IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 22, 2019, accepted June 12, 2019, date of publication June 17, 2019, date of current version July 2, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2923599

A New Centralized Access Control Scheme for D2D-Enabled mmWave Networks

DANIELA PANNO, (Member, IEEE), AND SALVATORE RIOLO, (Student Member, IEEE)
Department of Electrical, Electronics and Computer Engineering, University of Catania, Catania 95125, Italy
Consorzio Nazionale Interuniversitario per le Telecomunicazioni (CNIT)

Corresponding author: Salvatore Riolo (salvatore.riolo@unict.it)

This work was supported in part by the Italian Ministry of University and Research "Servify," under Grant PON 03PE_00132_1- CUP E52114000550005 and in part by the Italian Ministry of Economic Development "SUMMIT—Supporto Multipiattaforma per Applicazioni IoT," under Grant F/050270/03/X32 - CUP B68I17000370008.

ABSTRACT In high-density indoor environments (e.g., airports or malls), it is expected an increase in data demand, in the number of applications having stringent Quality-of-Service requirements, and in the number of services among nearby users, such as video sharing or gaming. In order to satisfy these requirements, we adopt D2D communications, 60 GHz band transmissions, and adaptive beamforming techniques. The symbiosis of these technologies provides notable energy saving, high data rate, and strong interference reduction. Despite these benefits, a major issue is the limited communication range, due to the high path loss. To overcome this issue, we consider a Millimeter-Wave Mobile Broadband (MMB) system, consisting of several Access Points interconnected among themselves through a wireless backhaul network. Based on the specific features of the technologies adopted for the considered MMB system, we aim to improve the transmission efficiency, by exploiting concurrent transmissions. Therefore, we propose a new centralized access control scheme which jointly manages D2D communications and transmissions in the access and the backhaul network. Our control strategy aims to maximize the system throughput, to minimize the end-to-end delay, and to improve the fairness among users, while taking into account to keep the computational load as low as possible. To meet these goals, our proposed access control scheme is composed of a data flow management strategy and a multi-criteria scheduling algorithm based on greedy graph vertex-coloring techniques.

INDEX TERMS Concurrent transmission, D2D, graph theory, interference mitigation, mmWave, MMB, scheduling algorithm.

I. INTRODUCTION

Every year the amount of traffic in cellular networks increases because of the massive popularity of smartphones, tablets, smartwatches and new smart mobile devices representing a key feature for our life. In addition, there is an increasing demand for multi-Gbps data rate services and stringent Quality of Service (QoS), in terms of end-to-end delay and packet loss. Because the existing 4G networks will be unable to satisfy the ambitious requirements for future wireless networks, important improvements in cellular network architecture need to be made.

Among the main proposed key technologies for 5G mobile wireless system [1], we focus on Device-to-Device (D2D) communications, millimeter-Wave (mmWave) radio transmissions and adaptive beamforming techniques. D2D communications support direct transmission between two nearby mobile UEs without traversing eNodeBs. A large number of services among nearby users is expected, such as video sharing, gaming, proximity-aware services, popular content downloading. So, enabling D2D communications can increase significantly the overall system capacity and data rates, while reducing eNodeB data traffic, end-to-end latency and power consumption [2]. The main problem is the management of interferences, not only with traditional cellular communications, but also among D2D communications themselves. Since traditional cellular communications typically take place in licensed band, the first problem can be overcome by adopting Outband D2D communications in appropriate unlicensed spectrum bands. At this regards, we observe that the 2.4 GHz and 5.8 GHz unlicensed bands, currently used for massive Wi-Fi and Bluetooth services, are

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

characterized by an uncontrollable interference. Moreover, the 5.8 GHz spectrum can be also adopted for Licensed Assisted Access (LAA) and the extended version (eLAA) introduced by 3GPP in Release 13 and 14, respectively, which are Carrier Aggregations between the licensed LTE carrier and the unlicensed LTE carriers. In this context, we take into account the second key technology: transmission in millimeter-Wave band (i.e., the spectrum between 6 GHz and 300 GHz) [3]. The 60 GHz unlicensed band is largely uncongested compared to the 2.4 GHz and 5.8 GHz bands, and has not yet been investigated by 3GPP Release 16 to be supported by 5G New Radio-Unlicensed (NR-U), whose current target is to extend the applicability of 5G NR to sub 7 GHz unlicensed spectrum bands. So, we choose the 60 GHz band because of the ability to utilize a huge unlicensed bandwidth, so to provide multiple Gbps transmission rates [4]. However, because of the high frequency, 60 GHz transmissions are characterized by strong propagation loss. Although it is a drawback for long-range communications, on the other hand very short-range communications benefit from a reduced interference power of farther interfering transmitters. Finally, in order to further reduce interference from close communication links, we can use high-gain adaptive antenna array supporting directional transmission with narrow beam, exploiting adaptive beamforming techniques. Due to the smaller size of antennas at 60 GHz, in the future it will be feasible to enclose ultra-dense array structures in smartphones [5], so we can use adaptive beamforming techniques in both User Equipments (UEs) and Base Stations.

For the above reasons, the symbiosis of D2D communications, 60 GHz band and directional transmissions is regarded as one of the most promising solution to increase significantly the overall system capacity, by enhancing spatial reuse, and to reduce the end-to-end latency and the power consumption. However, despite these benefits, in environments with high density of UEs, most likely one or more D2D communications will be inside the beamwidth range of another direct communication. Therefore, it is necessary to develop proper radio access control schemes and efficient scheduling algorithms to manage the interference among directional D2D communication, in order to achieve high performance in terms of transmission efficiency, throughput, and end-to-end delay. As reported in [6], an analytical solution will take long computation time that is unacceptable for mmWave cells where the duration of one time slot is only a few microseconds. For this reason, a heuristic approach which does not take long computation time is needed.

A. RELATED WORK AND MOTIVATION

The issues described above have already received attention, and several access control schemes for D2D communications in 60 GHz band are available in literature. 60 GHz standards have been defined for indoor Wireless Personal Area Networks (WPANs), such as IEEE 802.15.3c [7], and for Wireless Local Area Networks (WLANs), such as IEEE 802.11ad [8]. An accurate description and comparative analysis of both protocols can be found in [9]. Since these standards adopt a slotted Time-Division Multiple Access (TDMA), a lot of research works propose a centralized control scheme based on slotted TDMA radio access with concurrent transmission support [6], [10]–[18].

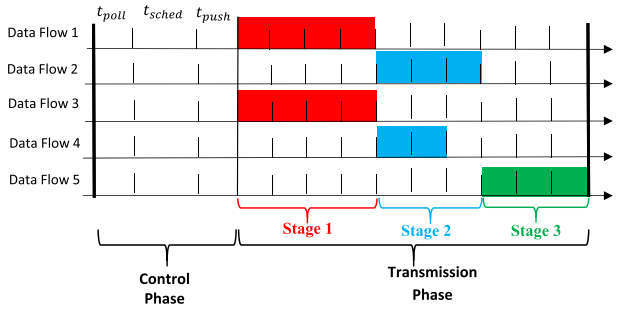

FIGURE 1. Time-line illustration of frame structure and an example of concurrent data flow transmissions.

In [10] and [11] the frame structure is based on the IEEE 802.15.3c, and each frame has a fixed length equal to 1000 time slots. During a time interval called Contention Access Period, each UE sends the transmission request to the PicoNet Controller (PNC) using the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) technology. After that, during a frame time, the controller runs a scheduling algorithm, and then delivers the relative scheduling information (i.e., time slots allocated) to UEs. In [10] concurrent transmission are allowed on the basis of an Exclusive Region (ER), centered on each receiver device, as function of the estimated interfering transmission power, without considering the effective signal-to-interference-plus-noise ratio (SINR). The ER is also adopted by Rehman *et al.* [11]. In order to improve the overall system throughput, they propose a graph-based scheduling algorithm, which favors short-range flows at the expense of a strong unfairness. The main weakness of these approaches is that each UE, once the request has been sent, has to wait for the next frame to start transmissions.

Wang *et al.* [12] the authors adopt a frame structure based on several scheduling periods, each one composed on a beacon period (for network synchronization), a control message exchanging period (in which each transmitter sends its requests and the PNC decides the allocation method for the transmitters), and a data transfer interval. At the aim of maximizing the network throughput per time slot, they present a vertex coloring based resource allocation algorithm. Like [10], [11], they propose to enable concurrent transmission considering the ER and the main lobe interference. In addition, they take into account the interference from the sidelobes, by defining a new power decision threshold in order to reduce this sidelobe interference.

Unlike the approaches analyzed above, the authors of [6] and [13]–[16] adopt a variable-length frame structure, which depends on the user demands. The frame structure is shown in Fig. 1. The central control unit receives transmission requests by means of a polling operation, takes

scheduling decisions during the scheduling time with the aim of minimizing the length of transmission phase, and sends transmission rules to UEs during the pushing time of the same frame. During the transmission phase, slots are organized into one or more stages of variable length, in which non-severe interfering data flows are concurrent transmitted. Despite the frame overhead being increased, in this way UE requests and transmissions occur in the same frame, thus reducing the wait before transmitting. Son *et al.* [13] propose an iterative graph-based scheduling algorithm, in which the mmWave network is considered as "pseudo-wired". For this reason, the concurrent transmission condition takes only into account that a device can transmit or receive to\from only one device at the same time. Starting from the algorithm in [13], Niu *et al.* [14] assume that the transmission rates of each wireless link is well-known and fixed at three discrete values (2, 4, and 6 Gbps) on the basis of the distance between UEs. Their scheduling algorithm allows concurrent transmissions only if the SINR of each link is larger than the minimum SINR to support its transmission rate.

Finally, we focus on some works adopting multi-hop mmWave D2D Communications among UEs. Niu *et al.* [15] propose a scheduling algorithm for popular content downloading in a mmWave small cell. They consider a scenario in which a single Access Point (AP) has to delivery the same packages to all UEs belonging to its coverage area. The AP, rather than delivering packets to each UE, sends packets only to nearby UEs; then, they will forward packets to the other UEs by means of multi-hop D2D communications. In the same context, Giatsoglou *et al.* [17] propose to increase popular content exchanges between the paired UEs by using a new policy for device caching. Wei *et al.* [18] investigate the relay transmission through other UEs if an entire light-of-sight (LOS) path is available, when the direct mmWave links are subject to random Bernoulli blockages.

Let us note that all above works consider a scenario limited to the coverage area of a single PNC or an AP, whose coverage radius is typically equal to 15*m*. Instead, Niu *et al.* [6] propose a centralized control scheme for a wider indoor scenario with a high density of UEs. In order to extend the advantages of 60 GHz D2D communications to the whole environment, they assume that several Access Points (APs) are installed inside the coverage area of a Macro eNodeB and interconnected among themselves through a high speed wireless connection at 60 GHz, as shown in Fig. 2. This solution implies that the access and the backhaul network use the same resources, therefore resource allocation management is more difficult than a single PNC/AP scenario because it should take into account also the interference between the access and the backhaul network. At the aim of maximizing the system efficiency by having the ability to allocate all available resources without restriction, they propose a central controller which considers both the access and the backhaul network jointly. The authors consider the Central Controller (CC) located inside one AP, so in this way, the proposed architecture acts as a stand-alone Millimeter-Wave Mobile Broadband (MMB)

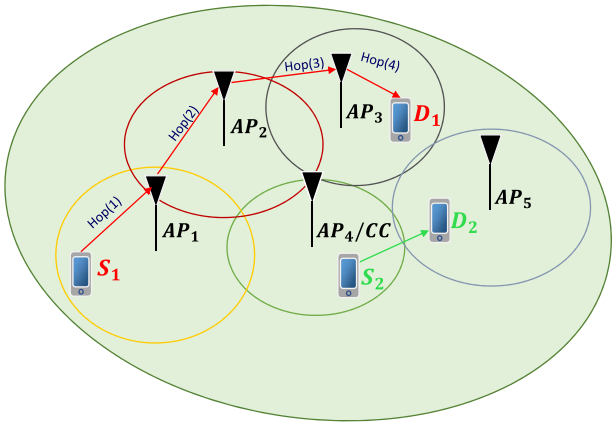

FIGURE 2. An MMB system for a wide indoor scenario with a high density of UEs. The communication between each pair of UEs inside the whole scenario can take place without traversing the core network via a multi-hop path (e.g., communication between Source S_1 and Destination D_1) or directly (e.g., communication between S_2 and D_2). The Central Controller (CC) is inside AP_4.

system [19], that is, a 5G system operating exclusively on mmWave bands. The controller also acts as a gateway, because it is connected to the Internet via a direct and high speed wired connection. The remaining APs in the scenario communicate with the gateway to send (receive) data to (from) Internet. The communication between each pair of UEs inside this scenario can take place without traversing the core network via a multi-hop path (through one or more APs) or directly (traditional D2D communication), as shown in Fig. 2. On the basis of their previous work [14], Niu *et al.* [6] present an improved access scheme termed D2DMAC, which consists of a path selection criterion to determine if a communication takes place through a multi-hop path or directly, and a concurrent transmission scheduling algorithm, which supports multi-hop communications. The basic strategy adopted in their scheduling algorithm is to deliver all packets in a frame time, whether delivery occurs directly or via a multi-hop path. Gao *et al.* [16] propose an improved D2DMAC scheme, termed directional D2D medium access control (D3MAC). Compared to the D2DMAC scheme, Gao *et al.* [16] introduce a contention graph in order to depict the contention relationship between flows. It is constructed in the way that two vertices are connected with an edge only if the maximum interference between them is greater than or equal to a fixed threshold value. On the basis of this graph, the heuristic transmission scheduling algorithm of D3MAC, which is the same of [6], reduces the number of iterations.

Although D3MAC scheme has addressed the multi-hop D2D transmission problem in a wide scenario, it shows two significant weaknesses. First, in the presence of one or more multi-hop paths, the packet delivery in a single frame may increase the frame length, so much that the next polling time could occur after several milliseconds. Consequently, this causes delays for new transmission requests, that could be too long for delay-sensitive services. Second, the D3MAC

scheme is not very efficient because it does not fully exploit the opportunities of concurrent transmission as we will show later on. In order to overcome the above problems, we proposed a preliminary work in a conference paper [20], and this work mainly extends it. More specifically, the path type selection criterion has been greatly improved, and all criteria of the scheduling algorithm have been revised and significantly extended to take into account the fairness among data-flows and to enhance, as much as possible, concurrent transmissions.

B. CONTRIBUTIONS

In this paper, we investigate the access control problem in an indoor D2D-enabled stand-alone MMB system with a high density of UEs for the purpose of interference mitigation. We aim to enhance the transmission efficiency, by exploiting concurrent transmissions, to maximize the throughput, to minimize the end-to-end delay, and to improve the fairness among users, while taking into account to keep the computational load as low as possible. To this end, we propose a new centralized access control scheme composed of a data flow management strategy and a multi-criteria scheduling algorithm based on a greedy graph vertex-coloring technique. Like D3MAC, our access control scheme is based on the variable-length frame structure of Fig. 1 and jointly manages D2D communications and transmissions in the access and the backhaul network. In addition to the D3MAC, our scheduling algorithm takes also into account the source rate and the service delay requirements.

The main contributions can be summarized as follows:

- As regards multi-hop transmissions, unlike D3MAC, our resource management strategy is based on performing only one hop per frame, with the aim of maximizing the number of possible concurrent transmissions, thus improving the transmission efficiency and the delay-sensitive service performance.
- For each transmission request, the Central Controller chooses the proper path type (i.e., direct or multi-hop path) through a new path selection criterion, which takes into account also the source rate and the performance differences between the direct path and the multi-hop path, by means of new distance-based metrics.
- We define a new multi-criteria scheduling algorithm, in order to reach the fairness, enhance transmission efficiency, maximize the system throughput, and reduce the end-to-end delay. These criteria are grouped in three phases. First, analysis of the interference relations among data transmissions, and creation of an interference graph. Second, assignment of different transmission stages to intolerant interfering sub-flows, by using a new graph multi-coloring method. Third, a new criterion for minimizing the number of time slots to be allocated to each stage.
- To further enhance transmission efficiency, the controller splits, wherever necessary, a data flow transmission in more transmission stages inside the same frame without increasing the number of stages; this minimizes the frame length for the same amount of data traffic transmitted.
- By simulation, we benchmark our approach against the D3MAC scheme [16]. The comparative analysis shows that our control scheme outperforms the reference scheme in terms of throughput, end-to-end delay and fairness in any simulated traffic condition.

The remainder of this paper is organized as follows. In Section II the system overview is given. Section III presents the proposed access control scheme, in which the flow control strategies are described. The transmission scheduling problem is formulated in Section IV. Then, the heuristic scheduling algorithm used in our access control scheme is explained in Section V. Section VI shows the performance analysis. Finally, the concluding remarks are given in Section VII.

II. SYSTEM OVERVIEW

We consider a wide indoor D2D-enabled MMB system, operating at 60 GHz, inside the coverage area of a M-eNB with a high density of UEs. The MMB system consists of N_{AP} APs and N_U UEs operating on the same spectrum. We assume that UEs are equipped with multi-standard technology transceivers (e.g., mmWave and LTE Advanced). So, a UE, while is connected at a low frequency with the macro eNodeB for traditional services, can communicate at 60 GHz with a device (UE or AP) inside the same MMB system through a direct communication or a multi-hop path across APs without traversing the Core Network.[1]

On the basis of the described scenario and the adopted architectural solution, in addition to the interference between D2D communications, the interference between mmWave access and backhaul network arises. So, in order to efficiently manage these interferences, as in [16], we choose a centralized control approach, considering both radio access and backhaul networks jointly. More specifically, we assume a local Central Controller (CC), inside an AP, which is the one which reduces the average one-way latency between each pair of APs. Other assumptions are listed below:

1) APs are in LOS between them and their position is fixed.
2) The Central Controller knows the static topology of APs and the up-to-date location of UEs. The latter information can be obtained by a maximum-likelihood classifier based on beam-space channel matrix [21].
3) Each UE and AP is equipped with ideal directional adaptive antennas [10] with the same beam angle: the antenna gain is constant within the beam angle θ ($G = \frac{2\pi}{\theta}$) and zero outside; the transmitter antenna is able to automatically adapt its radiation pattern, by putting the main lobe in the direction of the receiver antenna.

[1]The proposed control scheme could also work with traditional communications to the Internet, by considering an AP gateway as the destination device.

4) Each UE can be served by only one Access Point and is associated with the nearest AP.

A. BASICS OF THE RADIO ACCESS SCHEME

As in [13], [14] and [16], we adopt a slotted TDMA with concurrent transmission support. A fundamental constraint of TDMA approach is the high synchronization accuracy needed in the system. The clock of all APs is synchronized by the controller; then, each AP synchronizes the clock of UEs associated with it. However, to ensure time synchronization taking into account propagation delays, a radio network planning is needed. We have addressed this issue in another work [22].

In the considered scenario, where the beamforming technology is a key point, the management of UE request packets is not based on the CSMA/CA standard, but on the polling technique, because the conventional carrier sensing medium contention schemes do not work well with directional antennas, as outlined in [23].

Time is divided into time slots of equal length and organized in non-overlapping frames with variable length [13]. As shown in Fig. 1, a frame consists of two phases: control phase and transmission phase. The first phase contains three time intervals: polling time, scheduling time and pushing time.

During the polling time (t_{poll}), each AP sequentially polls all UEs associated with it to check whether any UE has data to transmit. Each UE must respond after a fixed interval (Short Inter Frame Space, SIFS) with a request packet or with a 'keep alive' response message if it does not have any data to transmit. Once this procedure is completed, each AP will report the information received from UEs to the Central Controller through the wireless backhaul network.

During the scheduling time (t_{sched}), the Central Controller applies an efficient scheduling algorithm based on received requests and location information, to meet the traffic demands.

Next, during the pushing time (t_{push}), the CC distributes scheduling rules to all APs through the wireless backhaul network. Afterwards, each AP forwards the scheduling information to UEs associated with it. Upon receipt of the information from its AP, each UE sends an acknowledgment message. The time required for the pushing operation is the same as the polling.

Finally, during the transmission phase, slots are organized into one or more stages of variable length. During a stage, non-severe interfering flows are concurrently transmitted. Let us note that each wireless device can not transmit and receive at the same time, as described in Section IV.

We set slot time $t_{SLOT} = 5\mu s$ and its payload time $t_{PAY} = 4\mu s$, as in [16]. In addition, in [22] we derived the maximum distance for mmWave communication in non-line-of-sight (NLOS) conditions, that is, $d_{\max,\mathrm{NLOS}} = 16.97\ m$. This choice guarantees all overhead parameters and a proper guard interval, to prevent the use of compensation techniques for time slot synchronization (e.g., timing advance).

B. 60 GHZ CHANNEL MODEL

The main characteristics of 60 GHz communications are high frequency and large bandwidth, so high path loss and transmission rate values. We adopt the path loss model for frequency spectrum above 6 GHz in Indoor-office scenarios provided by the 3GPP standard in [24]. The path loss in dB at distance d can be expressed as follows:

$$PL_{[dB]}(d) = PL_{[dB]}(d_0) + 10\alpha \log_{10}\left(\frac{d}{d_0}\right) + X_{\sigma_{[dB]}}, \quad (1)$$

where $PL_{[dB]}(d_0) = -10\log_{10}\left[\left(\frac{\lambda}{4\pi d_0}\right)^2 G_T G_R\right]$, λ is the wavelength, G_T and G_R are transmitter and receiver antenna gain, α is the path loss exponent, d_0 is the reference distance, d is the distance between transmitter and receiver, $X_{\sigma_{[dB]}}$ represents the large-scale fading, modeled as a log-normal distribution which in the dB-domain corresponds to a zero-mean Gaussian distribution with standard deviation $\sigma_{[dB]}$. The values of α and $\sigma_{[dB]}$ depend on the visibility condition, i.e., Line-Of-Sight (LOS) or Non-Line-Of-Sight (NLOS), and are reported in Table 1. The LOS state is considered when no buildings or walls block the direct path between TX and RX. The impact of objects, such as chairs, desks, office furniture, and so on, is modeled using the shadowing term $X_{\sigma_{[dB]}}$ [25].

In order to estimate the maximum achievable transmission rate of data flow i from transmitter TX_i to receiver RX_i, we adopt Shannon's theoretical formula,[2] as in [11]:

$$R_i = \eta W \log_2\left(1 + \mathrm{SINR}_i\right), \quad (2)$$

where η represents the transceiver efficiency ($\eta \in [0, 1]$), W is the system bandwidth, and SINR_i is the signal to interference plus noise ratio, calculated as:

$$\mathrm{SINR}_i = \frac{P_{T_i}/PL(d(TX_i, RX_i))}{N_0 W + \sum_{j\neq i} P_{T_j}/PL(d(TX_j, RX_i))}, \quad (3)$$

where P_{T_i} is the signal power of TX_i, PL is the path loss in (1) converted from the dB-domain to the linear domain, $d(TX_i, RX_i)$ is the distance between TX_i and RX_i, and N_0 is the power spectral density of the white Gaussian noise.

Let us note that $\sum_{j\neq i} P_{T_j}/PL(d(TX_j, RX_i))$ is the power from all interfering data flows.

III. OUR ACCESS CONTROL STRATEGY

For the mmWave network architecture and the basic radio access scheme introduced above, we propose a new resource management strategy aiming to fulfill the following objectives:

- enhance transmission efficiency, taking fully advantage of concurrent transmissions;
- maximize throughput;
- minimize end-to-end delay;
- keep computational complexity as low as possible.

[2] A future implementation of a proper modulation and code scheme will set accurate transmission rate values. However, the framework for our analysis would remain identical.

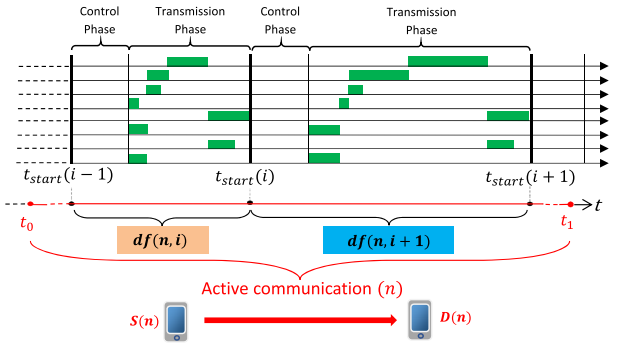

FIGURE 3. Graphic representation of active communication(n) and the related data flows.

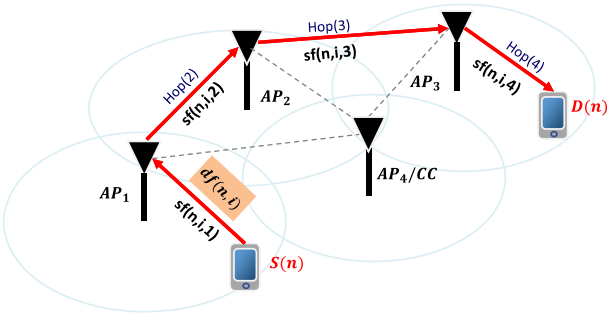

FIGURE 4. Graphic representation of data flow(n, i) and the related sub-flows.

These targets are often in opposition to each other. On the one hand, we try to offer best efficiency, enabling as many as possible concurrent transmissions. On the other hand, we try to maximize transmission rates, by minimizing the interference (i.e., by reducing the amount of concurrent transmissions). In addition, we need to establish the best order of transmission to minimize the end-to-end delay.

In this paper, several aspects are considered and several criteria are developed in order to propose a new Access Control scheme which aims to best accomplish each target. In the next subsection, before defining the Central Controller operations, the approach taken for the management of multi-hop flows is described and the motivations are explained.

A. MULTI-HOP TRANSMISSION MANAGEMENT

In Fig. 3 we show a data session between source UE, $S(n)$, and destination UE, $D(n)$, termed "active communication(n)"; it starts at time t_0 and ends at time t_1. Also, we denote with $t_{start}(i)$ the beginning time of frame i, and with "data flow(n, i)", $df(n, i)$, the amount of traffic data generated by source $S(n)$ during frame $(i-1)$, i.e., from $t_{start}(i-1)$ to $t_{start}(i)$. During the polling time of the ith frame, $S(n)$ will send to the related AP a request packet, containing, inter alia, $D(n)$ address and the amount of traffic data to be transmitted, $df(n, i)$. On the basis of the decisions made by the controller, $df(n, i)$ can be transmitted directly (direct path) or via Access Points (multi-hop path). In the latter case, $df(n, i)$ needs to be transmitted sequentially, hop-by-hop, as depicted in Fig. 4. We define "sub-flow(n, i, j)", $sf(n, i, j)$, as the transmission of $df(n, i)$ related to hop (j). The transmission of a direct data flow is a single sub-flow(n, i, 1).

As regards multi-hop transmissions, two different strategies can be considered:

1) complete data flow transmission from source to destination UE, during a single frame;
2) perform only one hop per frame.

D3MAC scheme adopts strategy 1, i.e., each AP stores the received $df(n, i)$ and forwards it in the next stage of the same frame (see Fig. 5a). Instead, with strategy 2, each AP stores the received $df(n, i)$ and forwards it in the next frame. The choice of strategy is relevant because significantly affects the system performance, as we show below.

Firstly, we consider a simple example of an active communication consisting of a single data flow, which is delivered through 4 hops (sub-flows). With strategy 1, the total overhead time is related only to the control phase of a single frame. Instead, with the second strategy, because only one sub-flow can be delivered during a frame, the total overhead time is related to four control phases. Therefore, if we restrict our analysis to just one multi-hop data flow, the best strategy seems to be the first one.

However, it is very likely that during the transmission phase, the source UE of the considered active communication continues to generate packets. Because sub-flows belonging to the same data flow are sequential transmissions, as shown in Fig. 5a, the frame length can be long. The longer the frame length, the greater the amount of traffic generated by source UE, i.e., the data flow size. The process of lengthening continues until the length of the frame reaches the steady state with a long duration. Conversely, with strategy 2, because in each frame all sub-flows belong to different data flows of the same active communication, some of these sub-flows can be scheduled concurrently, as shown in Fig. 5b. The frame length will be shorter compared to strategy 1 and, consequently, the traffic amount generated during a frame is reduced. As a result, an active communication is expected to be split into more data flows with a small number of packets that can exploit concurrent transmissions.

Consequently, with strategy 2, the time necessary to transmit the same amount of data is less than with the first one (i.e., the efficiency increases and the packet delay is reduced). In addition, we note that the control phase frequency is greater, so other new source UEs can benefit from a reduced wait for the next polling time (i.e., a reduced accumulated initial delay and an improved fairness among UEs). For these reasons, unlike D3MAC [16], in our control access scheme we adopt strategy 2.

B. CENTRAL CONTROLLER OPERATIONS

A high-level block diagram of the access control strategy is shown in Fig. 6.

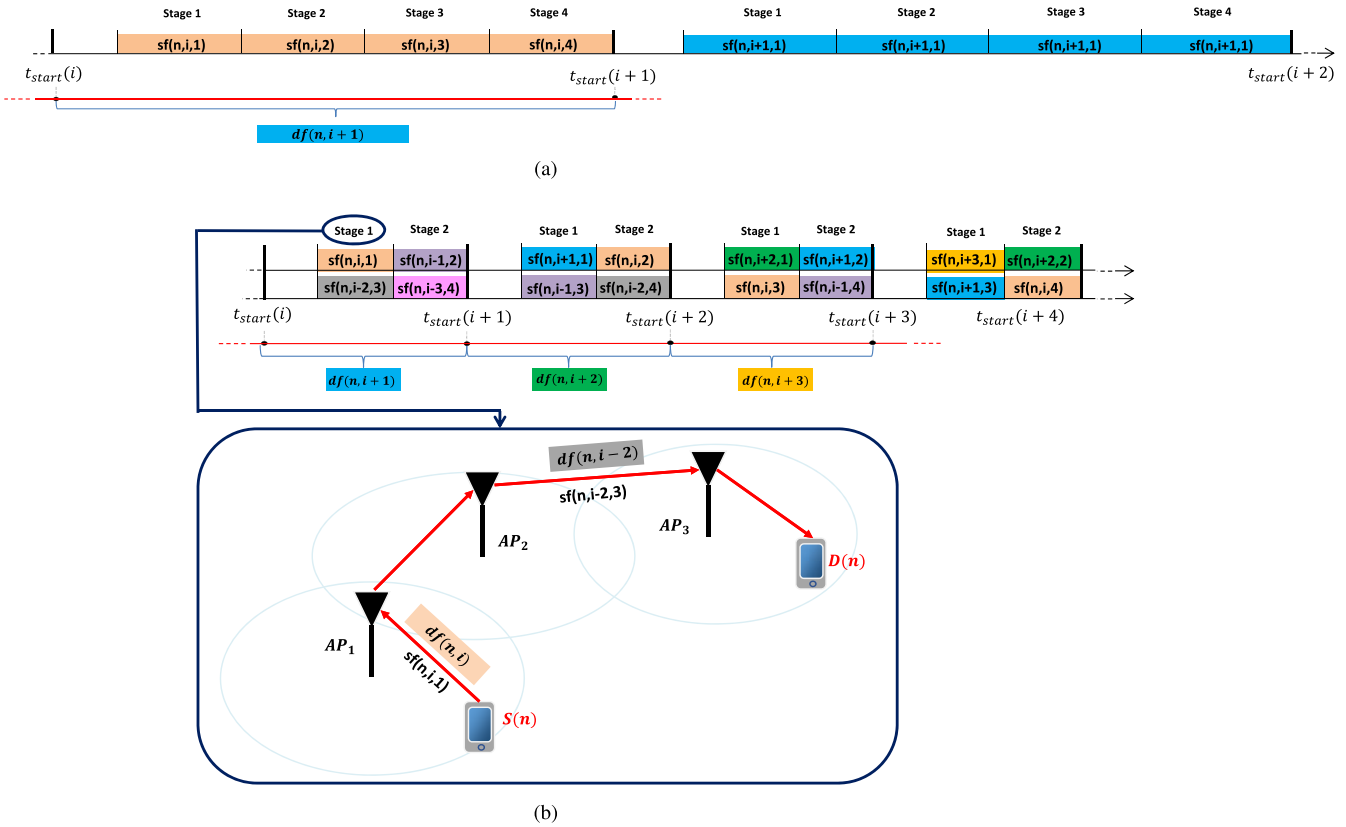

FIGURE 5. Active communication consisting of multi 4-hop data flows. Transmission in steady state with two strategies. (a) Strategy 1. (b) Strategy 2.

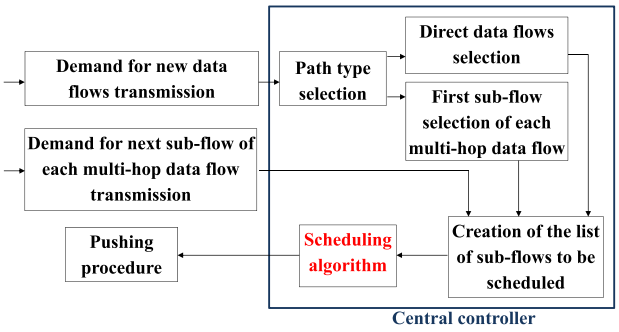

FIGURE 6. High-level block diagram of the access control strategy.

During the polling time, each AP obtains request packets from its UEs with new data flows to be transmitted, characterized by: destination UE; amount of data in terms of number of packets ($n_{packets}$), assuming a fixed packet size of $N_{\text{bytes,packet}}$ bytes; source rate (R_{source}), e.g., 2 or 4 Gbps; end-to-end delay threshold (d_{thr}), if any. Then, each AP will report to the AP/CC the requests for transmission of new data flows. In addition, each AP, which has received and stored a data flow in the previous frame, will send to the Central Controller a transmission request to forward it to the next hop. As regards active communications with delay-sensitive services, we assume that each AP discards all related packets which have accumulated a delay larger than the delay threshold, before sending its requests to the Central Controller.

Therefore, at the end of each polling time, the Central Controller knows all UEs and APs requests. For all new data flows, the controller selects the path type: it establishes whether data transmission occurs by a direct path (direct sub-flow) or a multi-hop path (more sub-flows). The multi-hop path is determined by a static routing. If the controller chooses the multi-hop path, then it selects only the first sub-flow for the current scheduling procedure. The path selection criterion is described in the next subsection.

Finally, the list of sub-flows to be scheduled is created. It is the input of the scheduling algorithm described in Section V. This list contains all direct sub-flows (single hop), the first sub-flow related to each new multi-hop data flow, and the next sub-flow related to each multi-hop data flow stored in APs.

C. PATH TYPE SELECTION

The objective of the path selection criterion is to choose the proper path type (i.e., direct or multi-hop path) for new data flows. Here, we propose an improved criterion in comparison with [20]. In addition to the constraint on the maximum communication distance in NLOS conditions ($d_{\text{max,NLOS}}$), which ensures the time slot synchronization, the new criterion takes also into account R_{source} and the performance differences between the direct path and the multi-hop path in terms of

achieved transmission rate. In order to meet the requirement of low computational complexity for this path type selection algorithm, we introduce new parameters based only on the distance metric.

For each new data flow(n, i), by using (2), we derive the minimum SINR value supporting the source rate $R_{\text{source},n}$, as follows:

$$\text{SINR}_{\text{min},n} = 2^{\frac{R_{\text{source},n}}{\eta W}} - 1. \tag{4}$$

So, we define $d_{\text{max,source},n}$ as the maximum distance between $S(n)$ and $D(n)$ supporting $R_{\text{source},n}$ in absence of interference. By using (3) in which $\sum_{j\neq i} P_{T_j}/PL(d(TX_j, RX_i)) = 0$ and $\text{SINR}_i = \text{SINR}_{\text{min},n}$, we obtain $d_{\text{max,source},n}$ as follows:

$$d_{\text{max,source},n} = d_0 \left(\frac{P_{T_n}/PL(d_0)}{\text{SINR}_{\text{min},n} N_0 W} \right)^{\frac{1}{\alpha}}, \tag{5}$$

where P_{T_n} is the transmission power of $S(n)$.

Also, we aim to compare the expected performance of the direct path against the one of the multi-hop path. We observe that, due to the direct visibility among APs, the performance of a multi-hop data flow is main limited by the first and the last hop, which are generally in NLOS. Between these two, it is likely that the one at greater distance has less favorable characteristics. For this reason, we define a multi-hop distance value as:

$$d_{\text{mh},n,i} = \max\left[d\left(S(n), AP_{S(n)}\right), d\left(AP_{D(n)}, D(n)\right) \right], \tag{6}$$

where $AP_{S(n)}$ and $AP_{D(n)}$ represent the Access Points to which $S(n)$ and $D(n)$ are attached, respectively. Finally, in order to consider the performance of the whole multi-hop path, we introduce a weight equal to the number of hops ($N_{\text{hop},n,i}$).

On the basis of $d_{\text{max,NLOS}}$ and the new parameters (5) and (6), we define a new path type selection algorithm, shown in pseudo-code 1. Therein, the distance of the direct path, $d\,(S(n), D(n))$, is compared with $d_{\text{max,NLOS}}$ to ensure the time slot synchronization, and with (5) and (6) to guarantee $R_{\text{source},n}$, as far as possible.

Pseudo-code 1 Path Type Selection

Definitions:

- F_N: set of new request for transmission (from UEs)
- $d_{\text{max,NLOS}}$: maximum NLOS distance
- for each data flow:
  - $d\,(S(n), D(n))$: distance from $S(n)$ to $D(n)$ of active communication(n)
  - $d_{\text{max,source},n}$: maximum distance supporting the source rate
  - $d_{\text{mh},n,i}$: multi-hop distance
  - $N_{\text{hop},n,i}$: number of hops of multi-hop path

Iteration:

```
for each element ∈ F_N do
  if d (S(n), D(n)) > d_max,NLOS then
    choose multi-hop path {time slot synchronization is not guaranteed with the direct path}
  else
    if d (S(n), D(n)) > d_max,source,n then
      if d (S(n), D(n)) < d_mh,n,i then
        choose direct path {source rate is not guaranteed with the direct path, but the multi-hop path has worse performance}
      else
        choose multi-hop path
      end if
    else
      if d (S(n), D(n)) > N_hop,n,i d_mh,n,i then
        choose multi-hop path {despite source rate being guaranteed with the direct path, the weighted multi-hop path still performs better}
      else
        choose direct path
      end if
    end if
  end if
end for
```

IV. TRANSMISSION SCHEDULING PROBLEM FORMULATION

In this Section, we present the problem formulation of the optimal scheduling decisions. The scheduling function has a key role in the Central Controller. As depicted in Fig. 6, a list of sub-flows to be scheduled in the current frame is the input of the scheduling algorithm. For each sub-flow i related to active communication(n), the Central Controller knows the related hop (i.e., transmitting device TX_i and receiving device RX_i), the amount of packets to be transmitted ($n_{\text{packets},i}$), and the source rate $R_{\text{source},n}$.

We consider N_{SF} sub-flows to be scheduled in the transmission phase of a single frame. As reported in Section II-A, the transmission phase is divided into N_{stage} non-overlapping stages, and the kth stage lasts N_{ts_k} time slots. In each stage, multiple sub-flows could be scheduled to transmit concurrently. Given the traffic demand, in order to maximize the transmission efficiency, the optimal scheduling decisions should accommodate the traffic demand ($n_{\text{packets},i}$) of each sub-flow i with the minimum number of time slots. This objective can be formulated as follows:

$$\min \sum_{k=1}^{N_{\text{stage}}} N_{\text{ts}_k}. \tag{7}$$

Let us underline that N_{stage} and $N_{\text{ts},k}$ are both unknowns of the problem. $N_{\text{stage}} \in \{1, \ldots, N_{SF}\}$, i.e., $N_{\text{stage}} = 1$ when all sub-flows are transmitted concurrently, and $N_{\text{stage}} = N_{SF}$ when no sub-flow is transmitted concurrently. $N_{\text{ts}_k} \in \left\{1, \ldots, \max_{\forall i}\left[n_{\text{packets},i}\right]\right\}$, because in one time slot each sub-flow can transmit one or more packets, based on its transmission rate.

Now, we analyze the system constraints. Let s_i^k be a boolean variable which indicates whether sub-flow i is scheduled in stage k. If it is, $s_i^k = 1$; otherwise $s_i^k = 0$.

Because transmitting and receiving antennas of a wireless device (UE or AP) operate at the same carrier frequency, each wireless device can only transmit or receive at the same time from at most one device. This constraint can be formulated as

follows.

$$s_i^k + s_j^k \leq 1, \quad \forall k,\ \forall i \neq j \text{ if } RX_i = TX_j; \tag{8}$$

$$s_i^k + s_j^k \leq 1, \quad \forall k,\ \forall i \neq j \text{ if } RX_i = RX_j; \tag{9}$$

$$s_i^k + s_j^k \leq 1, \quad \forall k,\ \forall i \neq j \text{ if } TX_i = TX_j. \tag{10}$$

To exploit concurrent transmissions, the signal to interference plus noise ratio experienced by each flow i in stage k (SINR_i^k) should be at least equal to $\text{SINR}_{\text{min},i}$, which is the minimum SINR value supporting the source rate $R_{\text{source},n}$ of the related active communication(n). This constraint is formulated as follows:

$$\text{SINR}_i^k \geq \text{SINR}_{\text{min},i}, \quad \forall i, k \text{ so that } s_i^k = 1, \tag{11}$$

where

$$\text{SINR}_i^k = \frac{s_i^k P_{T_i} / PL(d(TX_i, RX_i))}{N_0 W + \sum_{j \neq i} s_j^k b_i^j P_{T_j} / PL(d(TX_j, RX_i))}, \tag{12}$$

in which $b_i^j = 1$ if RX_i is inside the beamwidth range of TX_j; otherwise, $b_i^j = 0$.

The traffic demand ($n_{\text{packets},i}$) of each sub-flow i must be fully accommodated in the current transmission phase. It means that:

$$\sum_{k=1}^{N_{\text{stage}}} \left\lfloor \frac{s_i^k N_{\text{ts}_k} t_{PAY} R_i^k}{8 N_{\text{bytes,packet}}} \right\rfloor \geq n_{\text{packets},i}, \forall i, \tag{13}$$

where:

$$R_i^k = \eta W \log_2 \left(1 + \text{SINR}_i^k\right), \forall i, k. \tag{14}$$

As regards constraint (13), we need to make the following considerations. Let us note that Gao *et al.* [16] assume that each sub-flow is activated just once in the transmission phase, that is,

$$\sum_{k=1}^{N_{\text{stage}}} s_i^k = 1, \forall i. \tag{15}$$

By adopting this assumption, as depicted in Fig. 7a, each stage k ends only when each sub-flow i, so that $s_i^k = 1$, has transmitted all of its packets, e.g., in the figure N_{ts_2} ends when sub-flow 5 is completed. This means that, to satisfy constraint (13), for each stage k the N_{ts_k} value is established by the sub-flow requesting more time slots in it (i.e., the sub-flow with a large amount of data and/or a low transmission rate).

Unlike [16], in this work, to better achieve objective (7), we propose to split, wherever necessary and possible, a sub-flow into more transmission stages. For instance, if sub-flows 1 and 5 can be transmitted concurrently in stage 3 (i.e., if $\text{SINR}_1^3 \geq \text{SINR}_{\text{min},1}$ and $\text{SINR}_5^3 \geq \text{SINR}_{\text{min},5}$), then N_{ts_2} can be properly reduced by enhancing concurrent transmissions, as shown in Fig. 7b. This assumption can be formulated as follows.

$$\sum_{k=1}^{N_{\text{stage}}} s_i^k \leq N_{\text{stage}}, \forall i. \tag{16}$$

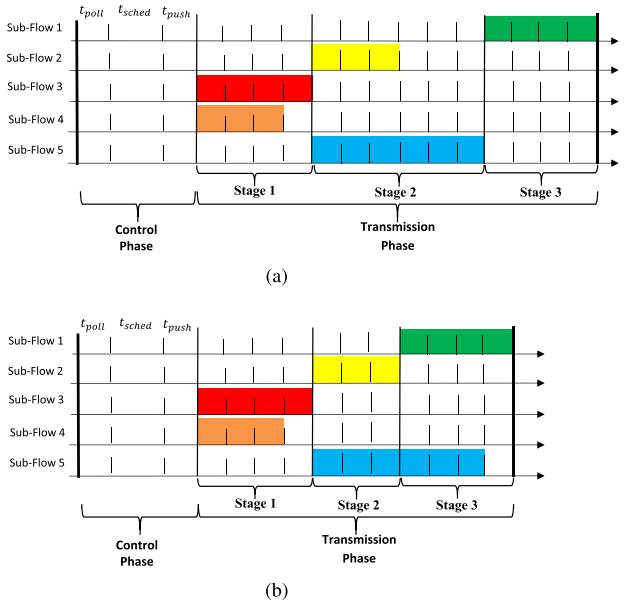

FIGURE 7. Benefit to transmit sub-flows into more stages. (a) Only one transmission stage per sub-flow. (b) More transmission stages per sub-flow.

It is clear that assumption (16) favors a more efficient transmission than assumption (15), at the expense of a greater computational complexity. In fact, considering assumption (15), constraint (13) corresponds to N_{SF} inequalities each one in a single unknown, while considering assumption (16), constraint (13) can be written as a system of N_{SF} inequalities each one in N_{stage} unknowns:

$$\begin{cases} \left\lfloor \dfrac{s_1^1 N_{\text{ts}_1} t_{PAY} R_1^1 + \cdots + s_1^{N_{\text{stage}}} N_{\text{ts}_{N_{\text{stage}}}} t_{PAY} R_1^{N_{\text{stage}}}}{8 N_{\text{bytes,packet}}} \right\rfloor \geq n_{\text{packets},1} \\ \left\lfloor \dfrac{s_2^1 N_{\text{ts}_1} t_{PAY} R_2^1 + \cdots + s_2^{N_{\text{stage}}} N_{\text{ts}_{N_{\text{stage}}}} t_{PAY} R_2^{N_{\text{stage}}}}{8 N_{\text{bytes,packet}}} \right\rfloor \geq n_{\text{packets},2} \\ \cdots \\ \left\lfloor \dfrac{s_{N_{SF}}^1 N_{\text{ts}_1} t_{PAY} R_{N_{SF}}^1 + \cdots + s_{N_{SF}}^{N_{\text{stage}}} N_{\text{ts}_{N_{\text{stage}}}} t_{PAY} R_{N_{SF}}^{N_{\text{stage}}}}{8 N_{\text{bytes,packet}}} \right\rfloor \geq n_{\text{packets},N_{SF}}. \end{cases} \tag{17}$$

In addition, this system in general has not a unique solution, because $N_{\text{stage}} \leq N_{SF}$. Therefore, it is evident that the degrees of freedom of our problem formulation are greatly increased compared to [16] as well as the computational complexity of the problem.

In summary, the objective of the optimal scheduling is (7) s.t. constraints (8)-(11), and (13). It is easy to observe that the formulated problem is a Mixed Integer Nonlinear Program (MINLP), which is generally NP-hard. Using the branch-and-bound algorithm, it will take significantly long computation time [13], and it is unacceptable for practical mmWave cells where the duration of one time slot is only a few microseconds. For this reason, in the next Section we

propose a multi-criteria heuristic scheduling algorithm to obtain near-optimal solutions which will not take significantly long computation time.

V. OUR PROPOSED SCHEDULING ALGORITHM

In this section, we propose a greedy algorithm that breaks up the global scheduling problem into a series of non-overlapping sub-problems, to make the locally optimal choice at each phase, with the intent of finding a global optimum. Our multi-criteria scheduling algorithm consists of three phase, as shown in Fig. 8, and in Lines 11-19 of pseudo-code 2, which summaries the overall access control procedure. For each phase, the details are described in the following subsection.

Pseudo-code 2 Overall Access Control Procedure

Definitions:

- $\boldsymbol{F_N}$: set of new requests for transmission (from UEs)
- $\boldsymbol{F_O}$: set of requests for next sub-flow of multi-hop data flow transmission (from APs)

Initialization:

1: $\boldsymbol{F_D} = \emptyset, \boldsymbol{F_{MH}} = \emptyset, \boldsymbol{L} = \emptyset$
2: **for** each data flow $\in \boldsymbol{F_N}$ **do**
3: select the path type by following pseudo-code 1.
4: **if** direct path has been selected **then**
5: Insert it in vector $\boldsymbol{F_D}$.
6: **else**
7: Insert it in vector $\boldsymbol{F_{MH}}$.
8: **end if**
9: **end for**
10: Insert each direct data flow $\in \boldsymbol{F_D}$, the first sub-flow of each multi-hop data flow $\in \boldsymbol{F_{MH}}$, and the sub-flows $\in \boldsymbol{F_O}$ in List $\boldsymbol{L}$.
11: **for** each subflow $\in \boldsymbol{L}$ **do**
12: Create lists $\boldsymbol{L_{1,i}}$, $\boldsymbol{L_{2,i}}$, and $\boldsymbol{L_{3,i}}$ by following the criteria of Step 1.1.
13: **end for**
14: Create an undirected interference graph, G(V,E), by following Step 1.2.
15: Color the graph G(V,E) by following pseudo-code 3.
16: Apply the multi-color criterion to graph G(V,E) by following Step 2.2.
17: Establish Color Transmission Order by following Step 2.3.
18: Calculate the initial stage lengths by following pseudo-code 4.
19: Extend the stage lengths, if necessary, by following pseudo-code 5.
20: The scheduling rules are pushed to all APs.

A. PHASE 1

The objective of this phase is to identify the critical interference relations among all sub-flows in input.

Step 1.1. The first step is to get for each sub-flow two final lists: the list of intolerable interfering sub-flows and the list of tolerable ones. At this aim for the ith sub-flow, firstly the Central Controller derives $\boldsymbol{L_{1,i}}$, $\boldsymbol{L_{2,i}}$ and $\boldsymbol{L_{3,i}}$ as the lists of

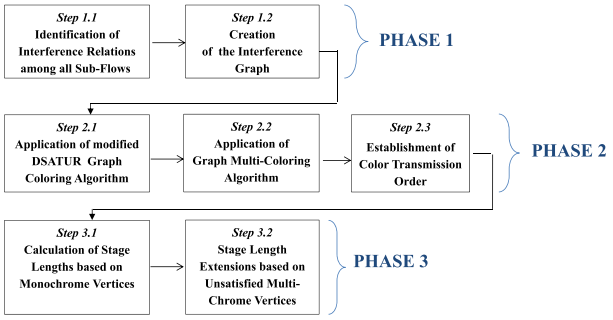

FIGURE 8. Multi-criteria scheduling algorithm phases.

interfering sub-flows, by using three criteria described in the following.

Keeping in mind constraint (8)-(10),if receiver RX_i or transmitter TX_i of sub-flow i matches with RX_j or TX_j of sub-flow j, then sub-flows i and j must be scheduled in different stages. So, sub-flow j is inserted into list $\boldsymbol{L_{1,i}}$ and i into list $\boldsymbol{L_{1,j}}$.

The second criterion takes into account the antenna directivity, as in [11]. According to Fig. 9, the CC evaluates an angle α, by using the cosine law, as a function of distance metric. If $\alpha \leq \frac{\theta}{2}$, then RX_i is inside the beamwidth range of TX_h (i.e., h is an interfering sub-flow for i), consequently sub-flow h is inserted into list $\boldsymbol{L_{2,i}}$, otherwise not.

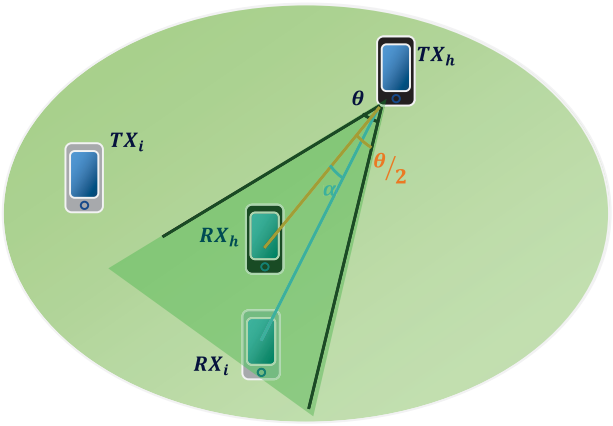

FIGURE 9. Second interference criterion.

Unlike list $\boldsymbol{L_{1,i}}$, list $\boldsymbol{L_{2,i}}$ may contain sub-flows that are **non-severe** interfering with sub-flow i. The second criterion, in fact, takes into account only the antenna directivity, but does not quantify the received interference power. Because the high path loss at 60 GHz greatly reduces the interference level, it could be possible to transmit more interfering sub-flows concurrently, thus increasing the transmission efficiency.

For the above reason, we define a third criterion to establish what interfering sub-flows can be classified as tolerable ones for sub-flow i (i.e., their interference is so non-severe that they can transmit concurrently). The CC allows concurrent

transmissions only if $R_{\text{source},i}$ is guaranteed, as far as possible. At this aim, for sub-flow i, the SINR_i at receiver RX_i is estimated by using (3) where $j \in \boldsymbol{L_{2,i}}$. Next, the obtained SINR_i value is compared with $\text{SINR}_{\text{min},n}$, that is the value which guarantees $R_{\text{source},n}$, already defined in the previous section. If $\text{SINR}_i < \text{SINR}_{\text{min},n}$, the j-th sub-flow with the highest interference power is removed from list $\boldsymbol{L_{2,i}}$ and inserted into list $\boldsymbol{L_{3,i}}$. Then, SINR_i is recalculated. This process continues until $\text{SINR}_i \geq \text{SINR}_{\text{min},n}$.

Finally, for the ith sub-flow, we define $\boldsymbol{L_i} = \boldsymbol{L_{1,i}} \cup \boldsymbol{L_{3,i}}$ as the list of intolerable interfering sub-flows and $\boldsymbol{L_{2,i}}$ as the final list of tolerable interfering sub-flows.

Step 1.2. The second step is to create an undirected interference graph, G(V,E), where each sub-flow is a vertex and the edges represent the intolerable interference relations among all sub-flows. More specifically, if a sub-flow j is present in list $\boldsymbol{L_i}$, then an edge interconnects vertex i and vertex j.

B. PHASE 2

The objective of the second phase is to assign different transmission stages to intolerant interfering sub-flows and to sort the transmission stages.

Step 2.1. Firstly, it is important to minimize the number of stages (N_{stage}) exploiting concurrent transmissions. To solve this problem, we adopt a vertex coloring approach. As a matter of the fact, graph coloring has considerable application to a large variety of complex problems involving optimization. In particular conflict resolution can often be accomplished by means of graph coloring, as shown in [26]. Our target to minimize N_{stage} corresponds to color vertices of the interference graph G(V,E) by using the minimum number of colors such that no two adjacent vertices share the same color. Therefore, different colors (stages) must be assigned to vertices (sub-flows) interconnected between them in the graph. Despite the graph coloring problem being known to be NP-complete, in literature much attention has been focused on the development of heuristic algorithms which will usually produce a good, though not necessarily optimal, coloring for any graph in a reasonable amount of time. Among several algorithms available in literature, we consider the DSATUR graph-coloring method [27] because of its computational efficiency and low complexity. Starting with a single color, the original DSATUR algorithm assigns a new color to the selected vertex only if there are no available colors among those already assigned. It runs iteratively until all vertices are colored.

The main drawback is that the DSATUR algorithm considers all vertices having the same transmission priority. Instead, in our system, a vertex representing the first or an intermediate hop of a multi-hop data flow could be transmitted with low priority, because packets will not be delivered to the destination UE in the current frame, but will only be stored in an AP. In order to minimize the end-to-end delay, we take into account the priority of sub-flows, by proposing a modified version of the DSATUR algorithm, as described in pseudo-code 3. Our modification favors to assign the same color to high priority sub-flows, as far as possible, and this will positively affect the reduction of the end-to-end delay, as we will see in step 2.3.

Pseudo-code 3 *Step 2.1*. Color Graph Criterion (Modified DSATUR Algorithm)

Definitions:

- $\boldsymbol{C}$: Color pool, containing initially only one color ($|\boldsymbol{C}| = 1$)
- for each vertex i:
  - $\theta_{f,i}$: saturation degree (total number of different colors to which the vertex is connected)

Iteration:

```
while all vertices are colored do
    sort all uncolored vertices by decreasing order of θ_{f,i}
    select the vertex (vertices) having the maximum θ_{f,i} value
    if there are more vertices having the same θ_{f,i} value then
        select among them the vertex (vertices) having the maximum number of uncolored neighbors
        if there are more vertices with the above characteristics then
            select among them (if present) a vertex representing a high-priority sub-flow (i.e., the last hop of a multi-hop data flow or a direct sub-flow); otherwise a random choice is made
        end if
    end if
    if there are color(s) in C different from those assigned to the neighbors then
        color the vertex with one of the available colors
    else
        color the vertex with a new color and increase the size of C
    end if
end while
```

At the output of the modified DSATUR algorithm we have a color pool $\boldsymbol{C}$ and one color assigned to each vertex. The pool size $|\boldsymbol{C}|$ represents the number of stages.

Step 2.2. As described in Section IV, in order to further improve the transmission efficiency, our innovative idea is to split transmission of sub-flows into more stages, as formulated in assumption (16). Therefore, the next step is to color vertices with more than one color, while guaranteeing the restrictions imposed by the interference graph. Let us note that this criterion must not increase the number of stages (i.e., the color pool size), already determined in the previous step.

Fig. 10 shows an example of the benefit to apply more than one color to a vertex (i.e., to split data transmission of a sub-flow in more stages). We consider to schedule five sub-flows, whose intolerant interference relations are shown in the interference graph. We suppose that the sub-flow corresponding to vertex 5 needs a large number of time slots to transmit its data. As shown in the interference graph of Fig. 10a, yellow-colored vertex 5 could also be colored with green,

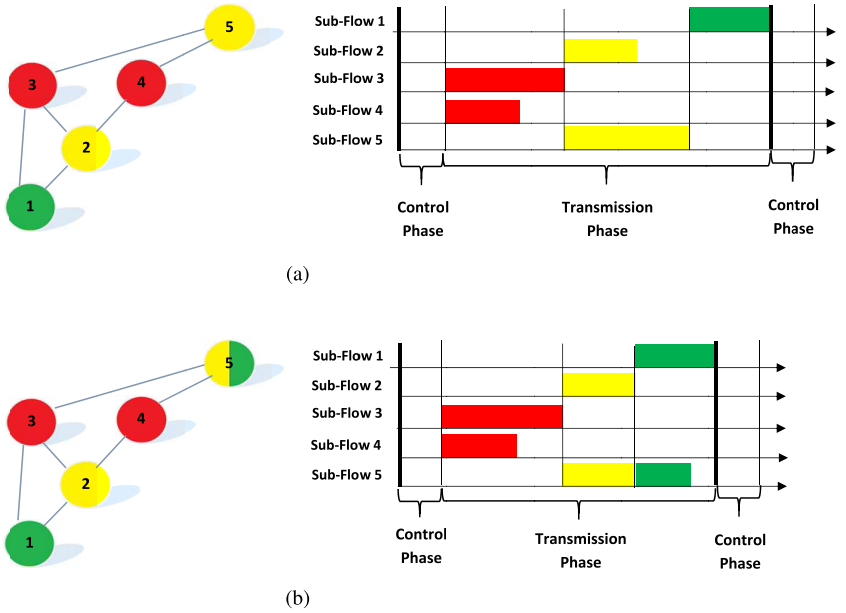

FIGURE 10. Benefit to apply more colors to a vertex. (a) Only one color per sub-flow. (b) More colors per sub-flow.

because its neighbors (intolerable interfering sub-flows) use only the red color, consequently its data transmission can be split, as in Fig. 10b, reducing the current frame length.

In this example, we also note that vertex 4 could be two-colored (red and green), instead of vertex 5. But vertex 4 completes its transmission by using less time slots than the only-red-colored vertex 3, so the frame length would remain the same of Fig. 10a. This means that we need to select, among all vertices that are candidates for multi-coloring, the ones corresponding to the sub-flows that require more time slots to transmit their data. With this aim in mind, we want to associate an easy-to-calculate weight to each vertex, thus to create a weighted colored graph. Because the amount of time slots depends on the number of packets to be transmitted ($n_{packets,i}$) by sub-flow i, the distance between transmitter and receiver, and also on the visibility conditions (α), for the ith vertex, the weight proposed is:

$$w_i = \frac{1}{|\boldsymbol{C_i}|}\left(\frac{dist\,(TX_i, RX_i)}{d_0}\right)^{\alpha} n_{packets,i}, \tag{18}$$

where $\boldsymbol{C_i}$ is the set of colors assigned to vertex i (at the beginning, $|\boldsymbol{C_i}| = 1$). In this way, a high-weighted vertex represents a sub-flow requiring a large number of time slots.

In conclusion, in Step 2.2, we define a new criterion aiming to multi-color one or more vertices, always guaranteeing that adjacent vertices cannot share the same color. This criterion works as follows. First, the controller creates a weighted colored graph and a list of vertices ($\boldsymbol{V}$) in descending order by weights. Next, starting from the first element of $\boldsymbol{V}$, if in $\boldsymbol{C}$ there are colors available for it (i.e., one or more colors in $\boldsymbol{C}$ are not assigned to any neighbor and to the vertex itself), then one additional color is assigned to the vertex. After that, its weight value and list $\boldsymbol{V}$ are updated so that its priority to obtain another color is reduced. Instead, if there are no available colors for it, the vertex is deleted from the list. This process continues until $\boldsymbol{V}$ is empty.

Step 2.3. Now, we have a list of stages (colors) and a many-to-many correspondence between colors and vertices. The last step in the second phase is to establish the transmission order of stages (colors), which influences the end-to-end delay. At this regard, we remember that in our modified DSATUR algorithm, we have given high priority to direct or last-hop sub-flows. For this reason, we establish to transmit the colors (stages) containing more high priority sub-flows first, the other ones then. In the event that there is an equality, the CC makes a random choice.

As output of the entire second phase, we have $|\boldsymbol{C}|$ colors in transmission order ($k = 1, 2, \ldots, |\boldsymbol{C}|$) and, for each color k, a list of vertices using this color ($\boldsymbol{L_{col_k}}$).

C. PHASE 3

The third and last algorithm phase focuses on calculating the number of time slots to be allocated to each color. It is important to make the most efficient allocation with the minimum number of time slots, while fulfilling all sub-flows. The issue is that the amount of required resources depends on transmission rate, which in turn depends on current interference.

Step 3.1. In order to establish in the first instance the duration of the first stage (color $k = 1$) in terms of time slots, among all vertices in list $\boldsymbol{L_{col_1}}$ the CC selects only the monochrome ones, because they cannot split the related data in more stages. For monochrome vertex i, the Central Controller calculates the estimated SINR_i, evaluated by using (3), where $j \in \boldsymbol{L_{2,i}} \cap \boldsymbol{L_{col_1}}$, i.e., by taking into account

only the sub-flows transmitting during stage 1, including the multi-color ones. Then, the transmission rate R_i can be estimated by using (2). Finally, the number of time slots necessary to transmit the data related to sub-flow i is:

$$n_{ts_{i,1}} = \left\lceil \frac{n_{packets,i} 8N_{\text{bytes,packet}}}{t_{PAY} R_i} \right\rceil. \tag{19}$$

After calculating (19) for each monochrome vertex in $\boldsymbol{L_{col_1}}$, the length of stage 1 (n_{ts_1}) is set equal to $\max_i \{n_{ts_{i,1}}\}$.

Before moving on the second color, the CC needs to analyze the multi-colored vertices $\in \boldsymbol{L_{col_1}}$, if any. They ensure more concurrent transmissions, as previously shown, but on the other hand they complicate the estimation of interference. For example, if sub-flow h corresponding to a two-colored vertex $\in \boldsymbol{L_{col_1}} \cap \boldsymbol{L_{col_j}}$ could meet its demand by using only color 1, then the CC deletes it from $\boldsymbol{L_{col_j}}$ because it will cause no interference to sub-flows transmitting during the j-th stage. Instead, if sub-flow h needs to use more than one color, then the CC evaluate how many sub-flow packets can be transmitted during stage (color) 1 (i.e., in n_{ts_1} time slots):

$$n_{packets_{h,col_1}} = \left\lfloor \frac{n_{ts_1} R_h t_{PAY}}{8N_{\text{bytes,packet}}} \right\rfloor. \tag{20}$$

Therefore, the Central Controller estimates the amount of remaining packets to be transmitted by sub-flow h during stage (color) j.

Once all vertices in $\boldsymbol{L_{col_1}}$ are analyzed, the CC can establish the duration of the second color and so on, in a similar manner to the first one. The main difference consists in the multi-colored vertices management. If the considered color represents the last available color of a multi-colored vertex, then the CC needs to evaluate whether all its remaining packets can be transmitted during this last stage. If it does not succeed in meeting its demands, then the CC includes this vertex (sub-flow) in a list ($\boldsymbol{L_{NotSatisfied}}$) containing all sub-flows needing to extend the length of one or more colors (stages) to transmit all the related data. The details of all operations described above are shown in pseudo-code 4.

Step 3.2. After the length of each stage has been established in the first instance by the monochrome vertices, the CC needs to analyze the elements in $\boldsymbol{L_{NotSatisfied}}$.

If the list contains a single sub-flow i, the CC selects the color whose length needs to be extended, as following:

- if it is a high-priority sub-flow, then its first color is selected;
- if it is a low-priority sub-flow, then its last color is selected.

Then, the length of the selected color is set out as the number of time slots necessary to fulfill data transmission of sub-flow i.

Finally, we describe the strategy adopted if $\boldsymbol{L_{NotSatisfied}}$ contains more than one sub-flow. We consider the graph in Fig. 11. We suppose that the transmission order derived in Step 2.3 is {RED, YELLOW, GREEN}, $\boldsymbol{L_{NotSatisfied}} = \{5, 6\}$, and both vertices correspond to direct sub-flows (high-priority). If the CC applies the above strategy for them, then

Pseudo-code 4 *Step 3.1*. Initial Estimate of Stage Lengths

Definitions:

- $\boldsymbol{C}$: Color Pool, in transmission order
- $\boldsymbol{L_{col_k}}$: list of vertices using color (stage) k
- n_{ts_k}: number of time slots to be allocated to color (stage) k
- for each vertex (sub-flow) i:
  - $\boldsymbol{C_i}$: set of assigned color(s), in transmission order
  - $n_{packets,i}$: number of packets
  - $n_{ts_{i,k}}$: number of time slots required in color (stage) k

Iteration:

```
for each color k ∈ C do
  for each vertex i|i ∈ L_col_k ∧ |C_i| = 1 do
    calculate n_ts_{i,k}
  end for
  set n_ts_k = max_i {n_ts_{i,k}}
  for each vertex i|i ∈ L_col_k ∧ |C_i| > 1 do
    if color k is the first element of C_i then
      calculate n_ts_{i,k}
      if n_ts_{i,k} < n_ts_k then
        delete vertex i from L_col_m, ∀m ∈ C_i, m > k
      else
        evaluate how many packets i can transmits in n_ts_k
      end if
    else if k is the last element of C_i then
      estimate the number of time slots necessary to fulfill data transmission of i in color k: n̂_ts_{i,k}
      if n̂_ts_{i,k} > n_ts_k then
        insert i in list L_NotSatisfied
      end if
    else
      calculate n̂_ts_{i,k}
      if n̂_ts_{i,k} < n_ts_k then
        delete vertex i from L_col_m, ∀m ∈ C_i, m > k
      else
        evaluate how many packets i can transmits in n_ts_k
      end if
    end if
  end for
end for
```

we obtain that sub-flow 5 requires to extend color YELLOW, and sub-flow 6 color RED.

We note that this strategy is not efficient, because the best solution is to extend the length of color GREEN they have in common. In a more complex situation, there may be more colors in common with all sub-flows in $\boldsymbol{L_{NotSatisfied}}$. In order to extend the minimum number of colors ensuring transmission efficiency, we propose the following greedy criterion. It is a sub-optimal solution to keep the computational load low.

We introduce a color matrix $\boldsymbol{M_C}$ of size $|\boldsymbol{L_{NotSatisfied}}| \times |\boldsymbol{C}|$, where each element $a_{i,j} = 1$ if the ith sub-flow ($i \in \boldsymbol{L_{NotSatisfied}}$) uses color j ($j \in \boldsymbol{C}$), $a_{i,j} = 0$ otherwise. Fig. 12 shows an example of this procedure. The controller selects

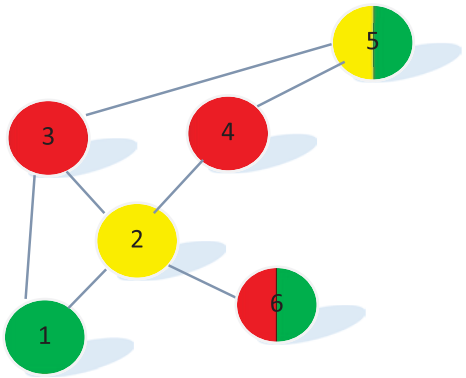

FIGURE 11. Example of more colors in $L_{NotSatisfied}$.

$$M_C = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$M'_C = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$L_{ColorExt} = \{7,4\}$$

FIGURE 12. Color matrix procedure.

the column with the maximum number of 1s (random choice between columns 5 and 7). In our example, column 7 is selected. Next, the controller deletes all rows containing 1 in the selected column to obtain a row-reduced matrix, and inserts color 7 in list $L_{ColorExt}$. This process continues until Color Matrix is empty. As output of the criterion, we obtain a list of color(s) to be extended ($L_{ColorExt}$).

Then, we perform the following steps:

1) for each color j in $L_{ColorExt}$, we select each sub-flow $i \in L_{NotSatisfied}$ using **only** color j, and estimate the amount of time slots necessary to fulfill the relative data transmission ($\widehat{n}_{ts_{i,j}}$). Then, we remove them from $L_{NotSatisfied}$, and set out the new length of color j as the maximum value of $\widehat{n}_{ts_{i,j}}$.
2) We verify whether sub-flow $i \in L_{NotSatisfied}$ (if any) can now complete its transmission with the new color length values. If so, then we remove it from $L_{NotSatisfied}$. After that, if $L_{NotSatisfied}$ is not empty, M_C is recalculated and the process starts again, otherwise it ends.

The details on extending the stage lengths of sub-flows in $L_{NotSatisfied}$ are shown in the pseudo-code 5.

VI. PERFORMANCE EVALUATION

In this section, we evaluate the performances of the proposed access control scheme in MATLAB environment.

Pseudo-code 5 *Step 3.2.* Stage Lengths to be Extended

Definitions:

- $L_{NotSatisfied}$: list of vertices (sub-flows) not satisfying their demands
- n_{ts_k}: number of time slots allocated to color (stage) k
- for each vertex (sub-flow) i:
  - C_i: set of assigned color(s), in transmission order
  - $n_{packets,i}$: number of packets

Iteration:

while $L_{NotSatisfied}$ is not empty **do**
  if $|L_{NotSatisfied}| == 1$ **then**
    if $i \in L_{NotSatisfied}$ represents a high-priority sub-flow **then**
      select the first color $m \in C_i$
    else
      select the last color $m \in C_i$
    end if
    estimate $\widehat{n}_{ts_{i,m}}$ (i.e., the number of time slots necessary to fulfill data transmission of sub-flow i in color m) and set $n_{ts_m} = \widehat{n}_{ts_{i,m}}$
  else
    create the Color Matrix M_C
    repeat
      select the column j with the maximum number of 1s;
      delete all rows containing 1 in column j and insert j in list $L_{ColorExt}$
    until M_C is empty
    for each color $j \in L_{ColorExt}$ **do**
      for each sub-flow $l \in L_{NotSatisfied}$ such that $j \in C_l \wedge |C_l \cap L_{ColorExt}| = 1$ **do**
        remove l from $L_{NotSatisfied}$ and estimate $\widehat{n}_{ts_{l,j}}$
      end for
      set $n_{ts_j} = \max_l \{\widehat{n}_{ts_{l,j}}\}$
    end for
    for each sub-flow $h \in L_{NotSatisfied}$ (i.e., $|C_h \cap L_{ColorExt}| > 1$), if any **do**
      if h can now complete its transmission with the new color length values **then**
        remove h from $L_{NotSatisfied}$
      end if
    end for
  end if
end while

A. SIMULATION ASSUMPTIONS

In our simulation scenario, as in [16] we consider a flat indoor area of 50 $m \times 50m$, covered with 9 APs uniformly distributed to form a regular grid. The CC is located inside the AP in the center of the scenario. We consider N_U UEs uniformly distributed within the whole area. We assume that transmissions among APs occur in LOS conditions, whereas any other ones in NLOS.

The set of parameters used in simulations is provided in Table 1. Each result is averaged over 100 independent

TABLE 1. System parameters.

Parameter	Symbol	Value
System bandwidth [11]	W	$1600MHz$
Carrier frequency	f_c	$60GHz$
Beam angle	θ	45°
UE and AP TX Power	P_T	$10mW$
Background Noise	N_0	$-114dBm/MHz$
PL exponent [24]	α	1.73 (LOS) 3.19 (NLOS)
Shadowing Standard deviation [24]	σ	$3dB$ (LOS) $8.29dB$ (NLOS)
Maximum NLOS distance	$d_{\text{max,NLOS}}$	$16.97m$
Control phase length	t_{con}	10 time slots
TX and RX Gain	G_R, G_T	8
Reference distance	d_0	$1m$
Slot time [16]	t_{SLOT}	$5\mu s$
Payload time [16]	t_{PAY}	$4\mu s$
Packet size [16]	$N_{\text{bytes,packet}}$	1000 bytes
Transceiver efficiency	η	1
Number of UEs	N_U	$20, 25, \ldots, 50$
Number of active commun.	N_C	$2, 5, 10, 15$
Delay threshold [16]	d_{thr}	$50ms$
Source bit rate	R_{source}	2 Gbps, 4 Gbps
Simulation time [16]	t_{SIM}	$500ms$

simulations. To take into account the frame overhead, we set each control phase length (t_{cont}) equal to 10 time slots [22].

B. TRAFFIC MODEL

We assume that a source UE generates an active communication for a destination UE at a random time (t_0) uniformly distributed in the continuous range $[0, t_{SIM}]$. An active communication consists of a variable number of packets, generated at a constant peak rate (R_{source}). We introduce 6 traffic classes and assume that the number of generated packets for an active communication of class h is uniformly distributed in a discrete range $\Delta_h = [5000(h-1), 5000h]$, with $h = \{1, 2, \ldots, 6\}$.

For each simulation, we set the number of active communications (N_C). Source and destination UEs are randomly chosen on N_U UEs in such a way that:

- a UE may be only source or destination UE;
- a source UE is, at most, the source of a single active communication;
- a destination UE may be the destination of more active communications.

C. PERFORMANCE METRICS

We compare our solution with the D3MAC scheme [16]. Unlike our control scheme, Gao *et al.* consider that the transmission rate of each wireless link is fixed and known (equal to 2, 4 or 6 Gbps, according to the distance between devices). In order to fairly assess the two control schemes, we adapt D3MAC scheme to our more accurate approach in which the transmission rates are variable according to the interference level. We analyze the system performance in several traffic conditions and any active communication requiring a stringent delay threshold, $d_{thr} = 50ms$, as in [16]. We introduce several metrics aiming to analyze not only performances of the overall system (packet level performances) but also of each active communication (session level performances).

Let P be the set of all generated packets, and P_d the set of packets delivered within the delay threshold, i.e., $P_d = \{p | p \in P \wedge d_{e2e,p} < d_{thr}\}$, where $d_{e2e,p}$ is the end-to-end delay of packet p. We measure:

$$\text{Successfully delivered packets}(\%) = \frac{|P_d|}{|P|}100, \tag{21}$$

$$\text{Average packets delay} = \frac{1}{|P_d|}\sum_{\forall p \in P_d} d_{e2e,p}. \tag{22}$$

In order to compare the behavior of two control schemes in the same traffic conditions, we define "Throughput Gain" of the ith active communication (ΔTh_i) as the difference between the throughput achieved by our control scheme (Th_i) and the one achieved by D3MAC (Th_i^{D3MAC}), normalized to the source rate ($R_{\text{source},i}$).

$$\Delta Th_i = \frac{Th_i - Th_i^{D3MAC}}{R_{\text{source},i}}, \quad \Delta Th_i \in [-1, 1]. \tag{23}$$

where Th_i is calculated as the ratio between the number of successfully delivered packets and the time interval from the first packet generation time to the last packet delivery time.

We also estimate the Average Throughput Gain, as:

$$\Delta Th = \frac{1}{N_C}\sum_{i=1}^{N_C} \Delta Th_i. \tag{24}$$

Finally, we estimate the Fairness among active communications, which is an important performance criterion in all resource allocation schemes. In literature the Jain's index [28] is a parameter which quantitatively measures the degree of fairness offered by a system allocating resources to N_C contending requests. The Jain's Index is defined as follows:

$$J_{index} = \frac{\left(\sum_{i=1}^{N_C} x_i\right)^2}{N_C \sum_{i=1}^{N_C} x_i^2}, \tag{25}$$

where x_i represents the resources allocated to ith contender. The results range from $1/N_C$ (i.e., all resources allocated to a single active connection) to 1 (i.e., each request receives the same amount of resources).

We can assume $x_i = |P_{d,i}|$, that is, the number of packets related to active communication i delivered within the delay threshold. However, J_{index} is a good fairness parameter only if all requests contend the same resources. Instead, in our system there may be some sub-flows that can use the entire resources without contenting with the other ones (because they do not have interfering sub-flows), while other sub-flows need to share some radio resources.

This means that in our scenario the J_{index} values are not representative in absolute terms. For this reason, we assume as fairness index "ΔJain's Index", the difference between

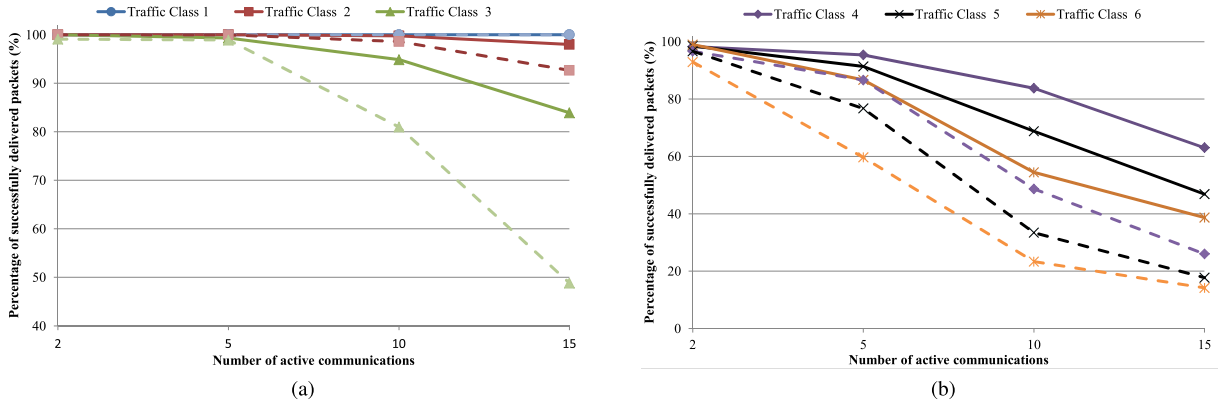

FIGURE 13. Percentage of successfully delivered packets, under different traffic classes and number of active communications. Source rate of 2 Gbps for each active communication. Solid line represents our control scheme performance, dashed line D3MAC performance. (a) Traffic class 1, 2, and 3. (b) Traffic class 4, 5, and 6.

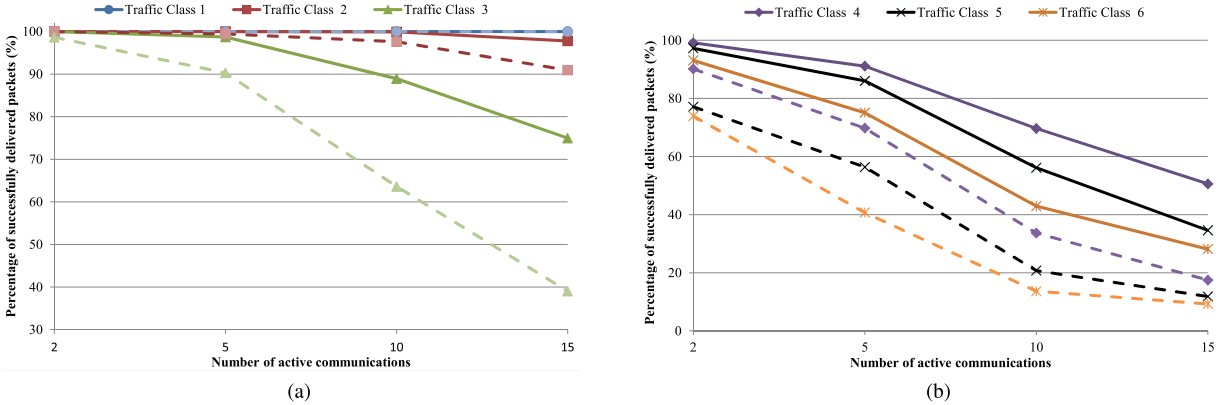

FIGURE 14. Percentage of successfully delivered packets, under different traffic classes and number of active communications. Source rate of 4 Gbps for each active communication. Solid line represents our control scheme performance, dashed line D3MAC performance. (a) Traffic class 1, 2, and 3. (b) Traffic class 4, 5, and 6.

the fairness achieved by our proposal (J_{index}) and the one achieved by D3MAC scheme (J_{index}^{D3MAC}):

$$\Delta J_{index} = J_{index} - J_{index}^{D3MAC}, \tag{26}$$

where $\Delta J_{index} \in \left[\frac{1}{N_C} - 1, 1 - \frac{1}{N_C}\right]$

D. PERFORMANCE ANALYSIS

1) PACKET LEVEL PERFORMANCE

Under several traffic conditions ($N_U = 30$, $N_C = \{2, 5, 10, 15\}$ and Traffic class $h = \{1, 2, \ldots, 6\}$), the percentage of successfully delivered packets by adopting our control scheme and D3MAC scheme are shown in Figs. 13 and 14, where source rate of any active communication is equal to 2 Gbps and 4 Gbps, respectively.

In the case of very light traffic load (traffic class 1), both our algorithm and the D3MAC seem to show the same excellent performances. However, when the traffic load increases our solution outperforms the D3MAC scheme, the more the traffic load the more the difference in performance. Of course, in the case of very heavy traffic load, both algorithms show worse performance because the network is close to saturation, but our proposal is still better.

We also report the average delay measures in Figs. 15 and 16, when source rate of any active communication is equal to 2 Gbps and 4 Gbps, respectively. We note that already under very light load conditions (traffic class 1) our control scheme is better than D3MAC. The delay difference increases as the traffic load increases. In the case of heavy traffic conditions ($N_C = 15$, $R_{\text{source}} = 2$ Gbps and traffic class 6) this difference comes down. Let us note that the measured statistics take into account only packets delivered within the threshold, so in heavy traffic conditions our control scheme achieves similar performance in average delay compared to D3MAC but with a very larger number of delivered packets (about 150% more). In the case of very heavy traffic load (e.g., $N_C = 15$, $R_{\text{source}} = 4$ Gbps and traffic class 4 to 6) the delay performance of D3MAC seems to be better, but only because the network is close to saturation and only a small amount of packets (20%) has been successfully delivered.

Now, as in [16], we assess the control schemes under several number of UEs, uniformly distributed in the whole area.

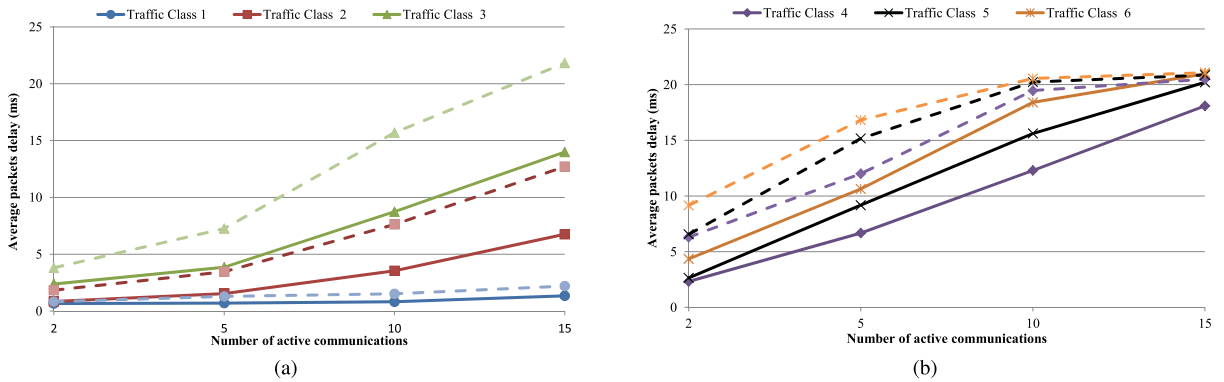

FIGURE 15. Average packets delay, under different traffic classes and number of active communications. Source rate of 2 Gbps for each active communication. Solid line represents our control scheme performance, dashed line D3MAC performance. (a) Traffic class 1, 2, and 3. (b) Traffic class 4, 5, and 6.

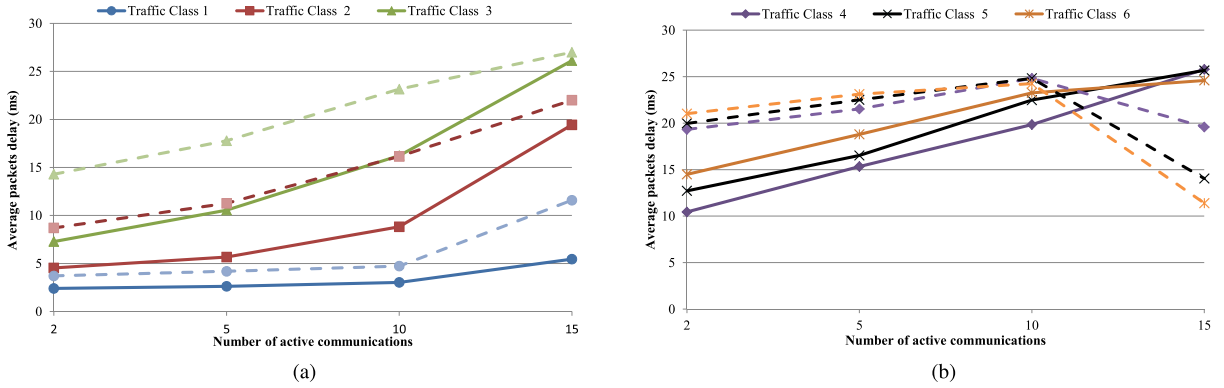

FIGURE 16. Average packets delay, under different traffic classes and number of active communications. Source rate of 4 Gbps for all active communications. Solid line represents our control scheme performance, dashed line D3MAC performance. (a) Traffic class 1, 2, and 3. (b) Traffic class 4, 5, and 6.

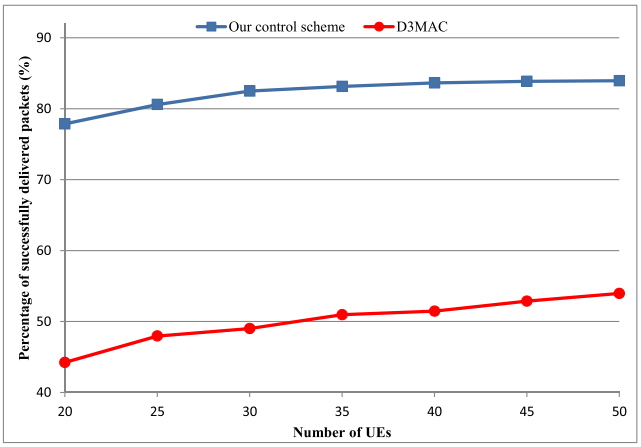

FIGURE 17. Percentage of successfully delivered packets, under different number of UEs. 10 active communications, $R_{\text{source}} = 2$ Gbps and Traffic class 3 for any active communication.

This simulation is characterized by an invariable total offered load: 10 active communications, $R_{\text{source}} = 2$ Gbps and traffic class 3. Fig. 17 shows that the percentage of successfully delivered packets increases with N_U for both algorithms. The explanation for this is that, for equal N_C, the smaller the number of UEs, the higher the probability that the same UE will be the destination of two or more active communications, thus increasing the number of intolerable interfering sub-flows, so the percentage of successfully delivered packets within the threshold is reduced. This phenomenon is all the more evident when the number of UEs is less than or closer to three times the number of active connections. Similar results are obtained by considering different total offered traffic loads.

2) SESSION LEVEL PERFORMANCE

In order to better understand the behavior of our control scheme compared to the one of D3MAC, in Fig. 18, we show the throughput gain of each active communication in two single representative simulation tests, with fixed traffic conditions, consisting on $N_U = 30$, $N_C = 15$, $R_{\text{source}} = 2$ Gbps, and traffic class 3.

In the first simulation test, our control scheme achieves an Average Throughput Gain value (represented as the horizontal red line) of about 50 percentage points compared to D3MAC scheme, and all active communications are

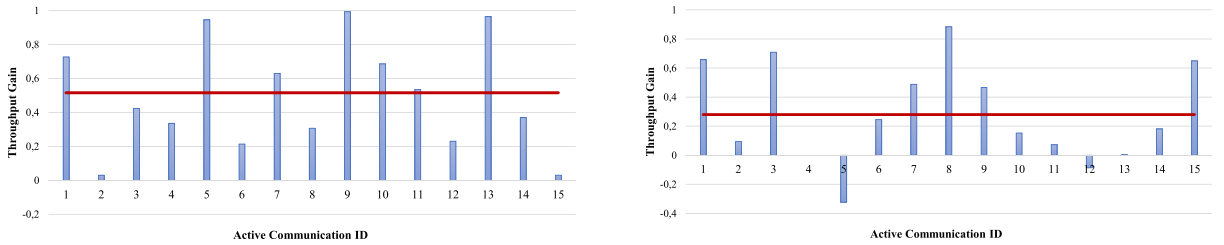

FIGURE 18. Throughput Gain for each active communication and average value (horizontal red line). 15 active communications, $R_{\text{source}} = 2$ Gbps and Traffic class 3 for any active communication. (a) Simulation test 1. (b) Simulation test 2.

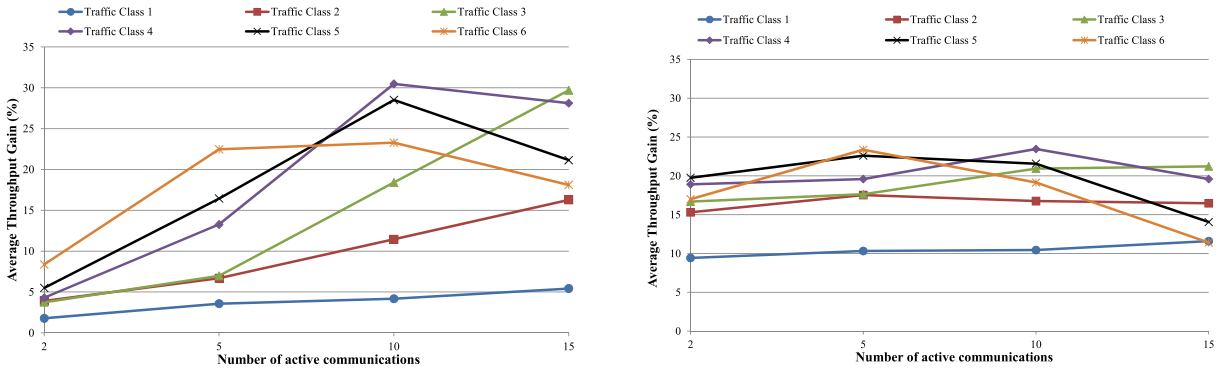

FIGURE 19. Average Throughput Gain under different Traffic Classes and number of active communications. (a) Source rate of 2 Gbps for any active communication. (b) Source rate of 4 Gbps for any active communication.

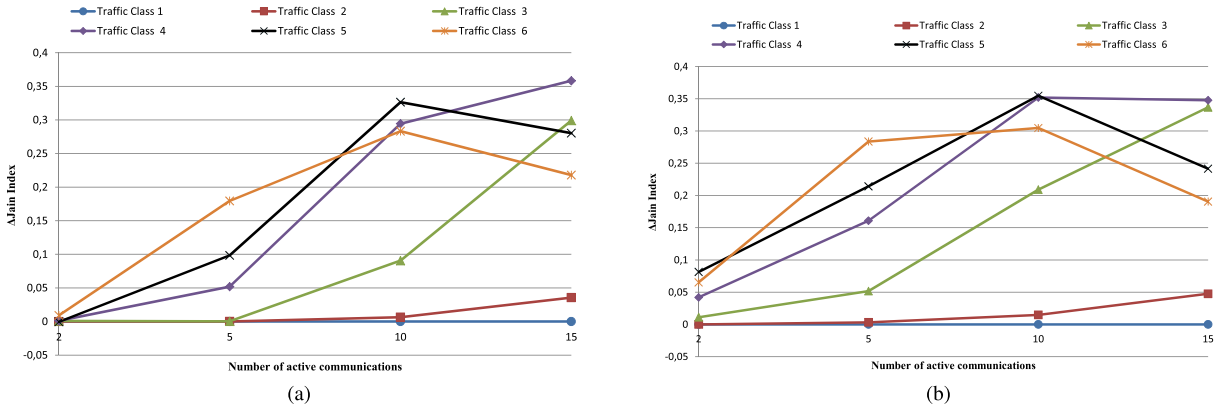

FIGURE 20. ΔJ_{index} under different traffic loads and number of active communications. (a) Source rate of 2 Gbps for all active communications. (b) Source rate of 4 Gbps for all active communications.

characterized by $\Delta Th_i \geq 0$. More specifically, three active communications ($i = \{5, 9, 13\}$) show $\Delta Th_i \approx 1$. This means that with our control scheme these three active communications achieve almost the maximum throughput, while with D3MAC their throughput is approximately zero (i.e., their transmission is inhibited and a very small number of packets is delivered on time).

In the second simulation test, we achieve an Average Throughput improvement of 27.9% compared to D3MAC. Unlike the first one, the improvement does not occur on each single active connection, so we experience $\Delta Th_i < 0$ for two active communications ($i = \{5, 12\}$). However, let us note that in the worst case $\Delta Th_5 \approx -0.32$, it means that our control strategy has not inhibited any communication, while allowing three active connections ($i = \{3, 8, 15\}$) to achieve significantly better performance than the D3MAC scheme and improving the Average Throughput Gain.

These two simulation tests show also that the system throughput depends strongly on the randomness of configurations (i.e., the distribution of source and destination

UEs). However, our control schemes always outperforms the D3MAC, as shown in Fig. 19, where the Average Throughput Gain, ΔTh (%), in average over 100 independent simulations, is always positive, whatever the traffic class and the number of active communications.

The analysis related to Fig. 18 suggests that our scheme may have higher degree of fairness than the benchmark scheme. At this regard, in Fig. 20, we report ΔJ_{index} under the traffic conditions previously considered. We note that under very light load conditions (traffic class 1), regardless of the number of active communications, $\Delta J_{index} = 0$, that is, J_{index} is about 1 for both control scheme. As the traffic class and the number of active communications increase, the probability of interfering sub-flows increases. So, the probability of many sub-flows competing for the same resources increases. In these conditions, ΔJ_{index} is always greater than zero, so our control scheme results more fair than D3MAC and the advantage is getting higher as the traffic load rises, up to a medium-heavy load conditions. Finally, in heavy traffic conditions, this difference comes down because the network is close to saturation.

VII. CONCLUSION

In this paper, we propose a Centralized Access Control Scheme for MMB systems, which jointly manages D2D communications and transmissions in the access and the backhaul networks. Our strategy provides for an innovative flow management scheme and a multi-criteria scheduling algorithm, based on multi graph-coloring techniques.

We benchmark our centralized control scheme against the D3MAC scheme [16], one of the most complete works available in literature. Unlike D3MAC, our criteria take also into account service requirements, in terms of source rate and end-to-end delay. In addition, we adapt the D3MAC to our more accurate approach in which the transmission rates are variable according to the interference level. Extensive simulations, under various traffic classes and number of active communications, demonstrate that our access scheme has notably improved the concurrent transmission efficiency and outperforms the considered reference scheme in terms of throughput, end-to-end packet delay, and fairness.

REFERENCES

[1] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tut.*, vol. 18, no. 3, pp. 1617–1655, 3rd Quart., 2016.

[2] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 1801–1819, Nov. 2014.

[3] D. Wu, J. Wang, Y. Cai, and M. Guizani, "Millimeter-wave multimedia communications: Challenges, methodology, and applications," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 232–238, Jan. 2015.

[4] A. Mastrosimone and D. Panno, "Moving network based on mmWave technology: A promising solution for 5G vehicular users," *Wireless Netw.*, vol. 24, no. 7, pp. 2409–2426, Oct. 2018.

[5] S. K. Yoo, S. L. Cotton, R. W. Heath, and Y. J. Chun, "Measurements of the 60 GHz UE to eNB channel for small cell deployments," *IEEE Wireless Commun. Lett.*, vol. 6, no. 2, pp. 178–181, Apr. 2017.

[6] 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.

[7] *IEEE Standard for High Data Rate Wireless Multi-Media Networks*, IEEE Standard 802.15.3-2016 (Revision IEEE Std 802.15.3-2003, Jul. 2016, pp. 1–510.

[8] *ISO/IEC/IEEE International Standard for Information Technology-Telecommunications and Information Exchange Between Systems-Local and Metropolitan Area Networks–Specific Requirements-Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 3: Enhancements for Very High Throughput in the 60 GHz Band*, Standard ISO/IEC/IEEE 8802-11:2012/Amd.3:2014(E), IEEE 802.11ad-2012, Mar. 2014, pp. 1–634.

[9] S. Scott-Hayward and E. Garcia-Palacios, "Multimedia resource allocation in mmwave 5G networks," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 240–247, Jan. 2015.

[10] 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.

[11] W. U. Rehman, J. Han, C. Yang, M. Ahmed, and X. Tao, "On scheduling algorithm for device-to-device communication in 60 GHz networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2014, pp. 2474–2479.

[12] L. Wang, S. Liu, M. Chen, G. Gui, and H. Sari, "Sidelobe interference reduced scheduling algorithm for mmWave device-to-device communication networks," *Peer-to-Peer Netw. Appl.*, vol. 12, no. 1, pp. 228–240, Jan. 2019.

[13] I. K. Son, S. Mao, M. X. Gong, and Y. Li, "On frame-based scheduling for directional mmWave WPANs," in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 2149–2157.

[14] Y. Niu, Y. Li, D. Jin, L. Su, and D. Wu, "A two stage approach for channel transmission rate aware scheduling in directional mmWave WPANs," *Wireless Commun. Mobile Comput.*, vol. 16, no. 3, pp. 313–329, 2016.

[15] Y. Niu, L. Su, C. Gao, Y. Li, D. Jin, and Z. Han, "Exploiting device-to-device communications to enhance spatial reuse for popular content downloading in directional mmWave small cells," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5538–5550, Jul. 2016.

[16] C. Gao, Y. Li, H. Fu, Y. Niu, D. Jin, S. Chen, and H. Zhu, "Evaluating the impact of user behavior on D2D communications in millimeter-wave small cells," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 6362–6377, Jul. 2017.

[17] N. Giatsoglou, K. Ntontin, E. Kartsakli, A. Antonopoulos, and C. Verikoukis, "D2D-aware device caching in mmWave-cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 2025–2037, Sep. 2017.

[18] N. Wei, X. Lin, and Z. Zhang, "Optimal relay probing in millimeter-wave cellular systems with device-to-device relaying," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 10218–10222, Dec. 2016.

[19] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[20] S. Riolo, D. Panno, and A. Di Maria, "A new centralized access control for mmWave D2D communications," in *Proc. 13th IEEE Int. Conf. Wireless Mobile Comput., Netw. Commun. (WiMob)*, Rome, Italy, Oct. 2017, pp. 1–8.

[21] H. Deng and A. Sayeed, "Mm-wave MIMO channel modeling and user localization using sparse beamspace signatures," in *Proc. IEEE 15th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2014, pp. 130–134.

[22] D. Panno and S. Riolo, "On mmWave radio network planning based on a centralized access control," in *Proc. Int. Conf. Sel. Topics Mobile Wireless Netw. (MoWNeT)*, Jun. 2018, pp. 1–8.

[23] S. Singh, F. Ziliotto, U. Madhow, E. M. Belding, and M. Rodwell, "Blockage and directivity in 60 GHz wireless personal area networks: From cross-layer model to multihop MAC design," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1400–1413, Jun. 2009.

[24] *Study on Channel Model for Frequency Spectrum Above 6 GHz*, document TR 38.901, Version 15.0.0, 3GPP, Jun. 2018.

[25] K. Haneda, L. Tian, H. Asplund, J. Li, Y. Wang, D. Steer, C. Li, T. Balercia, S. Lee, Y. Kim, and A. Ghosh, "Indoor 5G 3GPP-like channel models for office and shopping mall environments," in *Proc. IEEE Int. Conf. Commun. Workshops (ICCW)*, May 2016, pp. 694–699.

[26] F. T. Leighton, "A graph coloring algorithm for large scheduling problems," *J. Res. Nat. Bureau Standards*, vol. 84, pp. 489–506, Nov. 1979.

[27] D. Brélaz, "New methods to color the vertices of a graph," *Commun. ACM*, vol. 22, no. 4, pp. 251–256, Apr. 1979.

[28] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer system," Eastern Res. Lab., Digit. Equip. Corp., Maynard, MA, USA, Res. Rep. TR-301, Sep. 1984.

DANIELA PANNO received the laurea degree (Hons.) in electrical engineering from the University of Catania, Italy, in 1989, and the Ph.D. degree in electronic engineering and computer science engineering from the University of Palermo, Italy, in 1993. In 1989, she joined the Department of Electrical, Electronics, and Computer Engineering with the University of Catania, where she is currently an Associate Professor of telecommunications. She has attended several international workshops and symposia as invited speaker. She has served as a Guest Editor for *Computer Communications*. Her research interests include radio resource management, green networking, mmWave D2D communication, SDN and NFV technologies for 5G cellular networks, and mobile services in IoT environments.

SALVATORE RIOLO received the bachelor's degree in electronics engineering and the master's degree (Hons.) in telecommunications engineering from the Department of Electrical, Electronics and Computer Engineering, University of Catania, Italy, in 2012 and 2017, respectively, where he is currently pursuing the Ph.D. degree in systems, energy, computer, and telecommunications engineering. From 2017 to 2018, he was an early-stage Researcher in high bit rate device-to-device services for 5G mobile networks with the University of Catania. His scientific interests include radio resource management for 5G mobile networks, green networking, mmWave D2D communications, and access protocols for massive machine type communications.