL MODEL ANALYSIS

In this subsection, we model the protocol and calculate its saturated throughput in theory. Because we design a downlink protocol in this section, only downlink data is sent; only the AP sends data to the user node, and no node transmits data to the AP. Therefore, there is no need to consider the competition. The AP sends data to the user node at the beginning of each time frame.

We assume that the number of channels is N, the AP is equipped with K antennas, and the number of users in the

downlink is U. The downlink frame duration is

$$T_{DL-frame} = T_{MU-RTS} + \left\lceil \frac{U}{N} \right\rceil \times T_{CTS} + T_{DATA} + \left\lceil \frac{U}{N} \right\rceil \times T_{ACK}, \quad (10)$$

where T_{MU-RTS} represents the MU-RTS frame duration.

The AP can obtain the channel matrix $\mathbf{G}_n$ of the channel n accurately through channel estimation. Using $\mathbf{G}_n$ to pre-code, we can implement SDMA. Due to the use of massive MIMO, the AP will perform well only by using a simple linear precoding technique. In this chapter, we use matched filtering to pre-code and use G_n^H as a precoding matrix, where "$(\cdot)^H$" represents conjugate transposition. Then the signal received by the user in channel n is:

$$\mathbf{y} = \mathbf{G}_n \mathbf{G}_n^H \mathbf{s} + \mathbf{w}, \quad (11)$$

where $\mathbf{y}$ and $\mathbf{s}$ represent receiving and transmitting signal respectively. $\mathbf{w} \sim \mathcal{N}(0, \sigma^2)$ is an additive Gaussian noise vector. The users in channel n are determined by pre-allocation. For simplicity, we make the number of users U in the downlink an integer multiple of the number of channels N. Then, the number of users in each channel is $L = U/N$. Therefore, $\mathbf{y}$, $\mathbf{s}$ and $\mathbf{w}$ are $L \times 1$ column vector. $\mathbf{G}_n$ is $L \times K$ matrix and $\mathbf{G}_n^*$ is $K \times L$ matrix.

According to [37], the channel matrix $\mathbf{G}_n$, which combines a small-scale fading matrix $\mathbf{H}_n$ with a large-scale fading matrix $\mathbf{D}$, is given by

$$\mathbf{G}_n = \mathbf{H}_n \mathbf{D}^{1/2}. \quad (12)$$

The coefficient g_{nku} of matrix $\mathbf{G}_n$ is:

$$g_{nku} = h_{nku} \sqrt{\beta_u}, \quad k = 1, 2, \ldots, K, \quad (13)$$

where h_{nku} represents the fast fading coefficient (small-scale fading coefficient) of the user u to antenna k of the AP in channel n. We assume that the small-scale fading coefficients are i.i.d complex Gaussian, i.e., $h_{nku} \sim \mathcal{CN}(0, 1)$. h_{nku} remains unchanged in a downlink frame. $\sqrt{\beta_u}$ is the slow fading coefficient (large-scale fading coefficient) of the user u to the AP. $\sqrt{\beta_u}$ of different users are independent of each other. If d_u is the distance between user u and the AP, b_0 is the path-loss at reference distance d_0, α is the path-loss exponent, and z_{mk} is the channel gain due to shadowing noise, we model the large-scale fading coefficient β_u as [38]

$$\beta_u = b_0 d_u^{-\alpha} 10^{\frac{z_{mk}}{10}}. \quad (14)$$

For simplicity, we ignore the effect of shadowing noise.

According to [39], the SINR (signal to interference plus noise ratio) of the received signal of user u in channel n is given by:

$$SINR_u = \frac{\|\mathbf{p}_u \mathbf{q}_u\|^2}{\sigma^2 + \sum_{l=1, l \neq u}^{l=L} \|\mathbf{p}_u \mathbf{q}_l\|^2}, \quad (15)$$

where $\mathbf{p}_u$ represents the line u of the matrix $\mathbf{G}_n$, i.e., the channel vector of the user u to the AP and $\mathbf{q}_l$ represents the precoding vector for user l.

According to the law of large numbers, $\sum_{l=1, l \neq u}^{l=L} \|\mathbf{p}_u \mathbf{q}_l\|^2$ tends to zero when the AP uses a large number of antennas (K≫L) ideally [34], which means that interference between users in a cell is eliminated. This is the reason why we can achieve a good SINR even using simple linear precoding.

So (15) can be simplified as

$$SINR_u = \frac{\|\mathbf{p}_u \mathbf{q}_u\|^2}{\sigma^2}. \quad (16)$$

For simplicity, we use binary phase shift keying (BPSK) modulation. The bit error ratio (BER) of receiving terminal of the user u obtained by coherent detection method in a Rayleigh channel is given by

$$BER_u = \frac{1}{2}\left(1 - \sqrt{\frac{SINR_u}{1 + SINR_u}}\right). \quad (17)$$

The probability that a binary sequence with a bit rate of B has v errors during T, which follows a Poisson distribution, is given by

$$P(v) = \frac{(B \cdot T \cdot BER)^v}{v!} e^{-B \cdot T \cdot BER}, \quad (18)$$

where BER is not for a specific user.

We assume that the packet length is R bit, and the probability of at least one error bit within R bit is $1 - P(v = 0)$. We make the user drop a packet as long as one bit is wrong in this packet, i.e., the packet loss ratio for user u is:

$$PER_u = 1 - e^{-R \cdot BER_u}. \quad (19)$$

The data is sent to U users in one downlink frame, that is, there are U data packets. The average number of received packets is:

$$NUM = \sum_{u=1}^{U} (1 - PER_u). \quad (20)$$

The saturated throughput is given by

$$Throughput = \frac{NUM \times payload}{T_{DL-frame}}, \quad (21)$$

where payload represents the length of a data packet, that is, the length of the information portion after the auxiliary information of both the packet header and the packet trailer is removed.

V. THE MLSMWN-MAC PROTOCOL

We use the method of resources pre-allocation in the protocol proposed in Section IV. The user determines the channel used during data transmission and the space user group to which it belongs in advance, according to the destination address list in the MU-RTS frame. Because the AP can only obtain the CSI of the user in a single channel,

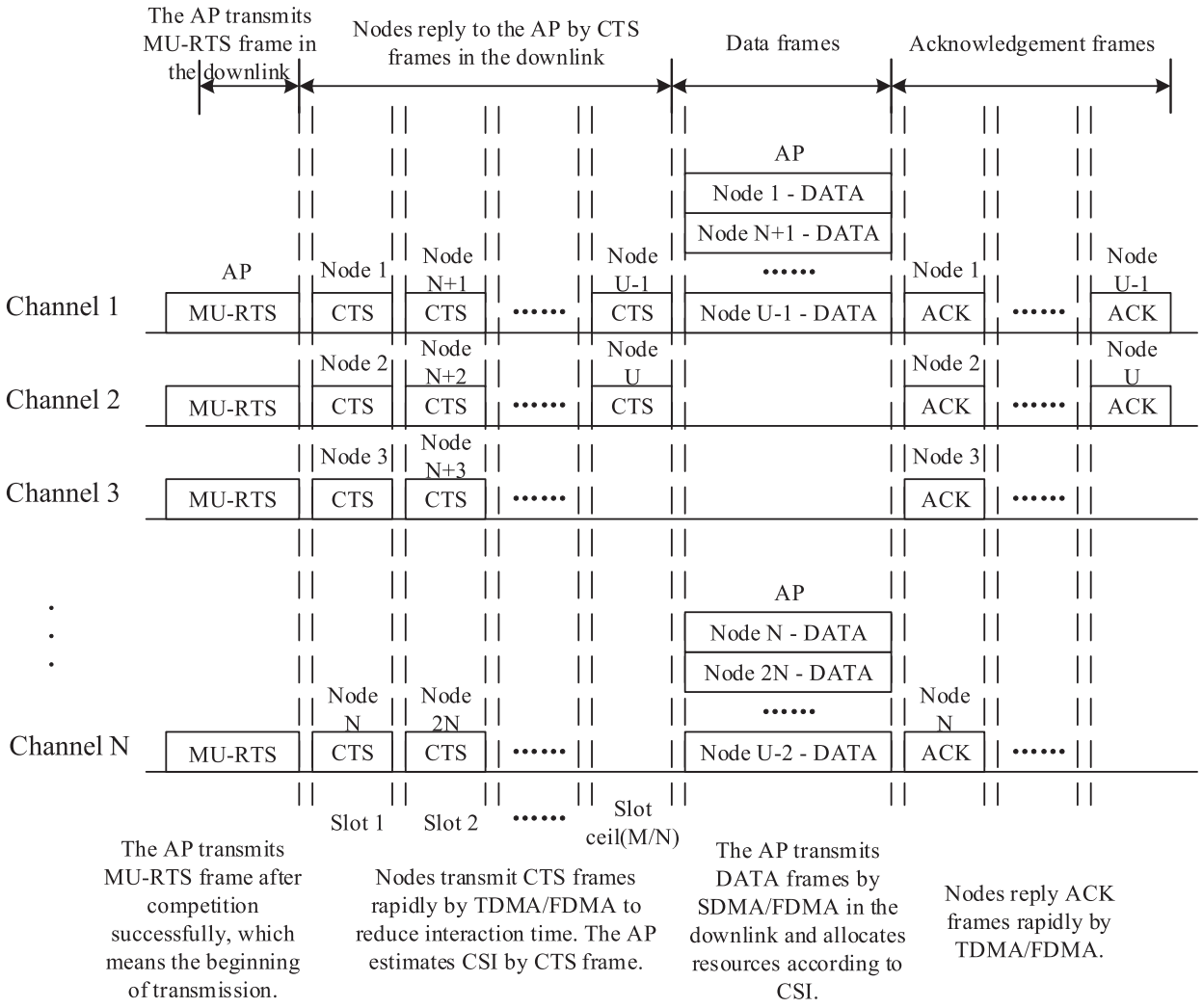


FIGURE 3. Time frame structure of MLSMWN-MAC protocol in the downlink. (The main difference of time frame between MLSMWN-MAC algorithm and LSMWN-MAC algorithm is in phase one.)

we cannot optimize the global resource allocation under the condition of limited CSI; such a fixed manner of resource allocation cannot adapt to channel changes dynamically and optimize resource allocation well. Therefore, we propose the MLSMWN-MAC (Modified Large-scale MIMO Based Wireless Network MAC) protocol, which can optimize resource allocation by allocating resources dynamically and improve throughput.

A. PROTOCOL STRUCTURE

The structure of the MLSMWN-MAC protocol is almost the same as that proposed in subsection IV-B. The difference is that we no longer use the AP but the user node to estimate the channel state in the MLSMWN-MAC protocol in order to ensure the AP obtains enough channel state information for resource allocation. The channel state is fed back to the AP by the CTS frame.

We assume that the user node can estimate the channel state information in each channel and feedback to the AP accurately. At the same time, the channel state does not change within a time frame. As shown in Figure 3, the transmission process of the protocol is divided into four phases.

Phase 1: The AP obtains the opportunity for downlink transmission through competition. The MU-RTS frame is transmitted in all channels if the competition is successful. The MU-RTS frame contains the addresses of $U(U < M)$ downlink destination nodes.

Phase 2: U destination nodes receive the MU-RTS frame in N channels, thereby estimating channel matrixes of N channels and feeding back to the AP respectively by encapsulating them into CTS frames. In this protocol, in order to achieve global optimization, we obtain all CSI of U users on N channels by channel estimation through users, which is similar to the channel estimation method of the FDD mode. Although the channel estimation overhead increases with the increase of the number of antennas, the complexity of downlink channel estimation and the amount of feedback data of the massive MIMO OFDMA system used in the protocol

are both low, based on the channel sparsity characteristics of massive MIMO OFDMA system.

The user replies to the CTS frame in order, according to the destination address list, which is the same as the method described in subsection IV-B. In this way, the AP obtains all CSIs of the U users in N channels and can optimize resource allocation. Although this method increases the complexity of the user device, it can be used to improve the network performance if conditions permit.

Phase 3: In this phase, the AP allocates resources to optimize network throughput first. For simplicity, we assume that U is an integer multiple of N. Due to the advantages of massive MIMO, precoding can eliminate interference within users in the SDMA system. The user groups in different spaces do not affect the performance of users. Therefore, this protocol does not consider spatial optimization in resource allocation, but only considers the optimization of spectrum resource allocation; then, users will be grouped into different user groups automatically.

The AP allocates $R = \frac{U}{N}$ users in each channel. According to the resource allocation method of OFDMA system, the AP can obtain the CSI of U users on N channels respectively, so we will calculate the instantaneous SNR of them according to these CSI. We put these users on the corresponding channels, in descending order of SNR. Therefore, $\frac{U}{N}$ users on each channel form a spatial user group automatically, which can be used for SDMA.

For the LSMWN-MAC protocol, we cannot use it for global resource optimization because all users transmit CTS frames on one subchannel. The AP estimates only the channel matrices $\mathbf{G}$ of users on this channel, and does not know $\mathbf{G}$ of them on other channels. However, for the MLSMWN-MAC protocol, the AP knows $\mathbf{G}$ of all users on each subchannel, which can be used for global resource optimization.

We make the AP precode according to the channel matrix of the corresponding channel and use beamforming for transmitting data frames to the spatial user groups simultaneously. As shown in FIGURE 3, the AP conducts beamforming in N channels at the same time and transmits DATA frames to U users simultaneously by SDMA/FDMA.

Phase 4: The user who received the DATA frame replies to the ACK frame via the same method as described in phase 2.

VI. PERFORMANCE EVALUATION

A. PARAMETER SETTINGS

We assume that the channel state does not change within a downlink frame but changes at its beginning. Table 1 summarizes the parameters used in the simulations.

B. SIMULATION RESULTS IN LSMWN-MAC

Here, we first verify the feasibility of the model. Next, we evaluate the performance of the proposed LSMWN-MAC protocol. Consider an mmWave massive MIMO system. We analyze the impact of several parameters, i.e., transmit SNRs, the number of channels, and the number of downlink users. The overall saturated throughput of the system can be calculated according to (21).

TABLE 1. Parameters used in simulations.

Parameter	value
Path-loss parameters (3GPP UMi [42])	d_0 = 10m l_0=-47.5dB α=2 (free space)
Channel rate	1 Gb/s
Noise power	1 mW
Propagation delay	1 μs
Transmission time interval	1 ms
Simulation time	10 s
Omnidirectional beamforming gain at BS	0 dB
Max beamforming gain for the BS assuming $K = 10$	10 dB
Actual payload length of packet	512 byte
MAC header length	224 byte

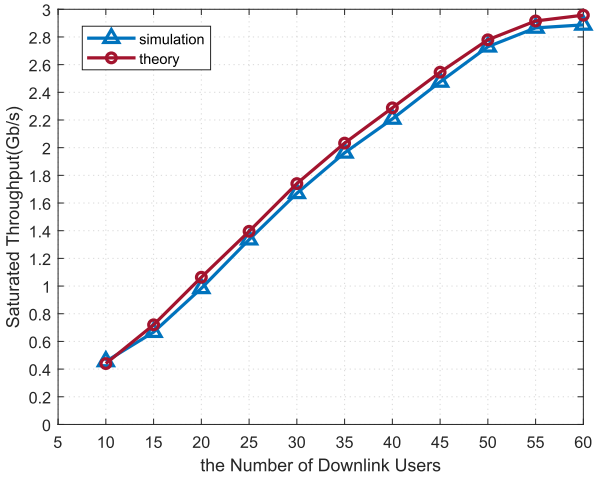


FIGURE 4. Comparison between theoretical model and simulation result for the SNR = 20dB and the number of channels $N = 6$.

1) VERIFY THE FEASIBILITY

In the following, we compare the theoretical model with the simulation results. When the SNR of the transmitting signal is set to 20dB and the number of channels to 6. The saturated throughput obtained by the theoretical model is nearly similar to that of the simulation results, indicating that the theoretical model in this paper is correct. Figure 4 shows the comparison between the theoretical model and the simulation result.

2) IMPACT OF TRANSMIT SNRs

Next, we investigate the impact of transmit SNRs on the saturated throughput by the different number of antennas for different transmit SNRs. When the number of channels is set to 6, and the number of downlink users to 10, according to a typical mmWave massive MIMO system, as shown in Figure 5, the curves increase as transmit SNR increases, and all then tend to be stable. The reason is that the higher transmit SNR, the higher receive SNR of users, and the lower the bit error rate and packet loss rate, resulting in higher throughput. However, when we increase the SNR to

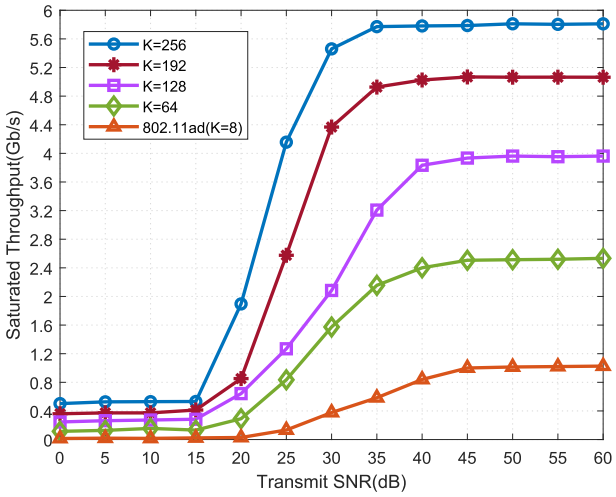


FIGURE 5. Comparison of saturated throughputs for systems with the different number of antennas in terms of SNR for the number of channels $N = 6$ and the number of downlink users $U = 10$.

the threshold, the packet loss rate is already low enough and almost no longer changes even though the SNR is still increased, resulting in a stable saturated throughput.

In addition, when the transmit SNR is low, massive MIMO can only increase the saturated throughput slightly (for example, when $K = 256$, the throughput increases by $0.49Gb/s$ compared with 802.11ad; when $K = 196$, it only improves $0.35\ Gb/s$). This is because the noise power is approximately equal to the transmission power when the transmit SNR is low. At this time, the increase in the number of antennas cannot compensate for the interference of noise on the transmission process, resulting in a high bit error rate (BER). However, when the SNR increases gradually, the greater the number of antennas, the larger the slope of the curve, that is, the faster the saturated throughput increases. As the SNR continues to increase, the saturated throughputs of mmWave massive MIMO systems are several times that of 802.11ad (for example, when $K = 256$, it increases by 4.8 times; when $K = 196$, it increases by 3.9 times). When the number of antennas increases, the sum of single channels capacities increases, causing a decrease in BER and an increase in throughput.

Furthermore, with the increase of the number of antennas, the transmit SNR corresponding to the saturated throughput threshold decreases (for example, when $K = 256$, the SNR is about 30; however, $K = 8$, SNR=45). This is because when K is small, the inter-user interference (IUI) is relatively large, and the transmission of data packets is not only affected by noise but also by other users. As K increases, the channel hardening characteristic emerges, resulting in the decreasing of IUI, and the system is almost only interfered by noise. Therefore, for mmWave massive MIMO systems, a slight increase in SNR can make the signal power larger than the sum of noise power and the interference power, so that the saturated throughput reaches the threshold earlier than 802.11ad. Hence, when the AP uses large-scale antenna arrays, a small transmit SNR can make the systems achieve good performance.

Finally, we can observe that as K increases, the gaps of saturated throughput thresholds become smaller (for example, when K increases from 8 to 64, the throughput threshold increases by 2.94 Gb/s; however, when K increases from 192 to 256, that only increases by $0.75Gb/s$), that is, the system performance improves slower gradually. The reason is that in terms of spatial resource, the antenna directivity is not high, and the coverage areas between beams are large when K is small. When K increases, the antenna directivity improves, and the beams become narrow, resulting in small coverage areas and low inter-beam interference. However, when K is relatively large, the antenna directivity is already high, and the degree of beam overlap is small. Increasing K on this basis cannot improve system performance significantly, leading to a result that the gaps in saturated throughput thresholds become smaller. In practice, the increase of K will make hardware more complex. Therefore, it is necessary to consider both system performance and hardware complexity comprehensively to determine the number of AP antennas.

3) IMPACT OF THE NUMBER OF CHANNELS

Here, we analyze the impact of the number of channels on the saturated throughput. From Figure 5, we can conclude that there are no apparent differences among all saturated throughputs of mmWave massive MIMO systems with the different number of antennas when the transmit SNR is less than 20dB. However, when it increases to 20dB, the saturated throughput increases dramatically. Hence, we set the transmit SNR to 20dB to investigate the impacts of other parameters on system performance.

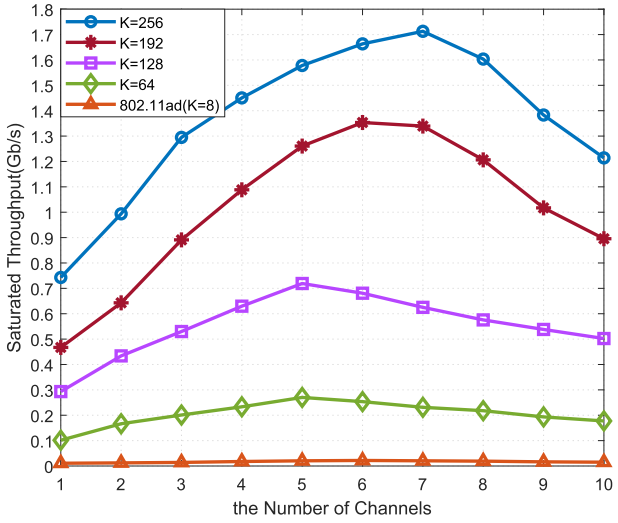


FIGURE 6. Comparison of saturated throughputs for systems with the different number of antennas in terms of the number of channels for the transmit SNR = 20dB and the number of downlink users $U = 10$.

As shown in Figure 6, we compare saturated throughputs for systems with a different number of antennas in terms of channel number. We can see that the saturated throughputs

first increase and then decrease, as the number of channels increases. The reason for increased throughput is in the second phase of the protocol, U users need $[U/N]$ time slots to reply to the CTS frame. When the number of users remains unchanged, the required timeslots will decrease as the number of channels increases, resulting in a decrease in total downlink frame time used. Therefore, the number of transmitted downlink frames and received data packets increases in a fixed time, increasing throughput.

However, with the increasing of channel number, the inter-channel interference (ICI) will increase rapidly, leading to a result that the negative impact of interference gradually approaches and exceeds the positive effect of the increase of the number of channels. When the two are equivalent, the saturated throughput reaches a threshold and then decreases as the number of channels increases. Therefore, it is critical to select the appropriate number of channels to achieve optimal system performance in practice.

In addition, when K increases, the number of channels corresponding to the saturated throughput thresholds also increases (for example, when $K = 64$, the number of channels is 5, and when $K = 256$, that is 7). Therefore, mmWave massive MIMO systems are more resistant to ICI.

4) IMPACT OF THE NUMBER OF DOWNLINK USERS

Now, we investigate the impact of the number of channels on the saturated throughput. From Figure 6, we can conclude that when the number of channels is 6, the saturated throughput of mmWave massive MIMO systems with the different number of antennas are all close to the maximum approximately. Hence, we set the number of channels to 6, and the transmit SNR to 20dB, in order to investigate the impact of the number of downlink users on system performance.

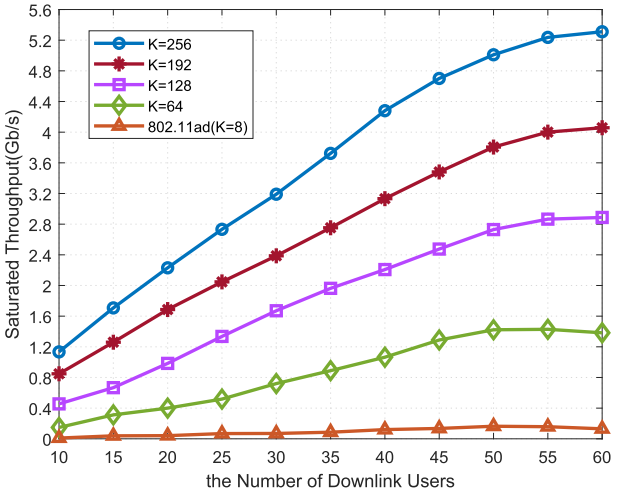


FIGURE 7. Comparison of saturated throughputs for systems with the different number of antennas in terms of the number of downlink users for the transmit SNR = 20dB and the number of channels $N = 6$.

As shown in Figure 7, we compare saturated throughputs for systems with a different number of antennas in terms of the number of downlink users. All curves show an upward trend as the number of users increases. The reason is that as the number of users increases, the sum of data packets received by all users increases. Hence, the received SNR increases, leading to the system saturated throughput improvement. Furthermore, the greater the number of antennas, the more total packets that users can receive, and the higher the system saturated throughput.

Moreover, the curves all become flat gradually or even decline as the number of users increases. The reason is that when the number of users increases, the inter-user interference(IUI) increases too. Although the massive MIMO system can reduce the inter-cell user interference to a certain extent owing to its channel-hardening characteristic, when the number of users increases, the user distribution becomes denser, and the beams allocated to different users will have serious conflicts. As a result, the probabilities of collision and loss of the data packet increase during transmitting, that is, the packet loss rate increases and the system throughput tends to be stable or worse.

C. SIMULATION RESULTS IN MLSMWN-MAC

Then we discuss the performance of MLSMWN-MAC protocol. Similarly, we analyze the impact of several parameters, i.e., transmit SNRs, the number of channels, and the number of downlink users.

1) IMPACT OF TRANSMIT SNRs

Firstly, we investigate the impact of transmit SNRs on the saturated throughput by the different protocols for different transmit SNRs when the number of channels and downlink users are 6 and 10, respectively. Solid lines depict the saturated throughputs of the systems using the LSMWN-MAC protocol, and dashed lines depict that of the systems using the improved protocol - MLSMWN-MAC protocol.

From Figure 8, we can observe that for different K, the saturated throughputs of systems using the

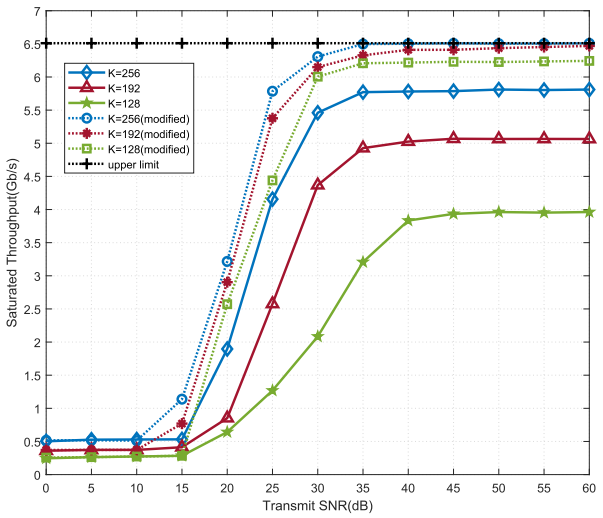


FIGURE 8. Comparison of saturated throughputs for systems with the different number of antennas for different protocols in terms of SNR for the number of channels $N = 6$ and the number of downlink users $U = 10$.

MLSMWN-MAC protocol are always greater than those using the LSMWN-MAC protocol when the SNR is increased gradually. For example, when K = 128, the saturated throughput increases by 2.2 Gb/s; when K = 256, that increases by 0.7 Gb/s. This is because each user sends its own CSI of all channels to the AP in the MLSMWN-MAC protocol so the AP can allocate resources dynamically according to all CSIs. However, in the LSMWN-MAC protocol, the AP can only obtain the CSI of the user of a certain channel, so resource allocation cannot be considered globally. Hence, the performance of the system using the modified protocol greatly improves. Moreover, to further illustrate the good performance that this modified protocol can achieve, we compare the maximum throughput that the system with the modified protocol can achieve and the upper limit of the mmWave massive MIMO system. We use the black dashed line in Figure 8 to show the upper limit. We can conclude that as the number of antennas increases, the system throughput increases gradually and approaches the upper limit.

In addition, the saturated throughput of systems with a different number of antennas has a similar result. Compared to the LSMWN-MAC protocol, the differences in saturated throughput due to an increase in the number of antennas in the MLSMWN-MAC protocol are reduced. The reason is that for the MLSMWN-MAC protocol, the saturated throughput of systems equipped with a different number of antennas all almost reach the theoretical optimal value. However, for the LSMWN-MAC protocol, there is a large gap between the saturated throughput of the system and the theoretical optimal value, and the number of antennas is still a critical variable affecting throughput. Hence, the MLSMWN-MAC protocol more significantly improves performance of the systems with a relatively small number of antennas. Therefore, for the modified protocol, excellent system performance can be achieve with a small number of antennas.

2) IMPACT OF THE NUMBER OF CHANNELS

Then, we compare saturated throughputs for systems with the different number of antennas for different protocols in terms of the number of channels. As shown in Figure 9, for different K, the curves still all increase first, then decline as the number of channels increases. In addition, the system throughputs of the MLSMWN-MAC protocol are all greater than that of the LSMWN-MAC protocol. For example, when K=128, the saturated throughput increases by 1.5Gb/s after using the modified protocol; when K=256, that increases by 1.0Gb/s. Hence, we can also conclude that using the modified protocol greatly improves the performance of the system.

In addition, the number of channels corresponding to the saturated throughput threshold of the system using the MLSMWN-MAC protocol declines relative to the LSMWN-MAC protocol, because the CSI of all users of each channel are needed for resource allocation. When the number of channels increases, ICI also increases, resulting in an increase in the probability of packet error or loss for transmitting CSI. Therefore, the increased number of channels has a certain degree of effect on throughput.

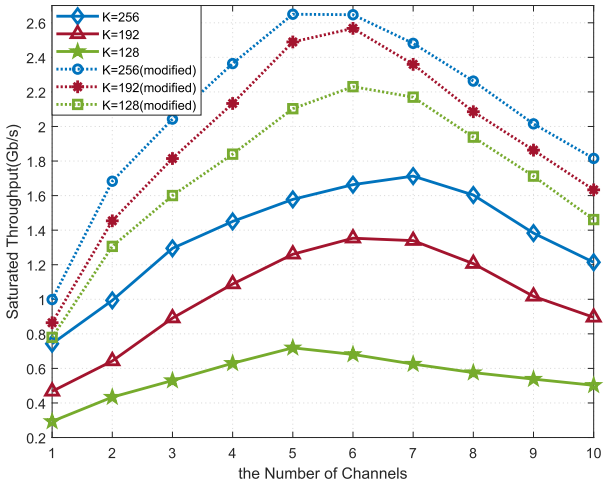


FIGURE 9. Comparison of saturated throughputs for systems with the different number of antennas for different protocols in terms of the number of channels for the transmit SNR = 20dB and the number of downlink users $U = 10$.

3) IMPACT OF THE NUMBER OF DOWNLINK USERS

As shown in Figure 10, we compare saturated throughputs for systems with a different number of antennas for different protocols in terms of the number of downlink users. We can conclude that when the number of downlink users is relatively small, the saturated throughputs of the systems with different antenna numbers are all improved significantly, that is, about 2.4 Gb/s. However, as the number of downlink users continues to increase, the curves become flat gradually and then decline. The reason is that in the MLSMWN-MAC protocol, all users receive MU-RTS frames on N channels,

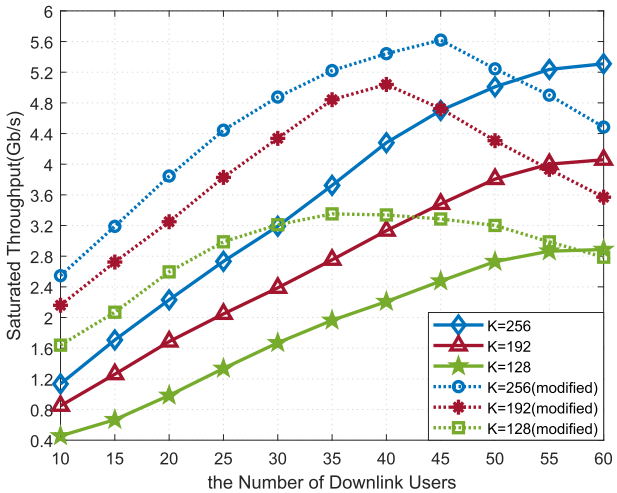


FIGURE 10. Comparison of saturated throughputs for systems with the different number of antennas for different protocols in terms of the number of downlink users for the transmit SNR = 20dB and the number of channels $N = 6$.

thereby estimating the channel matrix of N channels, and then feeding back to the AP. When the number of users is too large, the collision probabilities of data packets are so large that packet loss is more frequent, resulting in a decrease in saturated throughput.

Therefore, the MLSMWN-MAC protocol is more suitable for systems with fewer users. However, when the number of users increases to a certain extent, the static protocol, LSMWN-MAC protocol, is needed to ensure the throughput.

VII. CONCLUSION

In this paper, we have introduced mmWave massive MIMO technology into WLAN with MEC service, proposed a method for the pre-allocating wireless resource, and designed the LSMWN-MAC protocol in mmWave massive MIMO systems. We have used TDMA/FDMA to reduce interaction time between the AP and users and SDMA/FDMA to transmit data, which makes full use of the mmWave spatial resources and improves network throughput significantly. Then, we have proposed the modified protocol, MLSMWN-MAC protocol, and the improved resource allocation method for the shortcomings of the pre-allocation scheme of the LSMWN-MAC protocol. We have used the method of dynamic resource allocation to make the systems adapt to real-time changes in the channel. Finally, we have analyzed and simulated the two protocols separately. The simulation results show that both protocols significantly improve saturated throughput, thereby meeting user demand for high-speed Internet. MLSMWN-MAC performs better than the LSMWN-MAC protocol in most cases. However, when the number of users is large, the static LSMWN-MAC protocol achieves better system performance. Therefore, for the networks with MEC, the proposed protocols can increase its communication performances significantly, which will solve the rate mismatch between high-speed cloud computing and low-speed uploading and downloading of devices.

Future work will delve into how users interact with the AP in the uplink. At the same time, and the wireless resource allocation in uplink.

REFERENCES

[1] Y. Sun, S. Zhou, and J. Xu, "EMM: Energy-aware mobility management for mobile edge computing in ultra dense networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 11, pp. 2637–2646, Nov. 2017.

[2] S. Barbarossa, E. Ceci, M. Merluzzi, and E. Calvanese-Strinati, "Enabling effective mobile edge computing using millimeterwave links," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Paris, France, May 2017, pp. 367–372.

[3] P. Wang, C. Yao, Z. Zheng, G. Sun, and L. Song, "Joint task assignment, transmission, and computing resource allocation in multilayer mobile edge computing systems," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2872–2884, Apr. 2019.

[4] J. Liu, P. Li, J. Liu, and J. Lai, "Joint offloading and transmission power control for mobile edge computing," *IEEE Access*, vol. 7, pp. 81640–81651, 2019.

[5] X. Gao, L. Dai, and A. M. Sayeed, "Low RF-complexity technologies to enable millimeter-wave MIMO with large antenna array for 5G wireless communications," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 211–217, Apr. 2018.

[6] Q. Xue, X. Fang, and C.-X. Wang, "Beamspace SU-MIMO for future millimeter wave wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1564–1575, Jul. 2017.

[7] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[8] S. B. Ramteke, A. Y. Deshmukh, and K. N. Dekate, "A review on design and analysis of 5G mobile communication MIMO system with OFDM," in *Proc. 2nd Int. Conf. Electron., Commun. Aerosp. Technol. (ICECA)*, Coimbatore, India, 2018, pp. 542–546.

[9] P. Zhou, K. Cheng, X. Han, X. Fang, Y. Fang, R. He, Y. Long, and Y. Liu, "IEEE 802.11ay-based mmWave WLANs: Design challenges and solutions," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 1654–1681, 3rd Quart., 2018.

[10] K. Satyanarayana, M. El-Hajjar, A. A. M. Mourad, and L. Hanzo, "Multi-user hybrid beamforming relying on learning-aided link-adaptation for mmWave systems," *IEEE Access*, vol. 7, pp. 23197–23209, 2019.

[11] Y. Zhao, L. Huang, T.-Y. Chi, S.-Y. Kuo, and Y. Yao, "Capacity analysis for multiple-input multiple-output relay system in a low-rank line-of-sight environment," *IET Commun.*, vol. 6, no. 6, pp. 668–675, 2012.

[12] X. Li, C. He, and J. Zhang, "Spectral efficiency and energy efficiency of distributed antenna systems with virtual cells," *AEU-Int. J. Electron. Commun.*, vol. 96, pp. 130–137, Nov. 2018.

[13] T.-Y. Kim, T. Song, W. Kim, and S. Pack, "Phase-divided MAC protocol for integrated uplink and downlink multiuser MIMO WLANs," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3172–3185, Apr. 2018.

[14] Y. Ghasempour, C. R. C. M. da Silva, C. Cordeiro, and E. W. Knightly, "IEEE 802.11ay: Next-generation 60 GHz communication for 100 Gb/s Wi-Fi," *IEEE Commun. Mag.*, vol. 55, no. 12, pp. 186–192, Dec. 2017.

[15] X. Du, M. Z. Zhang, K. E. Nygard, S. Guizani, and H.-H. Chen, "Self-healing sensor networks with distributed decision making," *Int. J. Sensor Netw.*, vol. 2, nos. 5–6, pp. 289–298, 2007.

[16] H. Shariatmadari, R. Ratasuk, S. Iraji, A. Laya, T. Taleb, R. Jäntti, and A. Ghosh, "Machine-type communications: Current status and future perspectives toward 5G systems," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 10–17, Sep. 2015.

[17] S. Dutta, M. Mezzavilla, R. Ford, M. Zhang, S. Rangan, and M. Zorzi, "MAC layer frame design for millimeter wave cellular system," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, Athens, Greece, 2016, pp. 117–121.

[18] X. Wang, L. Kong, F. Kong, F. Qiu, M. Xia, S. Arnon, and G. Chen, "Millimeter wave communication: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 1616–1653, 3rd Quart., 2018.

[19] 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.

[20] G. Lee, Y. Sung, S. Mumtaz, J. Rodriguez, and L. Dai, "MAC layer design for mmWave massive MIMO," in *mmWave Massive MIMO: A Paradigm for 5G*. London, U.K.: Academic, 2017, pp. 227–255.

[21] Z. Ning, P. Dong, X. Kong, and F. Xia, "A cooperative partial computation offloading scheme for mobile edge computing enabled Internet of Things," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4804–4814, Jun. 2019.

[22] C.-F. Liu, M. Bennis, M. Debbah, and H. V. Poor, "Dynamic task offloading and resource allocation for ultra-reliable low-latency edge computing," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4132–4150, Jun. 2019.

[23] H. Zhang, Q. Zhang, and X. Du, "Toward vehicle-assisted cloud computing for smartphones," *IEEE Trans. Veh. Technol.*, vol. 64, no. 12, pp. 5610–5618, Dec. 2015.

[24] X. Cao, F. Wang, J. Xu, R. Zhang, and S. Cui, "Joint computation and communication cooperation for energy-efficient mobile edge computing," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4188–4200, Jun. 2019.

[25] P. Zhou, X. Fang, X. Wang, Y. Long, R. He, and X. Han, "Deep learning-based beam management and interference coordination in dense mmWave networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 592–603, Jan. 2019.

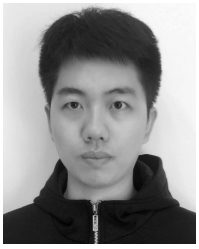
[26] X. Du, Y. Xiao, S. Ci, M. Guizani, and H.-H. Chen, "A routing-driven key management scheme for heterogeneous sensor networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Glasgow, U.K., Jun. 2007, pp. 3407–3412.

[27] X. Ma, F. Yang, S. Liu, J. Song, and Z. Han, "Design and optimization on training sequence for mmWave communications: A new approach for sparse channel estimation in massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1486–1497, Jul. 2017.

[28] L. You, X. Gao, G. Y. Li, X.-G. Xia, and N. Ma, "BDMA for millimeter-wave/terahertz massive MIMO transmission with per-beam synchronization," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1550–1563, Jul. 2017.
[29] B. Zhou, A. Liu, and V. Lau, "Successive localization and beamforming in 5G mmWave MIMO communication systems," *IEEE Trans. Signal Process.*, vol. 67, no. 6, pp. 1620–1635, Mar. 2019.
[30] D. Zhang, Z. Zhou, C. Xu, Y. Zhang, J. Rodriguez, and T. Sato, "Capacity analysis of NOMA with mmWave massive MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1606–1618, Jul. 2017.
[31] H. Zhang, S. Chen, X. Li, H. Ji, and X. Du, "Interference management for heterogeneous networks with spectral efficiency improvement," *IEEE Wireless Commun.*, vol. 22, no. 2, pp. 101–107, Apr. 2015.
[32] B. Satchidanandan, S. Yau, P. R. Kumar, A. Aziz, A. Ekbal, and N. Kundargi, "TrackMAC: An IEEE 802.11ad-compatible beam tracking-based MAC protocol for 5G millimeter-wave local area networks," in *Proc. 10th Int. Conf. Commun. Syst. Netw. (COMSNETS)*, Bengaluru, India, 2018, pp. 182–185.
[33] A. Akhtar and S. C. Ergen, "Directional MAC protocol for IEEE 802.11ad based wireless local area networks," *Ad Hoc Netw.*, vol. 69, pp. 49–64, Feb. 2018.
[34] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
[35] Z.-H. Wu, B. Wang, C. Jiang, and K. J. R. Liu, "Downlink MAC scheduler for 5G communications with spatial focusing effects," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 3968–3980, Jun. 2017.
[36] M. Huang, F. Yuan, H. Cheng, and X. Zhang, "Design of low-latency uplink MAC scheduling for Massive MIMO-OFDM systems," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Paris, France, May 2017, pp. 632–638.
[37] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.
[38] K. N. R. S. V. Prasad, E. Hossain, and V. K. Bhargava, "Machine learning methods for RSS-based user positioning in distributed massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8402–8417, Dec. 2018.
[39] L. X. Cai, H. Shan, W. Zhuang, X. Shen, J. W. Mark, and Z. Wang, "A distributed multi-user MIMO MAC protocol for wireless local area networks," in *Proc. IEEE Global Telecommun. Conf. (IEEE GLOBECOM)*, New Orleans, LO, USA, Nov./Dec. 2008, pp. 1–5.
[40] *Further Advancements for E-UTRA Physical Layer Aspects (Release 9)*, document TS 36.814, 3GPP, Mar. 2010.

YIFENG ZHAO received the B.S. degree in communication engineering, the M.S degree in electronic circuit system, and the Ph.D. degree in communication engineering from Xiamen University, Xiamen, Fujian, China, in 2002, 2005, and 2014, respectively. He is currently an Assistant Professor of communication engineering with Xiamen University. His current research interests include Mmwave communication, massive MIMO, and machine learning applied in wireless communications.

XUETING XU received the B.S. degree from the Nanjing University of Science and Technology, Nanjing, China. She is currently pursuing the master's degree with Xiamen University, China. Her research interests include wireless communications and massive MIMO.



YUHAN SU received the B.S. degree from Huaqiao University, Xiamen, China. He is currently pursuing the Ph.D. degree in communication and information system with Xiamen University, China. His research interests include wireless communications, congestion control, and network coding.

LIANFEN HUANG received the B.S. degree in radio physics and the Ph.D. degree in communication engineering from Xiamen University, Xiamen, China, in 1984 and 2008, respectively. She was a Visiting Scholar with Tsinghua University, Beijing, China, in 1997, and a Visiting Scholar with The Chinese University of Hong Kong, Hong Kong, in 2012. She is currently a Professor of communication engineering with Xiamen University. Her research interests include wireless communications, wireless network, and signal process.

XIAOJIANG DU received the B.S. and M.S. degrees in electrical engineering from Tsinghua University, Beijing, China, in 1996 and 1998, respectively, and the M.S. and Ph.D. degrees in electrical engineering from the University of Maryland, College Park, in 2002 and 2003, respectively. He is currently a Professor with the Department of Computer and Information Sciences, Temple University, Philadelphia, PA, USA. He received over U.S. five million research grants from the U.S. National Science Foundation, the Army Research Office, Air Force, NASA, the State of Pennsylvania, and Amazon. His research interests include security, wireless networks, and systems. He has authored over 320 journals and conference articles in these areas, as well as a book published by Springer. He is a Life Member of the ACM. He received the Best Paper Award at the IEEE GLOBECOM 2014 and the Best Poster Runner-Up Award at the ACM MobiHoc 2014. He served as the Lead Chair of the Communication and Information Security Symposium of the IEEE International Communication Conference 2015 and a Co-Chair of the Mobile and Wireless Networks Track of the IEEE Wireless Communications and Networking Conference 2015.

NADRA GUIZANI received the Ph.D. degree in computer engineering from Purdue University, in 2018. She is currently a Lecturer in computer science with Gonzaga University. Her research interests include machine learning, mobile networking, large data analysis, and prediction techniques. She is an active member in the Eta Kappa Nu Honors Society, Women in Engineering Program, and Computing Research Association for Women.