A QoE-Aware Mode Selection Framework for Video Streaming in D2D Networks

IBTIHAL AHMED[1]**, MAHMOUD H. ISMAIL**[1,2]**, (Senior Member, IEEE),**
AND MOHAMED S. HASSAN[1]

[1]Department of Electrical Engineering, American University of Sharjah, Sharjah, UAE
[2]Department of Electronics and Communications Engineering, Faculty of Engineering, Cairo University, Giza 12613, Egypt

Corresponding author: Mahmoud H. Ismail (mhibrahim@aus.edu)

ABSTRACT Device-to-Device (D2D) communication is a novel approach that has been proposed to enhance the performance of traditional cellular networks. This work focuses on video streaming using D2D communications underlaying a Long Term Evolution (LTE) network where resources are allocated to each D2D pair in three modes of operation; cellular, dedicated and reuse. The objective is to maximize the sum-rate of the D2D pairs while maintaining continuous video playback with an acceptable quality level. Unlike previous works, which assign one channel to each user or divide the available resources equally between them or assume constant bit rate streaming (CBR), we estimate the number of Resource Blocks (RBs) needed for all users according to their rate and Signal-to-Noise Ratio (SNR) requirements and consider Scalable Video Coding (SVC). A complete Quality of Experience (QoE)-aware framework for D2D-based video streaming that includes resource allocation, power control, mode selection and SVC layer selection is then proposed while taking into consideration buffer occupancies and the video peak signal-to-noise ratio (PSNR). The efficacy of the proposed scheme is evaluated through extensive simulations and compared to the scenario where only one mode of operation is assigned to all D2D pairs. Furthermore, the effect of the number of D2D users as well as their mobility is also investigated. Simulation results show that the proposed framework improves the throughput of the D2D pairs when compared to a single mode of operation while at the same time providing better video quality and continuity.

INDEX TERMS Video streaming, D2D communication, mode selection, quality-of-experience, interference.

I. INTRODUCTION

The past few decades have witnessed an exponential growth in both the number of cellular subscribers as well as their traffic demands. Voice calls and low data rate services dominated wireless traffic in the past. However, today, the widespread usage of portable gadgets and mobile applications, especially video-based ones, have prompted a rapid growth in data traffic [1]. According to the recent Global Internet Phenomena Report released by Sandvine in 2019, video represents more than 60% of the total downstream volume of traffic on the Internet [2]. This rapid increase in data traffic, dominated by video, is expected to grow even more over the coming years as well. To enhance the performance of a traditional cellu-

lar network, the 3rd generation partnership Project (3GPP) investigated the idea of enabling Device-to-Device (D2D) communication where direct communication between two cellular users (CUs) is established without the involvement of the base station (BS). Consequently, D2D communication is an approach that has received much attention during the past years as a key solution to relieve network congestion and improve the system throughput [3]–[5]. The advantages of integrating D2D communication into a cellular network also include improving energy efficiency [6], extending the network coverage and enhancing the performance of edge users [7]. Lots of works in the literature also investigated the idea of exploiting D2D communication in video transmission to improve the quality of experience (QoE) of the users, which is related to two factors; quality and continuity of the video stream. Despite the above-mentioned benefits of

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

D2D communication, some challenges still exist and require further investigation as presented in a survey [7], including mode selection, interference coordination, resource allocation, security as well as mobility management.

In a D2D-enabled network, the D2D devices can communicate using three different modes of operation; the cellular, the dedicated and the reuse modes. In the cellular mode, a D2D connection is established through a BS using uplink (UL) and downlink (DL) channels as if the D2D devices were conventional CUs. In the dedicated (overlay) mode, a percentage of the cellular spectrum is exclusively dedicated to D2D communications and each D2D pair is allocated dedicated resources, whereas in the reuse (underlay) mode, the D2D pairs reuse the cellular spectrum and hence cause interference to the conventional CUs in the network. One of the problems that has received much attention over the past few years is that of mode selection where the BS needs to select one of the three aforementioned modes to establish a D2D connection. This problem has been especially investigated in the context of video streaming. For example, a joint mode selection and video coding algorithm with the consideration of energy consumption is proposed in [8] to maximize the video quality. First, the coding mode for each video frame is selected (I, P or B). Following that, the signal-to-noise ratio (SNR) and the bit error rate (BER) are calculated for each packet in each proposed transmission mode and then the modes are ranked based on their achieved BER. Finally, the transmission mode is assigned to each packet to maximize the quality without exceeding a predefined energy limit. Also, mode selection is considered jointly with power control for variable bit rate (VBR) video streaming in [9] with the goal of maximizing the overall throughput while taking into account buffer underflow and overflow events. Transmission power is determined in each mode and the mode that maximizes the throughput is selected. Obtained results showed that the proposed approach provides better results than a single mode of transmission. Further, mode selection is discussed in [10] where the mode is selected based on the channel quality indicator (CQI). However, only two modes were considered in that work, cellular and dedicated, to eliminate the interference to CUs in the reuse mode. Similarly, using outband D2D, the mode of communication is chosen to maximize the throughput subject to a packet delay constraint in [11] where packets are transmitted directly from the BS or through a relay via a D2D link based on channel quality.

The significant increase in global data traffic is partly due to duplicate downloads of common video files [12]. Hence, some researchers proposed the concept of caching popular video files in mobile devices and sharing them with other users through D2D communication. Generally, caching schemes may be divided into centralized or distributed. In the centralized approach, a central device manages resource allocation and caching decisions [13], [14], while in distributed caching each device independently decides which file to cache [15]–[17]. Golrezaei *et al.* [18], [19], [13] proposed a caching strategy where the cell is split into smaller square clusters. When the BS receives a video streaming request, it checks the availability of the video file in the clusters. If the video content is cached by one of the mobile devices in the cluster, the file is transmitted through D2D communication, otherwise, the request is served by the BS. Considering social behavior, the authors in [20] proposed a scheme for cluster head selection and cluster formation using social characteristics of the users to improve spectrum and energy efficiencies. Similarly, a social-aware caching and resource sharing scheme for video streaming in 5G networks is addressed in [21]. This resource sharing strategy considers the social relationship information of each D2D pair, device information, the target SNR and the information of video files to make the decisions. The goal is to maximize the data rate while considering the target signal-to-interference-plus-noise ratio (SINR) by selecting the optimal number of caching copies of each video file, the femtocells to cache the files, and the optimal resource sharing allocation between the downlink users and the D2D pairs. The obtained results reveal that social relationships significantly impact the performance and a better understanding of social relation improves system performance. Similarly, the authors in [22] discussed D2D-based video caching in 5G networks. They discuss the downlink resource sharing between CUs and D2D pairs for efficient video streaming over multi-hop multi-path channels in a dense area. Assuming the video segments are cached by multiple D2D helpers, the objective is to maximize the throughput of CUs sharing resources with D2D helpers by optimally selecting the CUs and D2D pairs sharing the same resources and allocating which segments to which path.

Clearly, for video-based applications, the minimum QoE requirements of the user should be met, therefore some researchers discussed improving QoE of video streaming over D2D networks. For example, in [23], a QoE-aware resource allocation algorithm is proposed for adaptive D2D video streaming to improve the QoE. The goal is to maximize the video quality taking into account the number of stall events. The video is encoded at multiple quality levels and the appropriate number of layers should be selected for each D2D pair. After calculating the throughput, higher quality is selected for users with large buffer occupancy while for users with empty buffers, lower qualities are assigned. The proposed scheme provides better performance compared to QoE-oblivious resource allocation mechanisms in terms of quality and number of stall events. Similarly, to enhance the user's experience, a QoE-aware power allocation scheme for video transmission is proposed in [24]. The authors formulated an optimization problem aiming at maximizing the video quality of all D2D pairs constrained to a minimum data rate required by each user, a maximum transmit power as well as a certain interference level that can be tolerated. Also, the authors in [25] presented a joint source selection and power control mechanism that aims at maximizing the video quality with low-latency constraints. This is achieved by selecting the best device to transmit the file when multiple devices have cached the demanded video

stream, then the optimal transmission power of each selected device is adjusted to improve video quality. Furthermore, in [26], a location-based mechanism that exploits location information of the transmitter and receiver to choose the optimal route that maximizes the QoE of video streaming is proposed. Real time multi-cast has been also discussed in [27] where the video is split into blocks, which are further divided into smaller chunks. Each user is able to receive the video through two simultaneous links; cellular and D2D. The authors proposed an algorithm to minimize transmission through the BS to save cost while considering the QoE, which was expressed in terms of the average number of received blocks. For an extensive survey on video steaming over D2D networks, the reader is referred to [28].

In this paper, we combine D2D communications underlaying a Long Term Evolution (LTE) network with scalable video coding (SVC) and propose a heuristic algorithm aiming at improving the D2D users' QoE. Particularly, the objective is to perform resource allocation, scheduling and mode selection, among the three modes mentioned earlier, to maximize the sum-rate of the D2D pairs. This is achieved while maintaining continuous video playback by guaranteeing that the buffer occupancy of any D2D pair does not fall below a specific buffer threshold. In more details, the mode that maximizes the D2D rate is selected for transmission taking into account the minimum rate requirements of both the CUs and D2D pairs, the D2D buffers occupancies and video quality. Different from the assumption of assigning one channel to each user in [9] and [29] or dividing the available resources equally between users in [10], here we estimate the number of Resource Blocks (RBs) to be assigned to CUs and D2D pairs according to the required rate and SNR of each user. In addition, instead of constant bit rate streaming (CBR) as in [9], we adopt SVC where each frame is encoded into one base layer and multiple enhancement layers, and the number of layers to be transmitted is determined by a proposed SVC layer selection algorithm. Moreover, unlike the works in [23]–[25], which discussed improving QoE of video streaming over the D2D network via only maximizing the video quality, in this paper, we consider both buffer occupancy and quality of the stream in D2D pairs scheduling. The main contributions of this paper can thus be summarized as follows:

- Design a complete framework for video streaming over D2D-enabled networks that performs resource allocation, scheduling and mode selection for a group of D2D pairs considering the available resources, minimum rate requirements, video quality as well as buffer occupancy.
- Propose a power control algorithm in the reuse mode when CUs and D2D pairs share resources to reduce the interference to CUs after selecting a compatible D2D pair to each CU.
- Moreover, using the achievable rate, the buffer occupancy and frame deadline, we implement an SVC layer selection algorithm to determine the number of

enhancement layers (ELs) to be transmitted to each D2D pair in order to improve the QoE.
- We evaluate the effects of mobility on the performance of the proposed system by adopting a random walk mobility model to represent the unpredictable movement of UEs.

The rest of the paper is organized as follows: Section II presents the system model, the resource allocation and rate calculation mechanisms. Section III describes the optimization problem formulation and presents the heuristic approach for solving it. The proposed algorithm is then evaluated through simulations and its performance is discussed in Section IV before the paper is finally concluded in Section V.

II. SYSTEM MODEL
A. NETWORK DESCRIPTION
In this work, a single LTE cell is considered where the BS is placed at the center of the cell as shown in Fig. 1. A set of C CUs and D D2D pairs are randomly distributed within the cell with the D2D transmitter and receiver belonging to the ith D2D pair labeled as dt_i and dr_i, respectively. The D2D pairs will be using the DL N_R RBs to establish their connections and scheduling is performed every transmission time interval (TTI). Let $h_{i,j}$ denote the channel gain between any two nodes $i, j \in \{dt_d, dr_d, bs, c\}$ where bs denotes the BS, $c \in \mathcal{C} = \{1, 2, \cdots, C\}$ denotes a CU and $d \in \mathcal{D} = \{1, 2, \cdots, D\}$. We assume that channel gains are invariant during one TTI and change independently from one TTI to another. Similarly, let $P_i, i \in \{dt_d, bs, c\}$ denote the corresponding transmit powers of the different nodes in the system. Also, assume that the maximum transmission power of any user (whether a CU or a transmitter of a D2D pair) and the BS are denoted by $P_{cu}^{\max}$ and $P_{bs}^{\max}$, respectively. In the same time, let the minimum transmit power of the transmitter of any D2D pair needed to
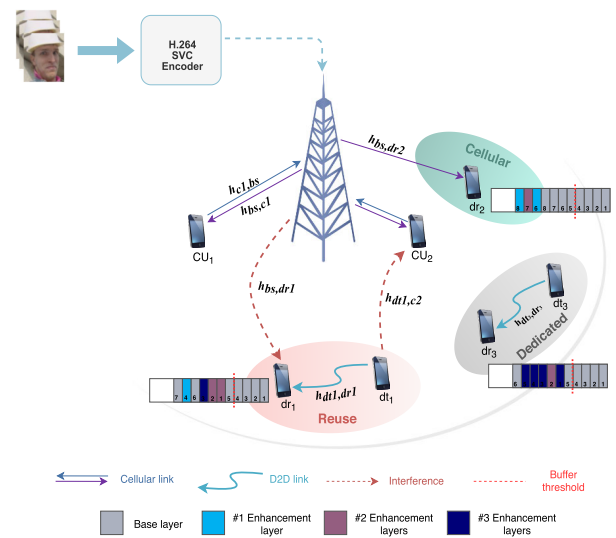

FIGURE 1. System model showing the details of the video streaming process.

properly establish and maintain a D2D connection be denoted by $P_{cu}^{\min}$. As in [29]–[31], equal power allocation is considered at the BS such that all RBs have the same power. Two scenarios will be considered in this work where all nodes are assumed to be either static or will follow a specific mobility model as will be explained later in the sequel.

As shown in Fig. 1, we assume that the BS is connected to a video-on-demand server that buffers and delivers a scalable-encoded video sequence using the H.264/SVC standard. Each video frame f is encoded into one base layer (BL) labeled as B^f with size $b^{f,0}$ bits and L ELs with different quality levels to accommodate bandwidth variations where $E^{f,l}$, the lth EL of the fth frame, has an associated quality $q^{f,l}$ and a size of $b^{f,l}$ bits. The video sequence is intended for the D2D receivers and not the CUs who have their own independent rate requirements based on the nature of their traffic. The video server is responsible for extracting the encoding information, e.g., frame rate, time duration, coding rate and layer quality. No-skip based streaming is adopted; if a BL cannot be received by its deadline, it will not be skipped and a stall event occurs. Assuming the video sequence is cached by the transmitter of a D2D pair and it is willing to collaborate, any D2D receiver can receive the video content either directly from the BS or from its transmitter through a D2D link. Accordingly, and as mentioned earlier, each D2D pair is allowed to operate in one of the three communication modes. In the cellular mode, the user receives the video sequence directly from the BS as conventional CUs while dedicated resources are allocated to each of the D2D links when the dedicated mode is chosen for communication. Further, the D2D pairs share DL resources with CUs in the reuse mode and thus introduce interference to them. Finally, resource allocation, scheduling and mode selection are all performed centrally at the BS. At the receiver side, we assume that the buffer occupancy of each user as well as its channel state information (CSI) are reported to the BS via a feedback mechanism. In the next subsection, we explain how resources are allocated to both the cellular and D2D users in each of the above-mentioned modes.

B. ALLOCATION OF RBs
LTE RBs have to be allocated to both CUs and D2D pairs. Therefore, resource allocation is performed in each mode taking into consideration the channel quality, the minimum rate required by each user as well as the possible mutual interference between the D2D links and CUs, if any. We consider two scheduling algorithms in this work; round robin (RR) and proportional fair (PF) and we describe the implementation details of these algorithms in each of the three modes of operation discussed earlier as follows.

1) CELLULAR MODE
In this mode, a D2D user communicates directly with the BS using DL resources as if it is a regular CU. Hence, the total number of CUs can be considered $K = C + D$ and RBs are assigned to those K CUs using RR and PF schedulers.

In RR scheduling, resources are allocated equally in a cyclic manner, while in PF, the goal is to strike a balance between fairness among the users and maximizing the total network throughput such that the user enjoying favorable channel conditions has a higher priority. Let the rate required by the kth user be R_k^{req}. For the dth D2D pair, this rate is calculated based on the size of the SVC layer being transmitted, viz., $R_d^{\text{req}} = b_d^{f,l}/\text{TTI Duration}$. The number of RBs, n_k^C, to be assigned to the user in the cellular mode can be estimated using [32]

$$n_k^C = \frac{\overline{\text{SNR}}_k}{\frac{1}{K}\sum_{j=1}^{K}\overline{\text{SNR}}_j} \times \frac{R_k^{\text{req}}}{\frac{1}{K}\sum_{j=1}^{K}R_j^{\text{req}}}, \tag{1}$$

where $\overline{\text{SNR}}_k$ is defined as

$$\overline{\text{SNR}}_k = \frac{1}{N_R}\sum_{n=1}^{N_R}\text{SNR}_{k,n} \tag{2}$$

with $\text{SNR}_{k,n}$ being the signal-to-noise ratio of the kth user on the nth RB during a specific TTI. In PF scheduling, RBs are allocated first to those users with the highest PF metric [33], which is defined for the kth user as

$$\psi_k = \frac{R_k^C}{\overline{R}_k^C}, \tag{3}$$

where R_k^C and $\overline{R}_k^C$ denote the instantaneous and average throughputs of the kth user in the cellular mode, respectively. Now, based on the number of available and total number of required RBs, two scenarios are possible. First, the number of available RBs is greater than the total required resources by the K users, i.e., $\sum_{k=1}^{K} n_k \leq N_R$, where clearly all of them (CUs and D2D) will be assigned RBs. The second scenario is when there are not enough RBs for all the users and consequently, those with low PF metric will not be scheduled.

2) DEDICATED MODE
In this mode, actual D2D connections are established without crossing the BS and a number of resources, n_d^D, is allocated to the dth D2D pair according to the RR or PF strategies as before. Two approaches are studied; in the first, a portion of the cellular spectrum is completely dedicated to the D2D users. This portion will be denoted by α and the value of which will be optimized to maximize the total system throughput as will be shown in the simulations result section. It is worth noting that if $\sum_{d=1}^{D} n_d^D > \alpha \times N_R$, some D2D pairs will not be scheduled. In the second approach, the CUs are first scheduled, then the remaining resources will be exclusively used by D2D users. As before, if all RBs are used by CUs, then no dedicated D2D links will be formed.

3) REUSE MODE
In reuse mode, the cellular spectrum is reused by D2D pairs in order to improve the spectral efficiency and consequently, CUs suffer from co-channel interference. We assume that n_d^R RBs will be allocated to the dth pair while allowing it to

reuse the RBs of only one CU. Consequently, for each CU, the most compatible D2D pair must be selected. In this work, we assume that $C \geq D$ in order to guarantee that all D2D pairs will be assigned to CUs. Different metrics in the literature have been proposed to decide on the compatibility of D2D pairs. These are detailed as follows [29], [34]:

- Random: as the name suggests, this metric assigns D2D pairs to CUs in a random fashion.
- Signal-to-Leakage-plus-Noise Ratio (SLNR): SLNR is a measure of how much power D2D links leak into CU ones. In particular, it is defined as the ratio between the desired signal power and the leakage signal power of a D2D link. The SLNR between the dth D2D pair and the cth CU sharing resources can thus be written as

$$\text{SLNR}_{d,c} = \frac{P_{dt_d}|h_{dt_d,dr_d}|^2}{P_{dt_d}|h_{dt_d,c}|^2 + N_o B_{CU}^R} \quad (4)$$

where $P_{dt_d}|h_{dt_d,c}|^2$ is the leakage signal power and B_{CU}^R is the bandwidth allocated to the CU of interest in the reuse mode.

- Exhaustive Search (ES): according to this metric, the cth CU shares its allocated RBs with the D2D pair that results in the highest exhaustive search metric (ESM) defined as

$$\text{ESM}_{d,c} = \frac{R_d^R + R_c^R}{\overline{R}_c^R} \quad (5)$$

where R_d^R and R_c^R are the instantaneous throughputs of the D2D pair and CU sharing the same RBs in the reuse mode, respectively and $\overline{R}_c^R$ is the average CU throughput assuming its RBs are not shared with any D2D pair and hence, no interference is encountered.

In the next subsection, details pertaining to the calculation of the achievable rates R^C, R^D and R^R, for the cellular, dedicated and reuse modes, respectively, based on the number of assigned RBs, level of interference (if any) and according to the LTE standard will be given.

C. LTE STANDARD BASED RATE CALCULATION

After calculating the number of RBs, n^m, to be allocated to each user in the different modes of operation $m \in \mathcal{M} = \{C, D, R\}$, the actual transmission rate R^m can be calculated based on LTE specifications as follows. According to LTE Release 13, a user is assigned a transport block (TB) over the PHY layer to transmit data, the size of which (in bits), referred to as transport block size (TBS), is decided upon based on the modulation and coding scheme (MCS) assigned to the user as well as n^m [35, Section 7.1.7.1]. The MCS is, in turn, chosen based on the CQI periodically reported by the user to the BS by mapping its measured SNR (or SINR) to a specific CQI [36]. TBS is obtained based on a mapping relation between the MCS index I_{MCS}, the TBS index I_{TBS} and n^m. Focusing on the dth D2D pair, its measured SNR in each of the modes of operation discussed before can be calculated as follows:

- Cellular Mode

$$\text{SNR}_d^C = 10 \log_{10} \frac{P_{bs}|h_{bs,dr_d}|^2}{N_o B_d^C}. \quad (6)$$

- Dedicated Mode

$$\text{SNR}_d^D = 10 \log_{10} \frac{P_{dt_d}|h_{dt_d,dr_d}|^2}{N_o B_d^D}. \quad (7)$$

- Reuse Mode

$$\text{SINR}_d^R = 10 \log_{10} \frac{P_{dt_d}|h_{dt_d,dr_d}|^2}{P_{bs}|h_{bs,dr_d}|^2 + N_o B_d^R}. \quad (8)$$

In the above equations, N_o is the additive white Gaussian noise (AWGN) power spectral density and B_d^m denotes the bandwidth allocated to the dth D2D pair in the mth mode given by

$$B_d^m = n_d^m \times BW, \quad (9)$$

where BW is the RB bandwidth in LTE. Finally, in each mode of operation, the achievable rate by the dth D2D pair can be found by

$$R_d^m = f(I_{MCS_d}^m, n_d^m), \quad (10)$$

where $f(I_{MCS_d}^m, n_d^m)$ is a mapping relation between the MCS and the number of RBs.

D. MOBILITY MODEL

As mentioned earlier, the CUs and D2D pairs are either assumed to be static at the same positions initiated randomly or will be allowed to be mobile. The mobility model used in this work to describe the unpredictable movement of mobile nodes is the Random Walk (RW) model [37], [38]. In the RW model, a node starts movement by choosing a speed and a direction from two sets of predefined ranges; $[0, V_{max}]$ and $[0, 2\pi]$, respectively. The mobile node then moves for a fixed time duration, which is equal to 1 TTI, after which a new speed and direction are chosen.

III. PROPOSED QoE-AWARE HEURISTIC ALGORITHM FOR D2D VIDEO STREAMING

A. OPTIMIZATION PROBLEM FORMULATION

As mentioned earlier, the main objective of this work is to maximize the sum-rate of D2D pairs while maintaining continuous video playback through proper resource allocation and mode selection. Based on the rates achieved by each D2D pair in each of the modes discussed earlier, the BS selects one mode of operation for each D2D pair, which results in the highest throughput. Let $\mathbf{x}^m$ be a D-dimensional binary assignment vector for the mth mode such that $x_d^m = 1$ if the dth D2D pair operates in the mth mode. Also, let $\mathbf{A}^m$ be a $N_R \times D$ binary assignment matrix where each element $a_{b,d}^m$ indicates whether or not the bth RB is assigned to the dth D2D pair in the mth mode. Finally, let $\mathbf{p}^m$ be a D-dimensional vector where the dth element represents the amount of transmit power by the

transmitter of the dth D2D pair in the mth mode. The problem can thus be formulated as

$$\max_{\{\mathbf{x}^m\}_{m\in\mathcal{M}},\{\mathbf{A}^m\}_{m\in\mathcal{M}},\{\mathbf{p}^m\}_{m\in\mathcal{M}}} \sum_{d=1}^{D}\sum_{m\in\mathcal{M}} x_d^m R_d^m, \quad (11a)$$

$$\text{where } R_d^m = f\left(I_{MCS_d}^m, \sum_{b=1}^{N_R} a_{b,d}^m\right)$$

$$\text{s.t. } x_d^m \in \{0,1\}, \quad \forall d \in \mathcal{D}, \quad (11b)$$

$$a_{b,d}^m \in \{0,1\}, \ \forall d \in \mathcal{D}, \ \forall b \in \{1,2,\cdots,N_R\} \quad (11c)$$

$$\sum_{m\in\mathcal{M}} x_d^m = 1, \ \forall d \in \mathcal{D}, \quad (11d)$$

$$\sum_{d=1}^{D} a_{b,d}^m = 1, \ \forall m \in \mathcal{M}, \ \forall b \in \{1,2,\cdots,N_R\}, \quad (11e)$$

$$\beta_d \geq \beta_{th}, \quad \forall d \in \mathcal{D}, \quad (11f)$$

$$P_{cu}^{\min} \leq P_{dt_d}^m \leq P_{cu}^{\max}, \quad \forall d \in \mathcal{D}, \ \forall m \in \mathcal{M}, \quad (11g)$$

In the above optimization problem, constraint (11d) indicates that any D2D pair can select exactly one of the three modes of operation. Also, constraint (11e) implies that any resource block can only be allocated to one D2D pair. Buffer occupancy of the D2D pair should not go below the buffer threshold as illustrated by constraint (11f). Finally, constraint (11g) indicates that the transmit power by any D2D transmitter should not exceed the maximum allowable power irrespective of the mode of operation. It should be noted that although the transmit power does not show up directly in the objective function, the rate indirectly depends on it through the SNR (SINR) to I_{MCS} mapping as indicated in Section II-C.

The optimization problem in (11) is formulated as a mixed-integer nonlinear program (MINLP), which is complex to find an optimal solution for. In general, MINLP problems can usually be solved using either exact or heuristic methods. One approach to find the exact solution is to use MINLP solvers such as BONMIN [39]. Other exact methods include solving a smooth continuous relaxation of the problem using methods such as branch-and-bound [40], the standard interior-point [41] or the generalized bender decomposition algorithms [42]. The use of heuristic approaches to find suboptimal solutions with low computational complexity for such problems is a very common approach in the literature as clearly used in [29], [43], [44], and many other works. This is done without having to report the optimal solution due to its prohibitive complexity from a practical point of view. Hence, in this paper, we propose a heuristic approach with lower computational complexity for solving the optimization problem in (11). In the next subsections, we provide details of this heuristic algorithm.

B. POWER CONTROL STEP
In the cellular and dedicated modes discussed before, since no interference is experienced between the CUs and D2D pairs in a single cell scenario like the one considered in this work, the maximum output power of the BS and D2D transmitter can be directly used to maximize the total D2D sum rate. However, in the reuse mode, the D2D pairs interfere with CUs links, which leads to degradation of the CUs throughput. Therefore, a power control algorithm is hereby developed to reduce the interference introduced to CUs while considering the minimum rate requirements of both the CUs and D2D pairs. At each TTI, the rates of each D2D pair and CU sharing the same RBs are first calculated. If the achievable rate by a CU is less than the minimum required while the D2D pair is exceeding the target rate, the power of the D2D transmitter is reduced by a specific step Δ until one of the following happens: either the CU reaches its minimum required rate, the D2D throughput declines to the minimum rate or the minimum transmit power, $P_{cu}^{\min}$, needed to maintain the D2D connection is reached. If the rates of both the D2D pair and the CU exceed the minimum rates, the power of the D2D pair is decreased by the same power step Δ as long as the total sum rate after power reduction is greater than that before the power control step or again, the minimum D2D power is reached. It is worth noting that we choose not to implement power control at the BS level since the interference from the BS to the D2D receiver is small due to the long distance between them. Therefore, applying power control by decreasing the power of BS will not result in throughput enhancement. The proposed power control mechanism is summarized in Algorithm 1.

C. MODE SELECTION STEP
The mode selection step proposed herein depends on the resource allocation and scheduling process taking into account channel conditions, the required rate by the D2D user, their instantaneous buffer occupancies as well as the frame quality (whenever possible). This is done in order to achieve a balance between video quality and continuity. Each D2D receiver sends feedback to the BS indicating its buffer occupancy and CQI. After allocating resources to the CUs and D2D pairs in each mode, the BS starts calculating the instantaneous rate of each D2D pair using its CQI and the mapping function in (10). In the cellular mode, the D2D pairs are scheduled as conventional CUs using RR or PF scheduler while in the reuse mode, the most compatible D2D pair is selected for each CU using ESM. In the dedicated mode, the BS compares the instantaneous buffer occupancy β_d of the dth D2D receiver to a specific buffer threshold β_{th} to determine the urgency of transmitting a video frame to this receiver. The D2D pairs are hence categorized into two groups; D2D pairs with buffer underflow and others with buffer occupancies that are equal or above the predefined threshold. Starting with the D2D pairs that are underflowing, the BS schedules them while giving higher priority to the pair with the lowest occupancy to avoid buffer starvation.

Algorithm 1 Power Control Algorithm for the Reuse Mode

1: **Initialization**: $P_{bs} = P_{bs}^{\max}$;
2: **for** each TTI and a CU c with a required minimum rate R_c^{req} sharing RBs with a D2D pair d with a required minimum rate R_d^{req} **do**
3:     Set $P_{dt_d} = P_{cu}^{\max}$;
4:     Calculate the D2D pair rate R_d using (10);
5:     Calculate the CU rate R_c using (10);
6:     **if** $R_d > R_d^{\text{req}}$ & $R_c < R_c^{\text{req}}$ **then**
7:         **repeat**
8:             Update $P_{dt_d} = P_{dt_d} - \Delta$;
9:             Calculate the D2D pair new rate R'_d using (10);
10:             Calculate the CU new rate R'_c using (10);
11:         **until** $R'_d = R_d^{\text{req}} \,||\, R'_c = R_c^{\text{req}} \,||\, P_{dt_d} = P_{cu}^{\min}$
12:     **else if** $R_d > R_d^{\text{req}}$ & $R_c > R_c^{\text{req}}$ **then**
13:         **repeat**
14:             Update $P_{dt_d} = P_{dt_d} - \Delta$;
15:             Calculate the D2D pair new rate R'_d using (10);
16:             Calculate the CU new rate R'_c using (10);
17:         **until** $R'_c + R'_d > R_c + R_d \,||\, R'_d = R_d^{\text{req}} \,||\, P_{dt_d} = P_{cu}^{\min}$
18:     **end if**
19: **end for**

If there are available RBs after scheduling those underflowing users, the BS schedules the D2D pairs with $\beta_d > \beta_{th}$ and gives higher priority to those pairs with higher quality and less number of required resources. This can be done based on a specific metric φ_d for the dth D2D pair, which can be calculated as

$$\varphi_d = \left(\frac{q_d^{f,l}}{\frac{1}{D}\sum_{d=1}^{D} q_d^{f,l}}\right) \times \left(\frac{n_d}{\frac{1}{D}\sum_{d=1}^{D} n_d}\right)^{-1} \qquad (12)$$

where $q_d^{f,l}$ is, as mentioned earlier, the quality of the lth layer of the fth frame transmitted to the dth user.

Finally, for each pair, the mode with the maximum throughput is selected for transmission. It is worth mentioning here that in the cellular mode, the buffer occupancy cannot be considered as a scheduling metric because if the D2D pairs are scheduled first based on β_d and the remaining resources are assigned to CUs, this will affect the performance of CUs and only D2D pairs might be scheduled. Similarly, for the reuse mode, since it is assumed that the number of CUs is greater than that of the D2D pairs as in Section II-B, all D2D users will be scheduled in reuse mode and thus, it is not needed to prioritize D2D pairs and determine which D2D pair to be served first.

D. SVC LAYER SELECTION

Given the achievable rate (as determined by the mode selection step), frame deadline and buffer occupancy of each D2D receiver, the number of video layers to be delivered can now be determined. Specifically, if for the dth D2D receiver, $\beta_d < \beta_{th}$, a BL is transmitted to maintain the continuity. On the other hand, if $\beta_d > \beta_{th}$ and there are available remaining RBs, it is possible to send either a BL or ELs to improve the video quality. We propose to first estimate the time by which the ELs will be received. We then compare the estimated buffer occupancy at that time to β_{th}. If the buffer occupancy after transmitting the ELs exceeds β_{th}, the ELs are transmitted. Otherwise, the BL of the next frame is transmitted to maintain video continuity. The buffer occupancy β_d^e of the dth D2D receiver after transmitting the ELs is estimated by subtracting the number of frames that will be played during the transmission of the ELs from the current buffer occupancy at the ith TTI, viz.,

$$\beta_d^e = \beta_d(i) - \frac{\sum_{l=1}^{L'} b_d^{f,l}}{\overline{R}_d} \times f_p, \qquad (13)$$

where $L' \leq L$ is the number of ELs that will be transmitted and f_p represents the playback rate in frames/sec. It is important to note that for an EL to be transmitted, all the lower-quality layers on which it depends need to have already been sent as well. Noting that L' ELs need to be received before the frame display deadline, this imposes the following condition that can be used to obtain the value of L'

$$t_i + \frac{\sum_{l=1}^{L'} b_d^{f,l}}{\overline{R}_d} < \text{display time of frame } f, \qquad (14)$$

where t_i is the start time of the ith TTI. The steps of the SVC layer selection are summarized in Algorithm 2 where

---

Algorithm 2 SVC Layer Selection Algorithm

1: **Input:** $R_d^{m_d}, b_d^{f,l}, f_p, \beta_d$;
2: **for** each TTI i **do**
3:     **for** each $d \in \mathcal{D}$ **do**;
4:         **if** $R_d^{m_d}(i) > R_d^{\text{req}}(i)$ **then**
5:             **if** $\beta_d(i) < \beta_{th}$ **then**
6:                 $S_d(i+1) = B^{f+1}$;
7:             **else**
8:                 Find the number of ELs (L') that can be transmitted before the deadline using (14);
9:                 Estimate the buffer occupancy after ELs transmission using (13);
10:                 **if** $\beta_d^e < \beta_{th}$ **then**
11:                     $S_d(i+1) = B^{f+1}$;
12:                 **else**
13:                     $S_d(i+1) = \{E^{f,l}\}_{l=1}^{L'}$;
14:                 **end if**
15:             **end if**
16:         **else**
17:             Transmit the next segment of currently transmitting layer;
18:         **end if**
19:     **end for**
20: **end for**

S_d indicates an SVC layer to be transmitted (whether BL or EL) to the dth D2D receiver. Also, the details of the overall heuristic algorithm performed by the BS, which involves resource allocation, power control, mode selection and SVC layer selection is illustrated in Algorithm 3.

Algorithm 3 Mode Selection Algorithm

1: **Input**: $N_R, \alpha, \beta_{th}, \beta_d, b^{f,l}, \text{CQI}_k$;
2: **Initialization**: $\mathcal{U} = \mathcal{N} = \phi, N_R^D = \alpha \times N_R, N_R^C = (1-\alpha) \times N_R, f = 1, R_d^{\text{req}}(i=1) = \frac{b_d^{f=1,0}}{1m}$;
3: **for** each TTI i **do**
4:     **%Dedicated Mode**:
5:     **for** $\forall c \in \mathcal{C}$ **do**;
6:         Calculate $n_c(i)$ using $R_c^{\text{req}}(i), N_R^C$ in (1);
7:         Find $R_c^D(i)$ using (10);
8:     **end for**
9:     Schedule the users in $\mathcal{C}$ according to the RR or PF metric in (3);
10:     **for** $\forall d \in \mathcal{D}$ **do**;
11:         Calculate n_d using $R_d^{\text{req}}(i), N_R^D$ in (1);
12:         Find $R_d^D(i)$ using (10);
13:         **if** $\beta_d(i) < \beta_{th}(i)$ **then**
14:           $\mathcal{U} = \mathcal{U} \cup \{d\}$;
15:         **else**
16:           $\mathcal{N} = \mathcal{N} \cup \{d\}$;
17:         **end if**
18:     **end for**
19:     Schedule the users in $\mathcal{U}$ in ascending order of their $\beta_d(i)$;
20:     Update $N_R^D = N_R^D - \sum_{u \in \mathcal{U}} n_u$;
21:     **if** $N_R^D > 0$ **then**
22:         Calculate $\varphi_n, \forall n \in \mathcal{N}$ using (12);
23:         Schedule the users in $\mathcal{N}$ in descending order of their φ_d;
24:     **end if**
25:     **%Cellular Mode**:
26:     Let $K = D + C$ and $\mathcal{K} = \{1, 2, 3, \ldots, K\}$;
27:     **for** $\forall k \in \mathcal{K}$ **do**;
28:         Calculate $n_k(i)$ using $R_k^{\text{req}}(i), N_R^C$ in (1);
29:         Find $R_k^C(i)$ using (10);
30:     **end for**
31:     Schedule the users in $\mathcal{K}$ according to the RR or PF metric in (3);
32:     **%Reuse Mode**:
33:     **for** $\forall c \in \mathcal{C}$ **do**;
34:         Calculate $n_c(i)$ using $R_c^{\text{req}}(i), N_R^C$ in (1);
35:         Find $R_c^R(i)$ using (10);
36:     **end for**
37:     Schedule the users in $\mathcal{C}$ according to the RR or PF metric in (3);
38:     **for** $\forall d \in \mathcal{D}$ **do**;
39:         Select a compatible CU for d D2D pair using the ES metric in (5);
40:         Perform power control as in Algorithm 1;
41:         Find $R_d^R(i)$ using (10);
42:     **end for**
43:     **%Mode Selection**:
44:     **for** $\forall d \in \mathcal{D}$ **do**;
45:         $m_d = \arg \max R_d^m$;
46:     **end for**
47:     Perform layer selection as in Algorithm 2;
48: **end for**

IV. SIMULATION RESULTS

In this section, we present the simulation results for the QoE-aware heuristic mode selection algorithm proposed in this work for D2D video streaming. As described in Section II, a single LTE cell is considered with the CUs distributed randomly in the cell. The D2D pairs are also assumed to be placed randomly around the BS, however, the D2D receivers are restricted to be within a disc around their transmitters with a maximum radius r. The urban microcell channel models introduced in [45] and [46] are used to model the different communication links where distance-based path loss, shadow fading modeled by a log-normal random variable as well as Rayleigh fading are all taken into account. Specifically, the D2D links pathloss model is given by $40 \log_{10}(d \text{ [km]}) + 49 + 30 \log_{10}(f \text{ [MHz]})$ and the BS-CU links pathloss model is given by $36.7 \log_{10}(d \text{ [m]}) + 22.7 + 26 \log_{10}(f \text{ [GHz]})$. The CUs' required rate is generated randomly between [0,1] Mbps based on service requirements as in [47], [48]. MATLAB has been used as the simulation environment and the different simulation parameters are summarized in Table 1. Based on the LTE standard [35], we have set the transmission power of the BS and UEs to the maximum permissible values. Also, the overall effect of TBs loss on video frame quality has been taken into account using the packet-layer model proposed in [49]. In this work, the quality of the received videos in different modes is presented in terms of the peak signal-to-noise-ratio (PSNR) metric. While the PSNR might not be sufficient for QoE evaluation, it still remains one of the most, if not the most, commonly used metric to evaluate the quality of compressed videos in the literature. Many works have also claimed that it indeed correlates positively with subjective quality methods like the mean opinion score (MOS) when the video content and codec are fixed [50], which is exactly the

TABLE 1. Simulation parameters.

Parameter	Value
Cell radius	500 m
Distance between D2D Tx and Rx (r)	15 m
UL frequency	2.56 GHz
DL frequency	2.68 GHz
TTI Duration	1 ms
Maximum transmit power of BS ($P_{bs}^{\max}$)	46 dBm
Maximum transmit power of D2D Tx ($P_{cu}^{\max}$)	23 dBm
Number of DL RBs (N_R)	100
Shadow fading standard deviation for D2D links	12 dB
Shadow fading standard deviation for cellular links	4 dB
CUs' required rates	between [0,1] Mbps

case considered in our work. Although MOS might also be used, it suffers from its own limitations including the need to control the viewing conditions such as the luminance of the used screens, the chromaticity of the background and the background room illumination in addition to proper selection of the viewers and the need to handle their participation in an ethical and professional manner, which is a very tedious process (the reader is referred to ITU recommendations P.910 [51], P.913 [52] and P.920 [53] for the specifics of such process). In summary, we believe that the PSNR is a useful and easily evaluated metric and hence, will be used in this paper. The different simulation results are presented in the following subsections.

A. RESOURCE ALLOCATION, SCHEDULING AND POWER CONTROL

First, we evaluate the performance of each communication mode independently without mode selection. Clearly, each mode of operation has its own parameters that can affect its performance. For example, the type of schedulers (whether RR or PF) in all of the modes, how the BW is dedicated to the D2D users in the dedicated mode as well as the power control algorithm in the reuse mode. Starting with the cellular mode, the cumulative distribution function (CDF) of the average throughput is shown in Fig. 2 assuming $C = 50$ and $D = 15$ users. The performance of the two schedulers, PF and RR, are also compared and it can be clearly seen that PF scheduling provides better performance. Hence, it is selected as the scheduling technique in all subsequent results.
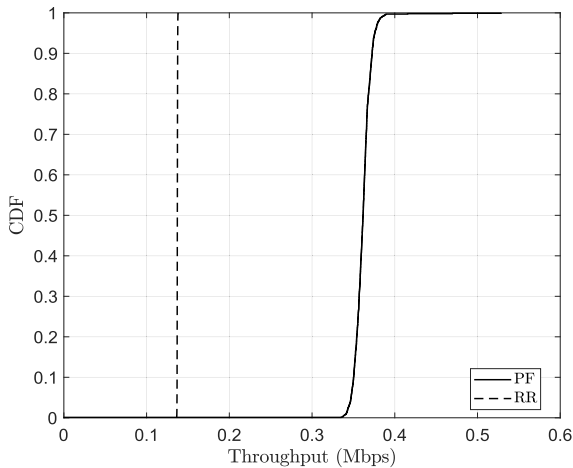

FIGURE 2. CDF of the average throughput in the cellular mode in the DL direction.

As for the dedicated mode, two approaches are examined as mentioned before using the PF scheduler. In the first one (referred to as PF-1), a portion α of the RBs is allocated to the D2D pairs. The network throughput versus α for different number of CUs and $D = 25$ is presented in Fig. 3. From the figure, as the number of CUs increases, the value of α
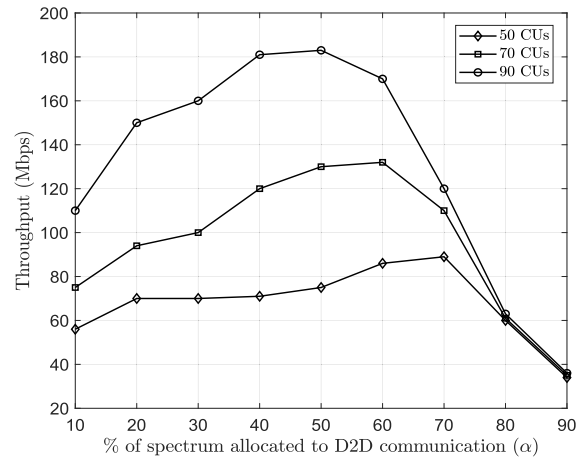

FIGURE 3. Total throughput in the dedicated mode for different values of α and different numbers of CUs with $D = 25$.

that maximizes the throughput declines. Although not shown in the figure, our investigations revealed that changing the number of D2D pairs does not affect the value of α needed for maximum throughput given the current network setup. For the given simulation parameters, the value of α that maximizes the system throughput was found to be around 70%. The second approach (referred to as PF-2) provides inferior performance because only the remaining resources after scheduling all CUs are dedicated to D2D communications. This is confirmed by the results illustrated in Fig. 4.
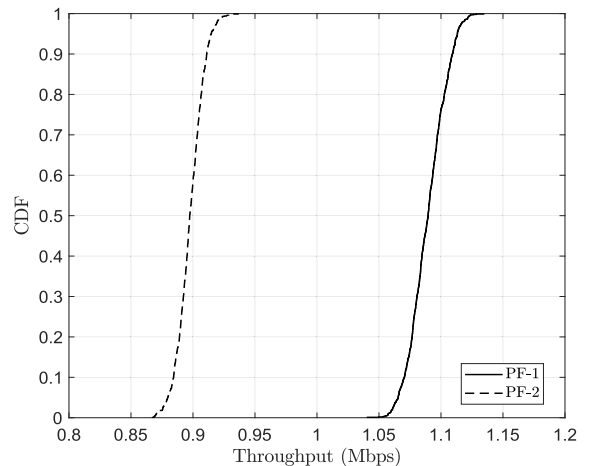

FIGURE 4. CDF of the average throughput in the dedicated mode using PF-1 with $\alpha = 70\%$ and PF-2.

As for the reuse mode, the power control algorithm (Algorithm 1) is implemented to reduce the interference from D2D pairs to CUs. Reducing the interference introduced by D2D pairs sharing the same resources with the CUs increases the throughput of those CUs, which in turn, enhances the total throughput. This is evident from Fig. 5, which also shows that the ES metric gives the best performance.
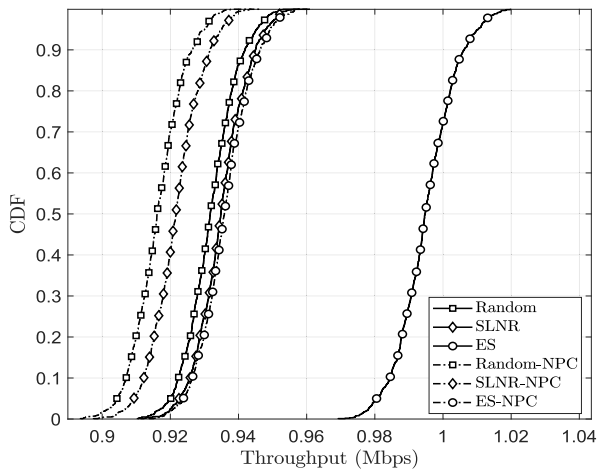
FIGURE 5. CDF of the average throughput of all UEs assuming the reuse mode with and without power control. NPC stands for no power control.

B. MODE SELECTION

In this section, results pertaining to the performance of the mode selection algorithm are presented. The "Crew" video sequence has been used from the trace file library available online in [54]. The chosen sequence has 177 frames at a 352×288 resolution and is encoded into 1 BL and 3 ELs with a playback rate of 15 fps. SNR scalability is adopted and the group of pictures (GOP) structure is IBBBPBBBPBBBPBBB, with hierarchical dependency between the frames of each layer. The buffer threshold β_{th} is set to 4 frames and $D = 20$.

1) THROUGHPUT, PSNR AND BUFFER OCCUPANCY

First, Fig. 6 depicts the average throughput of the D2D pairs for the "Crew" sequence assuming the three modes of transmission compared to when mode selection is used. As can be seen, the mode selection approach provides the highest
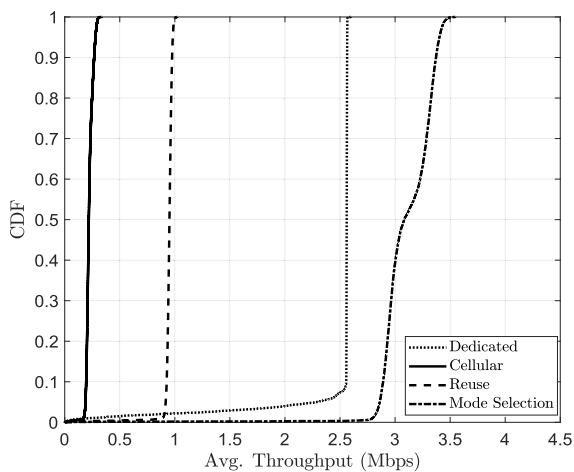
FIGURE 6. CDF of the average throughput of the D2D pairs streaming the "Crew" sequence in mode selection and the three individual modes of operation.

throughput simply because the mode that maximizes the throughput in each time slot is selected. In addition, in mode selection, the user transmits data in each time slot, unlike in dedicated and reuse modes, where some users may not be assigned resources when they have low priorities. Although the reuse mode also guarantees transmission in each slot, since we assume that the number of CUs is greater than D2D pairs, interference from CUs results in throughput degradation and mode selection still provides better performance. Fig. 7 then compares the average PSNR of the D2D pairs in different modes to that after encoding. The degradation in quality is due to the loss of ELs, either due to the loss of TBs or discarding a layer when it misses its deadline. Clearly, the values of the average PSNR using mode selection are higher compared to the other three modes with the dedicated mode coming next, while the cellular mode maintains an average quality value that is close to the BLs quality, which is estimated at a PSNR of 27.2 dB. Furthermore, mode selection provides high performance in terms of continuity of the playback as the instantaneous buffer occupancy is much higher than the predefined threshold as clearly seen in Fig. 8. The values of buffer occupancy in the cellular and reuse modes are, on the other hand, close to β_{th}. When the buffer occupancy is larger than the threshold, this allows the transmission of more ELs, which in turn improves the video quality.
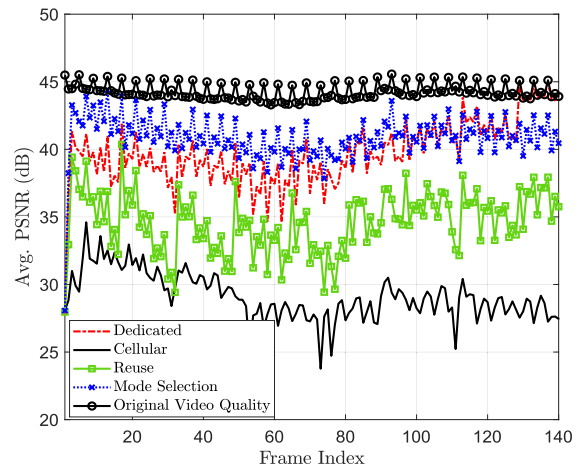

FIGURE 7. Average PSNR of the D2D pairs streaming the "Crew" sequence in mode selection and the three individual modes of operation.

2) EFFECT OF NUMBER OF D2D PAIRS

The effect of changing the number of D2D pairs is investigated by changing the number of D2D pairs from 10 to 40. As the number of D2D pairs increases, the average throughput decreases in all modes as illustrated in Fig. 9. As a consequence, the average PSNR also decreases as can be clearly noticed in Fig. 10 because when the throughput declines, less number of ELs can be transmitted within the deadline according to (14). It is worth mentioning that when the number of D2D pairs is only 10, the difference between
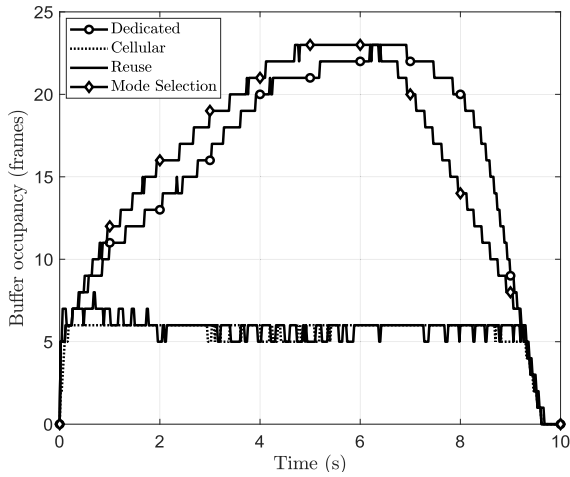
FIGURE 8. Average buffer occupancy for mode selection and the three modes of operation for D2D users streaming the "Crew" sequence.
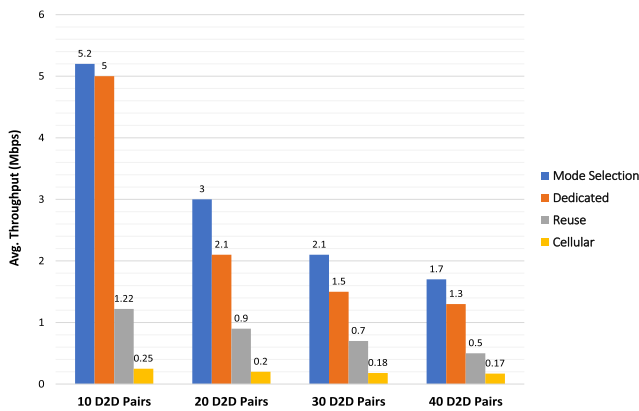

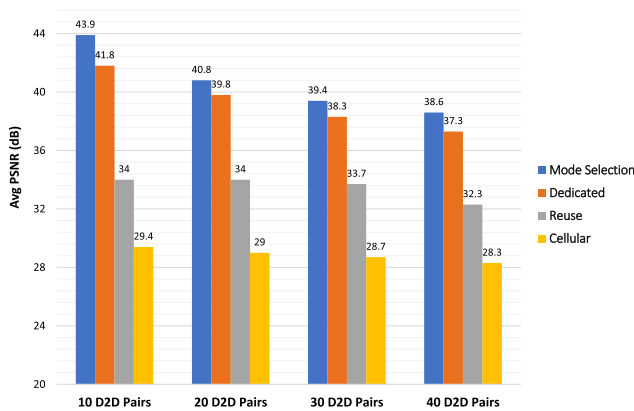**FIGURE 9.** Average throughput for different number of D2D pairs.

FIGURE 10. Average PSNR for different number of D2D pairs.

the throughput achieved by the dedicated and mode selection approaches is about 0.2 Mbps only, which is quite small. This is because for a few number of D2D pairs, all users can be scheduled and enough resources will be assigned to each user to achieve the required rate.

3) EFFECT OF BUFFER THRESHOLD VALUE

As shown in Fig. 11, there is a steady decline in video quality (in terms of the PSNR) as the buffer threshold increases for all modes of operation. The reason behind that is as the buffer threshold increases, more ELs will not satisfy the condition in (14) and will consequently miss the deadline, which results in quality degradation. Therefore, for a small number of users, the buffer threshold can be set to a small value to get better quality. In contrast, in a dense area or when there is a limited number of resources, the buffer threshold should be set to a high value to avoid starvation when no channels are assigned to the user.
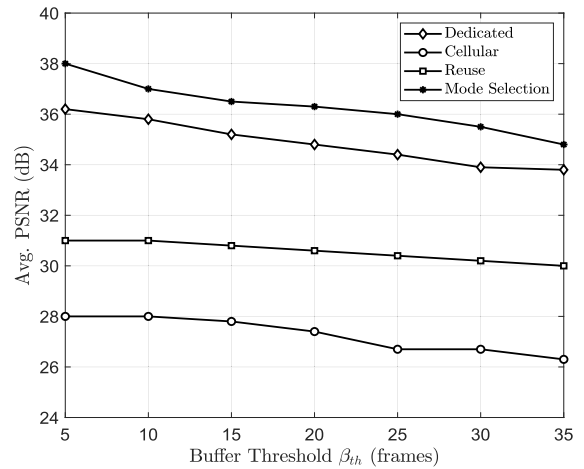

FIGURE 11. Average PSNR vs. buffer threshold.

4) EFFECT OF MOBILITY

In all the results presented so far, we examined the performance of the proposed approach when all the nodes are static at their initial positions randomly chosen at the start of the simulations. We now investigate the impact of user mobility as defined by the RW model in Section II-D on the system performance using two metrics; video quality (again in terms of PSNR) and the contact probability defined as the probability that a user is within the D2D range and data is received directly from the neighboring device using either the reuse or dedicated modes [55]. These metrics are depicted in Figs. 12 and 13 vs. mobility intensity represented in terms of the users' velocities in km/h. This study can be thought of as an investigation into the performance of video streaming in a vehicle-to-vehicle (V2V) communications scenario. From the figures, it can be seen that up to 40 km/h, the effect of mobility is not severe on both PSNR and contact probability. When the speed exceeds 40 km/h, however, the PSNR starts to drop gradually to reach about 35 dB at 130 km/h as shown in Fig. 12. Similarly, as illustrated in Fig. 13, the contact probability declines to 77% and some D2D pairs will start to receive video directly from the BS. One can conclude that low mobility does not affect the system performance, while for high-speed scenarios, video quality significantly degrades.
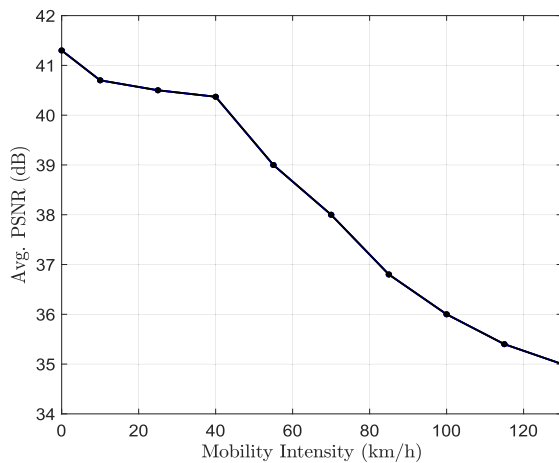
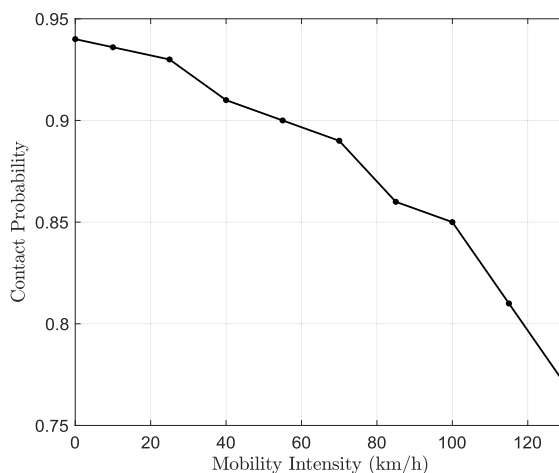**FIGURE 12.** Average PSNR vs. mobility intensity.

FIGURE 13. Contact probability vs. mobility intensity.

This is because proximity-based services lose their efficiency in providing D2D communications with a high capacity and the only alternative is to communicate through the cellular system.

V. CONCLUSION

The main focus of this paper was to propose a complete framework for video streaming over D2D networks. Resource allocation, mode selection and power control for multiple D2D pairs have been addressed with the goal of maximizing the throughput of the D2D users while considering the quality of experience requirements represented by video quality and continuity. Scalable video coding was employed to provide the video with high quality to users with good channel conditions while maintaining the basic video quality for users who experience bad channel conditions. Dynamic nodes instead of static ones have also been studied to examine the effect of mobility on mode selection and video quality. Simulation results revealed that the proposed mode selection scheme provides better performance in terms of average

throughput, quality and in maintaining the video playback continuity compared to a single mode of operation. Furthermore, we found that for low mobility scenarios, the degradation in video quality and reduction in contact probability is insignificant when compared to high mobility scenarios.

A final concluding remark is due here. Although this work has been presented in the context of 4G LTE networks, it still can be applied to 5G New Radio (NR) networks. This is because the 5G NR proposed access technology uses an RB grid that is very similar to that of LTE networks, where the spectrum is divided into a number RBs per time slot. Moreover, the operating frequency and frame structure in 5G NR are both similar to the one adopted in LTE. Specifically, the proposed approach can be implemented at band n7 in the frequency range 1 allocated to 5G NR [56] with slight modifications such as the number of resource blocks per time slot and modulation techniques. The core of the work, however, will not change.

REFERENCES

[1] L. Xu, C. Jiang, Y. Shen, T. Q. S. Quek, Z. Han, and Y. Ren, "Energy efficient D2D communications: A perspective of mechanism design," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7272–7285, Nov. 2016.

[2] Sandvice. (2019). *The Global Internet Phenomena Report*. [Online]. Available: https://www.sandvine.com/press-releases/sandvine-releases-2019-global-i%nternet-phenomena-report

[3] L. Aymen, B. Ye, and T. M. T. Nguyen, "Offloading performance evaluation for network coding-based cooperative mobile video streaming," in *Proc. 7th Int. Conf. Netw. Future (NOF)*, Nov. 2016, pp. 1–5.

[4] M. Sun, Y. Zou, M. Qian, J. Zhu, and Q. Wang, "Exploiting multiradio cooperation in heterogeneous wireless networks for absolute security against eavesdropping," in *Proc. ITU Kaleidoscope Challenges Data-Driven Soc. (ITU-K)*, Nov. 2017, pp. 1–6.

[5] Q. Zhao, Y. Mao, S. Leng, and H. Wang, "Multimedia traffic placement under 5G radio access techniques in indoor environments," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 3891–3896.

[6] 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, 4th Quart., 2014.

[7] F. Jameel, Z. Hamid, F. Jabeen, S. Zeadally, and M. A. Javed, "A survey of device-to-device communications: Research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2133–2168, 3rd Quart., 2018.

[8] Q. Wang, W. Wang, S. Jin, H. Zhu, and N. T. Zhang, "Joint coding mode and multi-path selection for video transmission in D2D-underlaid cellular network with shared relays," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2014, pp. 1–6.

[9] C. Ye, M. C. Gursoy, and S. Velipasalar, "Power control and mode selection for VBR video streaming in D2D networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2018, pp. 1–6.

[10] M. Klugel and W. Kellerer, "Leveraging the D2D-gain: Resource efficiency based mode selection for Device-to-Device communication," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–7.

[11] A. Asadi, V. Mancuso, and R. Gupta, "An SDR-based experimental study of outband D2D communications," in *Proc. INFOCOM 35th Annu. IEEE Int. Conf. Comput. Commun.*, Apr. 2016, pp. 1–9.

[12] A. S. Daghal and Q. Z. Ahmed, "Video content delivery using multiple devices to single device communications," in *Proc. IEEE 83rd Veh. Technol. Conf. (VTC Spring)*, May 2016, pp. 1–5.

[13] N. Golrezaei, M. Ji, A. F. Molisch, A. G. Dimakis, and G. Caire, "Device-to-device communications for wireless video delivery," in *Proc. Conf. Rec. Forty 6th Asilomar Conf. Signals, Syst. Comput. (ASILOMAR)*, Nov. 2012, pp. 930–933.

[14] L. Ye, Y. Hu, H. Tan, and L. Wang, "QoE-driven cross-layer design for device-to-device video delivery," in *Proc. IEEE 22nd Int. Conf. Parallel Distrib. Syst. (ICPADS)*, Dec. 2016, pp. 120–127.

[15] R. Amer, M. M. Butt, M. Bennis, and N. Marchetti, "Delay analysis for wireless D2D caching with inter-cluster cooperation," in *Proc. GLOBECOM IEEE Global Commun. Conf.*, Dec. 2017, pp. 1–7.

[16] N. Golrezaei, A. G. Dimakis, and A. F. Molisch, "Scaling behavior for device-to-device communications with distributed caching," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4286–4298, Jul. 2014.

[17] A. C. Kazez and T. Girici, "Interference-aware distributed device-to-device caching," in *Proc. IEEE Int. Black Sea Conf. Commun. Netw. (BlackSeaCom)*, Jun. 2017, pp. 1–5.

[18] N. Golrezaei, A. G. Dimakis, and A. F. Molisch, "Device-to-device collaboration through distributed storage," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2012, pp. 2397–2402.

[19] N. Golrezaei, P. Mansourifard, A. F. Molisch, and A. G. Dimakis, "Base-station assisted device-to-device communications for high-throughput wireless video networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 7, pp. 3665–3676, Jul. 2014.

[20] G. Zhang, K. Yang, and H.-H. Chen, "Socially aware cluster formation and radio resource allocation in D2D networks," *IEEE Wireless Commun.*, vol. 23, no. 4, pp. 68–73, Aug. 2016.

[21] M. Bui *et al.*, "Social-aware caching and resource sharing maximized video delivery capacity in 5G ultra-dense networks," *Mobile Netw. Appl.*, Jul. 2019, doi: 10.1007/s11036-019-01316-5.

[22] Q.-N. Tran, N.-S. Vo, M.-P. Bui, V.-C. Phan, Z. Kaleem, and T. Q. Duong, "Resource sharing and segment allocation optimized video streaming over multi-hop multi-path in dense D2D 5G networks," in *Proc. Int. Conf. Ind. Netw. Intell. Syst.* Cham, Switzerland: Springer, 2019, pp. 26–39.

[23] H. Zhu, Y. Cao, W. Wang, B. Liu, and T. Jiang, "QoE-aware resource allocation for adaptive device-to-device video streaming," *IEEE Netw.*, vol. 29, no. 6, pp. 6–12, Nov. 2015.

[24] N. Eshraghi, V. Shah-Mansouri, and B. Maham, "QoE-aware power allocation for device-to-device video transmissions," in *Proc. IEEE 27th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Sep. 2016, pp. 1–5.

[25] Q. Wang, W. Wang, S. Jin, H. Zhu, and N. T. Zhang, "Quality-optimized joint source selection and power control for wireless multimedia D2D communication using stackelberg game," *IEEE Trans. Veh. Technol.*, vol. 64, no. 8, pp. 3755–3769, Aug. 2015.

[26] A. Roy, P. De, and N. Saxena, "Location-based social video sharing over next generation cellular networks," *IEEE Commun. Mag.*, vol. 53, no. 10, pp. 136–143, Oct. 2015.

[27] N. Abedini, S. Sampath, R. Bhattacharyya, S. Paul, and S. Shakkottai, "Realtime streaming with guaranteed QoS over wireless D2D networks," in *Proc. 14th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, Aug. 2013, pp. 197–206.

[28] I. Ahmed, M. H. Ismail, and M. S. Hassan, "Video transmission using device-to-device communications: A survey," *IEEE Access*, vol. 7, pp. 131019–131038, 2019.

[29] R. L. Batista, C. F. M. Silva, T. F. Maciel, J. M. da Silva, and F. R. P. Cavalcanti, "Joint opportunistic scheduling of cellular and Device-to-Device communications," *J. Commun. Inf. Syst.*, vol. 32, no. 1, pp. 62–73, 2017.

[30] Y. Zhang and G. Liu, "Fine granularity resource allocation algorithm for video transmission in orthogonal frequency division multiple access system," *IET Commun.*, vol. 7, no. 13, pp. 1383–1393, Sep. 2013.

[31] F. Li, P. Ren, and Q. Du, "Joint packet scheduling and subcarrier assignment for video communications over downlink OFDMA systems," *IEEE Trans. Veh. Technol.*, vol. 61, no. 6, pp. 2753–2767, Jul. 2012.

[32] H. A. M. Ramli, K. Sandrasegaran, R. Basukala, R. Patachaianand, M. Xue, and C.-C. Lin, "Resource allocation technique for video streaming applications in the LTE system," in *Proc. 19th Annu. Wireless Opt. Commun. Conf. (WOCC)*, May 2010, pp. 1–5.

[33] R. D. Trivedi and M. Patel, "Comparison of different scheduling algorithm for LTE," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 4, no. 5, pp. 334–339, 2014.

[34] I. Ahmed, M. H. Ismail, and M. S. Hassan, "A resource allocation based mode selection scheme in D2D communication," in *Proc. Int. Conf. Commun., Signal Process., their Appl. (ICCSPA)*, Mar. 2019, pp. 1–5.

[35] *Physical Layer Procedures*, document 3GPP TS 36.213, 3GPP, 2016.

[36] Y.-W. Kuo and L.-D. Chou, "Power saving scheduling scheme for Internet of Things over LTE/LTE-advanced networks," *Mobile Inf. Syst.*, vol. 2015, pp. 1–11, Dec. 2015.

[37] C. Kumar, "Evaluation of MANET performance in presence of obstacles," *Int. J. Ad Hoc, Sensor Ubiquitous Comput.*, vol. 3, no. 3, pp. 37–46, Jun. 2012.

[38] A. Orsino, A. Ometov, G. Fodor, D. Moltchanov, L. Militano, S. Andreev, O. N. C. Yilmaz, T. Tirronen, J. Torsner, G. Araniti, A. Iera, M. Dohler, and Y. Koucheryavy, "Effects of heterogeneous mobility on D2D-and drone-assisted mission-critical MTC in 5G," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 79–87, Feb. 2017.

[39] C.-P. Chien, Y.-C. Chen, and H.-Y. Hsieh, "Exploiting spatial reuse gain through joint mode selection and resource allocation for underlay device-to-device communications," in *Proc. 15th Int. Symp. Wireless Pers. Multimedia Commun.*, 2012, pp. 80–84.

[40] D. Della Penda, L. Fu, and M. Johansson, "Energy efficient D2D communications in dynamic TDD systems," *IEEE Trans. Commun.*, vol. 65, no. 3, pp. 1260–1273, Mar. 2017.

[41] T. D. Hoang, L. B. Le, and T. Le-Ngoc, "Energy-efficient resource allocation for D2D communications in cellular networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 9, pp. 6972–6986, Sep. 2016.

[42] H. Meshgi, D. Zhao, and R. Zheng, "Optimal resource allocation in multicast device-to-device communications underlaying LTE networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 8357–8371, Sep. 2017.

[43] T. Han, R. Yin, Y. Xu, and G. Yu, "Uplink channel reusing selection optimization for device-to-device communication underlaying cellular networks," in *Proc. IEEE 23rd Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2012, pp. 559–564.

[44] M. Zulhasnine, C. Huang, and A. Srinivasan, "Efficient resource allocation for device-to-device communication underlaying LTE network," in *Proc. IEEE 6th Int. Conf. Wireless Mobile Comput., Netw. Commun.*, Oct. 2010, pp. 368–375.

[45] *Guidelines for Evaluation of Radio Interface Technologies for IMT-Advanced*, Standard ITU-R M.2135-1, ITU, Dec. 2009.

[46] H. Xing and S. Hakola, "The investigation of power control schemes for a device-to-device communication integrated into OFDMA cellular system," in *Proc. 21st Annu. IEEE Int. Symp. Pers., Indoor Mobile Radio Commun.*, Sep. 2010, pp. 1775–1780.

[47] P. Baptista, "Impact of HSUPA implementation on UMTS capacity and cell coverage," M.S. thesis, IST-UTL, Lisbon, Portugal, 2007.

[48] D. Ladeira, P. Costa, L. M. Correia, and L. Santo, "Planning of UMTS cellular networks for data services based on HSDPA," in *Proc. 13th Eur. Wireless Conf.*, Paris, France, Apr. 2007.

[49] F. Yang, J. Song, S. Wan, and H. R. Wu, "Content-adaptive packet-layer model for quality assessment of networked video services," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 6, pp. 672–683, Oct. 2012.

[50] Q. Huynh-Thu and M. Ghanbari, "The accuracy of PSNR in predicting video quality for different video scenes and frame rates," *Telecommun. Syst.*, vol. 49, pp. 35–48, Jan. 2012.

[51] *Subjective Video Quality Assessment Methods for Multimedia Applications*, Standard Recommended ITU-T P.910, International Telecommunication Union, Apr. 2008.

[52] *Methods for the Subjective Assessment of Video Quality, Audio Quality and Audiovisual Quality of Internet Video and Distribution Quality Television in Any Environment*, Standard Recommended ITU-T P.913, International Telecommunication Union, Mar. 2016.

[53] *Interactive Test Methods for Audiovisual Communications*, Standard Recommended ITU-T P.920, International Telecommunication Union, May 2000.

[54] P. Seeling and M. Reisslein. (May 2013). *Video Trace Library*. [Online]. Available: http://trace.eas.asu.edu/tracemain.html

[55] A. Orsino, D. Moltchanov, M. Gapeyenko, A. Samuylov, S. Andreev, L. Militano, G. Araniti, and Y. Koucheryavy, "Direct connection on the move: Characterization of user mobility in cellular-assisted D2D systems," *IEEE Veh. Technol. Mag.*, vol. 11, no. 3, pp. 38–48, Sep. 2016.

[56] *Technical Specification Group Radio Access Network, User Equipment (UE) Radio Transmission and Reception; Part 1: Range 1 Standalone (Release 16)*, document 3GPP TS 38.101-1, 3rd Generation Partnership Project, Dec. 2019.
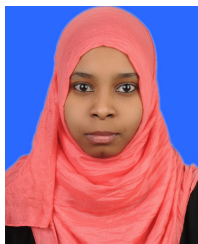
IBTIHAL AHMED received the B.Sc. degree in electrical and electronics engineering from the University of Khartoum, Sudan, in 2014, and the M.Sc. degree in electrical engineering from the American University of Sharjah in 2019. In 2015, she worked for one year as a Teaching Assistant at the University of Khartoum and then worked as a Post Processing Engineer for two years with Sabal Telecom Solutions Company Ltd. She is currently a Research Associate with the Department of Electrical Engineering, American University of Sharjah.

MAHMOUD H. ISMAIL (Senior Member, IEEE) received the B.Sc. degree (Hons.) in electronics and electrical communications engineering and the M.Sc. degree in communications engineering from Cairo University, Egypt, in 2000 and 2002, respectively, and the Ph.D. degree in electrical engineering from The University of Mississippi, MS, USA, in 2006. From August 2000 to August 2002, he was a Research and Teaching Assistant with the Department of Electronics and Electrical Communications Engineering, Cairo University. From 2004 to 2006, he was a Research Assistant with the Center for Wireless Communications (CWC), The University of Mississippi. He is currently a Full Professor with the American University of Sharjah, Sharjah, UAE, and a Full Professor (on leave) with the Department of Electronics and Electrical Communications Engineering, Cairo University. He was also a Systems Engineering Consultant at Newport Media Inc., now part of Microchip, from 2006 to 2014. His research interests include wireless communications with an emphasis on the performance evaluation of next-generation wireless systems and communications over fading channels. He was a recipient of The University of Mississippi Summer Assistantship Award in 2004 and 2005, the Best Paper Award presented at the Tenth IEEE Symposium on Computers and Communications (ISCC), La Manga del Mar Menor, Spain, in 2005, The University of Mississippi Dissertation Fellowship Award in 2006, and The University of Mississippi Graduate Achievement Award in Electrical Engineering in 2006.

MOHAMED S. HASSAN received the M.Sc. degree in electrical engineering from the University of Pennsylvania, Philadelphia, in 2000, and the Ph.D. degree in electrical and computer engineering from The University of Arizona, USA, in 2005. He is currently a Full Professor of electrical engineering with the American University of Sharjah. He was involved in multiple projects related to free space optical communications, electromagnetic shielding, demand response and smart grids, anti-static flooring, and fiber optic sensors for infrastructure health monitoring applications. His research interests include multimedia communications and networking, wireless communications, cognitive radios, resource allocation, and the performance evaluation of wired and wireless networks and next generation wireless systems.

• • •