Received January 14, 2019, accepted January 27, 2019, date of publication February 5, 2019, date of current version February 22, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2897647

A Video Streaming Transmission Scheme Based on Frame Priority in Device-to-Device Multicast Networks

QIQI REN, JIAN CHEN, (Member, IEEE), BO CHEN, AND LI JIN
School of Telecommunications Engineering, Xidian University, Xi'an 710071, China

Corresponding author: Jian Chen (jianchen@mail.xidian.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61771366, and in part by the 111 Project of China under Grant B08038.

ABSTRACT Owing to a dramatic increase of traffic, high demand for quality of service (QoS), and insufficient radio resources, real-time video streaming transmission with stringent delay constraints has been intensely concerned. By exploiting the device-to-device (D2D) multicast communications, this paper proposes a video streaming transmission scheme based on the frame priority (FP) to improve the QoS perceived by users and the users' satisfaction about the video quality. First, the FP strategy is proposed, which mainly considers the encoding characteristics of video streaming and users' feedback to ensure that the frames to be retransmitted are valuable for decoding. Then, in order to transmit a sufficient number of valuable frames within the delay constraints, the optimization of time consumption of the retransmitted frame is formulated. Based on this, the physical-layer resource allocation in the D2D multicast networks is discussed, where the relay selection, D2D subgroup forming, and channel allocation are jointly investigated with the aid of three-dimensional channel quality matrix. Furthermore, a heuristic algorithm is proposed to obtain a near-optimal performance with low complexity. The simulation results verify the advantages of our proposed transmission scheme in users' video reception quality and the satisfaction of all users.

INDEX TERMS D2D multicast, frame priority, resource allocation, real-time video.

I. INTRODUCTION

The proliferation of wireless video contents and services leads to a massive growth of multimedia data traffic in future cellular networks. According to the Visual Networking Index (VNI) forecasted by Cisco in 2017, the traffic of real-time video streaming (e.g. Skype multiplayer conferencing, NBA game live, video gaming [1]), is predicted to grow 39 times by 2021 [2]. The rapid growth of instantaneous data traffic and the higher quality requirements of service proposed by distinguishing features of future real-time video streaming (e.g. higher data rate and lower transmission delay) bring a great challenge to the transmission networks, especially at peak time [3]–[6]. Accordingly, it is necessary to design a more efficient transmission mechanism and expand the transmission system capacity of future cellular networks in order to cope with users' demands for advanced quality of service (QoS).

Video multicast is recognized as one of the most prospective solutions for real-time video streaming in the next generation of cellular networks [7]–[9]. By utilizing video multicast to simultaneously serve the different users with the same subscription on content, the redundant transmissions can be reduced and the system capacity can be efficiently improved [10]–[12]. However, wireless channel is error-prone and location-dependent due to path loss and fading [13]. To provide a reliable transmission for the user who experiences the worst channel condition, multicast serves all users in a group with the same perceived quality, consequently depriving the opportunities of the users who are in good channel conditions to enjoy a higher satisfactory video quality [14].

To overcome the limitation mentioned above, D2D cooperative technology has been implemented to video multicast [15]–[17]. In D2D, the mobile users who are

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

physically close establish a direct communication link, which enables the users to help each other to restore the unsuccessfully decoded video frames, and on the whole, both the QoS perceived by users and the users' satisfaction can be improved [18].

The design of D2D multicast scheme for video content has been investigated in recent researches [9], [18]–[22]. In [19], single-frequency D2D links are employed to increase the throughput of D2D-enhanced conventional video multicast network. However, the real-time video content with stringent delay requirements has not been considered in [19]. Although the complicated video processing techniques such as adaptive video coding [22] adopted by [9], [19], and [20] can be utilized to control the transmission delay, they also increase the computational complexity of video provider. A delay-guaranteed D2D multicast content delivery scheme has been designed in [21], where the content delivery and caching are intelligently scheduled. Wu *et al.* [18] have proposed an effective D2D multicast video distribution system with the consideration of users' mobility and social characteristics, where it concentrates on formulating social relationship to forward the video in D2D multicast manner without considering transmission link status. The aforementioned researches mainly focus on routing regular data streaming.

From the perspective of video streaming, it is necessary to consider the characteristics of encoding structure because the dependency of decoding among video frames is critical to video frame decoding and content reconstructing [23]. A successful decoding of the received video frame is premised on decoding its reference frames. Besides, the users' feedback about the reception of the video frames also should be involved when designing a real-time video streaming transmission scheme. In the presence of network dynamics and the mobile device heterogeneity, a video frame may be received unsuccessfully or dropped due to the interrupted connection or stringent delay constraints. Therefore, in order to ensure that the transmitted frames can be decoded successfully (i.e. be valuable for decoding), the characteristics of encoding and user's feedback about reception of video frames should be considered. Subsequently, the physical-layer resource allocation in D2D multicast networks should also be discussed for the purpose of transmitting as many valuable frames as possible within delay constraints.

Motivated by the observations above, in this paper, we design a flexible video transmission scheme based on frame priority (FP) in D2D multicast networks to improve users' perceived video quality and the satisfaction. The proposed scheme is implemented in two phases: Base station (BS) multicast phase and D2D multicast retransmission phase. In the first phase, BS multicasts high-quality video streaming to all users, and those users who are in good channel conditions can decode video successfully. In the second phase, the video streaming is retransmitted based on frame priority in a manner of D2D multicast, where the users who have decoded the video content forward the decoded data to the users failed to decode the video before, aiming at helping

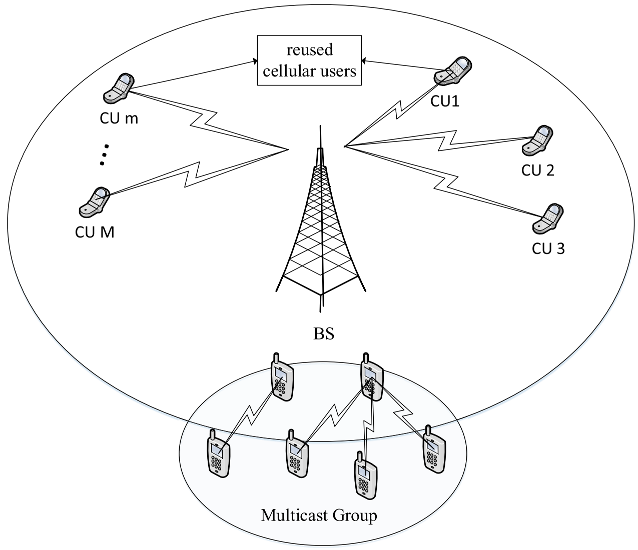


FIGURE 1. System model for the D2D multicast strategy.

them restore the video frames. The main contributions of this paper are summarized as follows:

1) For real-time video streaming, we utilize the priority of video frame with the consideration of the encoding structure of video frames and users' feedback about the reception of frames. This strategy ensures that the video frames to be retransmitted are valuable for decoding and can be decoded successfully.

2) To transmit as many valuable frames as possible within delay constraints, we formulate the minimization of the time consumption of single frame as the optimization problem, where the resource allocation in the physical-layer transmission of D2D multicast networks is discussed.

3) After the transformation of the original optimization problem, a heuristic algorithm with low complexity is proposed to solve the problem, in which the relay selection, D2D subgroup forming and channel allocation are considered jointly with the aid of a three-dimensional channel quality matrix (CQM).

4) Simulation results demonstrate that our proposed scheme can achieve a good performance.

The remainder of this paper is organized as follows. Section II introduces the system model and describes the details of the proposed FP-based video multicast strategy. In Section III, the optimization of time consumption for retransmitted frames is formulated based on the proposed three-dimensional CQM. Section IV describes a near-optimal heuristic algorithm to solve the optimization problem. The simulation results are discussed in Section V. Finally, Section VI concludes the content of this paper.

II. SYSTEM MODEL AND FP-BASED VIDEO MULTICAST STRATEGY

A. SYSTEM MODEL

We consider a single cellular network with D2D communications as shown in Fig. 1, where a group of D2D users

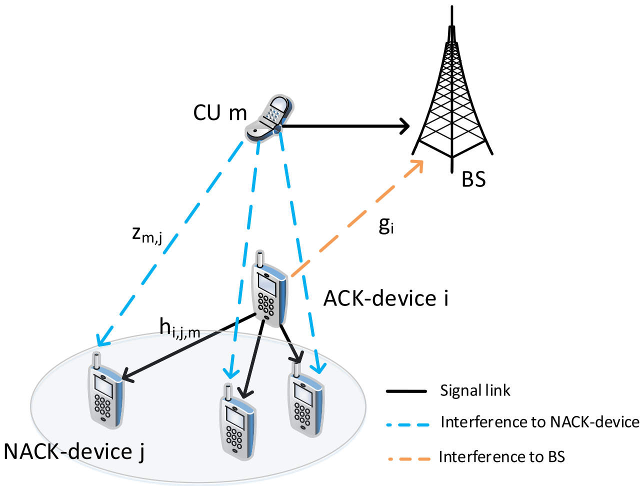


FIGURE 2. Channel reusing model.

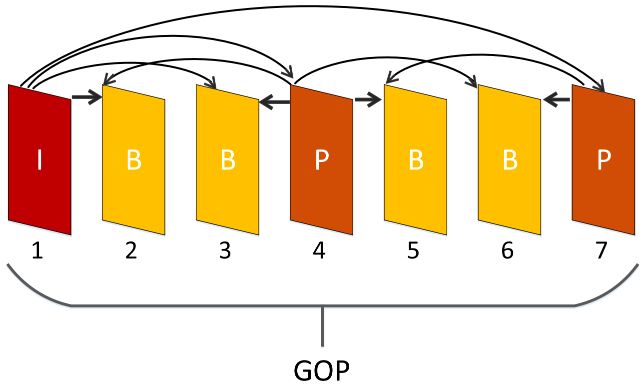


FIGURE 3. GOP structure.

located in a small area coexist with cellular users. The D2D users denoted by $\mathcal{U} = \{1, 2, 3, \ldots, N\}$ are requesting the same real-time video streaming. The BS is responsible for executing multicast transmission in the first phase and controlling the process of retransmission in the second phase. When BS receives the requests sent by D2D users, it will multicast video data to them. Due to the propagation loss, some D2D users may fail to decode the data from the BS, which are referred to as "NACK-devices", while other D2D users are referred to as "ACK-devices". Here, we denote the sets of ACK-devices and NACK-devices for the frame h as $\mathcal{U}_{ACK(h)}$ and $\mathcal{U}_{NACK(h)}$, respectively. Obviously, we have $\mathcal{U}_{NACK(h)} \bigcup \mathcal{U}_{ACK(h)} = \mathcal{U}$ and $\mathcal{U}_{NACK(h)} \bigcap \mathcal{U}_{ACK(h)} = \varnothing$. Hereafter, the frame index h is omitted for brevity of notation.

To transmit more frames, a number of ACK-devices are selected as relay users to forward the decoded data to the NACK-devices. Similarly, the retransmission is also performed in multicast manner. We assume that the D2D communications are progressed under the control of the BS and they reuse the uplink resources of cellular users (CUs) which is also declared in [24]. In addition, we assume that the BS can obtain the perfect global channel state information of all the links in this paper.

Let $\mathcal{M} = \{1, 2, \ldots m, \ldots M\}$ be the set of cellular channels to be reused, each of which is occupied by one CU, and the channels are orthogonal so that there is no interference among them [25]. Each of the selected relay users can only reuse one cellular channel to forward the decoded data to a portion of NACK-devices and each of the cellular channels is only reused by one relay user. The certain interference constraints are satisfied by power control [26] to mitigate the interference to the uplink receiver. As Fig. 2 shows, ACK-device i reuses the uplink channel occupied by CU m to forward data to a subgroup of NACK devices. In Fig. 2, the real line denotes the signal link towards the receiver, and the dotted line denotes the interference link towards the receiver. We assume that g_i is the power gain of interference link from ACK-device i to BS. To mitigate the interference to BS, the optimal result of power control for ACK-device i on channel m is concisely expressed as

$$P_i = \min(P_{\max}, \frac{Q_m}{g_i}), \tag{1}$$

where $P_{\max}$ stands for the power constraint of ACK-device i, and Q_m denotes the interference constraint of BS on channel m. It is assumed that the transmission power of the cellular users is fixed, which is denoted by P_c. The signal-to-interference-plus-noise ratio (SINR) of the link from ACK-device i to NACK-device j on channel m is expressed as

$$\gamma_{i,j,m} = \frac{h_{i,j,m} P_i}{z_{m,j} P_c + N_0}, \tag{2}$$

TABLE 1. Key parameters.

Parameter	Definition
$\mathcal{U}$	The set of D2D users
$\mathcal{U}_{ACK(h)}$	The set of ACK-devices for the frame h
$\mathcal{U}_{NACK(h)}$	The set of NACK-devices for the frame h
$\mathcal{M}$	The set of cellular channels to be reused
M	The number of cellular channels
N	The number of D2D users
I	The number of ACK-devices
J	The number of NACK-devices
P_{max}	The power constraint of ACK-device i (dBm)
Q_m	Interference constraint of BS on channel m (dBm)
N_0	The noise power (dBm)
$\gamma_{i,j,m}$	The SINR of the link from ACK-device i to NACK-device j on channel m
$c_{i,j,m}$	The CQI level of the link from ACK-device i to NACK-device j on channel m
T	The time duration of a GOP (second)
T_1	BS multicast phase (second)
T_2	D2D multicast phase (second)
a	The ratio of BS multicast phase
$R_{i,j,m}$	The transmission of the link from user i to j on channel m (bps)
$e_{i,j,m}$	The spectral efficiency of the link from user i to j on channel m (bps/Hz)
BW	Channel bandwidth (Hz)
$\mathcal{D}_{i_l}$	The set of NACK-devices that served by ACK-device i_l
D	The distance between BS and the center of multicast cluster (meter)

where P_i is the transmission power of ACK-device i, $h_{i,j,m}$ denotes the link power gain from ACK-device i to NACK-device j on channel m, P_c is the interference power of cellular user m, $z_{m,j}$ denotes the interference link power gain from cellular user m to NACK-device j, and N_0 is the noise power.

Based on the above SINR analysis, we obtain the three-dimensional CQM in Section III-B.

B. FP-BASED VIDEO MULTICAST STRATEGY

To ensure the video playback fluency as well as the basic-quality video of all D2D users, a real-time video streaming multicast strategy is designed by employing the characteristics of video streaming and users' feedback. We consider the video streaming encoded with H.264/AVC [27], which is typically encoded by using the Group-of-Picture (GOP) structure. As Fig. 3 shows, there are three kinds of frames contained in a GOP, i.e., the intra-coded (I) frame, the forward predictive coded (P) frame and the bi-directionally predictive coded (B) frame. The I frame acts as the reference frame of B frame and P frame, having a significant influence on decoding quality of the entire GOP. The P frame contains only part of the picture data and its decoding needs the prior I frame [28]. The B frame has the least importance and its decoding depends on both I frame and P frame [23]. Coding dependency between different kinds of frames will lead to decoding dependency among frames and we should adjust transmission order of frames in a GOP according to video encoding structure. Meanwhile, it is also important to take users' feedback about reception of frame into consideration. For instance, when all users receive the I frame, we will no longer transmit it in order to save time to transmit more frames, even if the I frame is the basis for encoding other frames in a GOP. Therefore, the target frame is valuable when both of following points are satisfied: 1) Before transmitting the target frame, its reference frames must have been decoded successfully by the receivers. 2) There are users who decoded the frame unsuccessfully. Our goal of strategy designing is transmitting as many valuable frames as possible to improve all users' reception quality of video streaming and the fairness among all users.

Based on the GOP structure described above, we regard the GOP as a transmission unit. The D2D users can acquire the basic-quality video via decoding the I frame in a GOP. With the development of the decoded frames' number in a GOP, a higher-quality video can be obtained. Due to the delay constraints, each GOP should be transmitted in a specified deadline denoted by T. As mentioned before, T is divided into two parts, i.e., the BS multicast phase T_1 and the D2D multicast phase T_2. In the first phase, the D2D users with good channel conditions can correctly decode more frame data transmitted by the BS. Then the indexes of unsuccessfully decoded frames of NACK-devices are returned to BS. In the second phase, BS selects relay users from the ACK-devices to retransmit those frames. The related process is shown in Fig. 4.

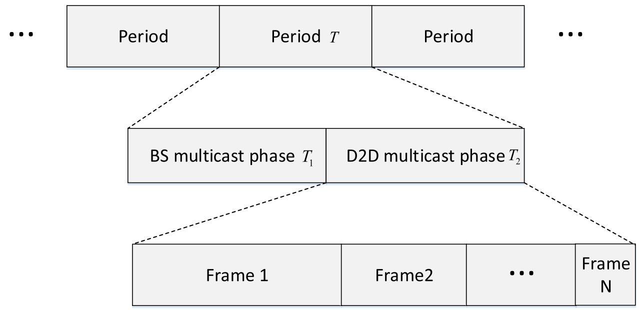


FIGURE 4. The transmission process of a GOP.

Next, we concentrate on the transmission order of frames in the D2D multicast phase. Considering the previously mentioned video encoding structure and users' feedback about reception of frames, we will research the transmission order based on frame priority (FP). Let $\mathcal{H} = \{1, 2, \ldots h, \ldots, H\}$ denote the indexes of unsuccessfully decoded frames in a GOP for all D2D users. In order to evaluate the priority, we set a weight coefficient for every single frame based on the characteristics of video coding structure combined with the received feedback information. We define the weight of frame h as

$$\omega_h = \lambda_h + \frac{n_h}{N}, \tag{3}$$

where n_h is the number of D2D users who fail to decode the frame h, N is the total number of D2D users and λ_h is the maximum number of frames in a GOP that cannot be decoded if frame h is missed. For example, in Fig. 2, we have $\lambda_1 = 6$, since decoding all of the remaining frames in a GOP needs the I frame; and $\lambda_4 = 4$ since decoding all of the 2_{th}, 3_{th}, 5_{th} and 6_{th} frames needs the 4_{th} frame. From the perspective of users, when frames have the same value of λ_h, the more users require the frame, the higher the priority of the frame has. Thus, the sorted ω_h represents the retransmission order of frames in a GOP. When frames have the same value of ω_h, the frames are retransmitted in their original order.

Through transmitting the video frames in the order of frame priority, FP strategy ensures that the frame to be retransmitted is valuable for decoding, thus improving the perceived video quality. Based on the determined retransmission order of the frames, the physical layer transmission is discussed to optimize the transmission time consumption for every single frame so as to transmit as many valuable frames as possible.

III. PHYSICAL-LAYER TRANSMISSION

In this section, we will discuss the resource allocation problem in the D2D multicast phase from the perspective of physical-layer transmission, aiming to transmit as many valuable video frames as possible and improve the perceptual video quality of the users with poor channel conditions. First, we will analyze the time consumption of retransmitted frame. Then the three-dimensional CQM will be introduced, followed by the optimal resource allocation problem formulation.

A. ANALYSIS OF TIME CONSUMPTION

To ensure more valuable frames can be retransmitted within the time constraints in the D2D multicast phase, it is necessary to optimize the time consumption of retransmitting the frames. In the D2D multicast phase, a portion of ACK-devices are selected as relay users to retransmit the data of GOP frame by frame in D2D multicast manner. Each relay user serves a portion of NACK-devices and each NACK-device should be served by only one relay user. The achievable multicast rate of each relay user is constrained by the user whose data rate is the worst one in the corresponding multicast subgroup. The BS is responsible for controlling the retransmission in D2D multicast phase, and only if all NACK-devices receive the data of current target frame correctly, the next frame will be retransmitted. Consequently, the time consumption of target frame h depends on the minimal achievable data rate of all D2D multicast subgroups. The detailed analysis is described as follows.

The time consumption for retransmitting frame h is given by

$$T_h = \frac{B_h}{\min R_{d2d}}, \tag{4}$$

where B_h denotes the number of bits contained in the frame h and $minR_{d2d}$ represents the minimal achievable rate of all D2D multicast subgroups, defined as

$$\min R_{d2d} = \min\{\min_{j_1 \in \mathcal{D}_{i_1}} (R_{i_1,j_1,m_1}), \ldots, \min_{j_L \in \mathcal{D}_{i_L}} (R_{i_L,j_L,m_L})\}. \tag{5}$$

Assuming that the number of D2D multicast subgroups is L, the set $\{i_1, .., i_l, .., i_L\}$ denotes the ACK-devices that act as relay users, the set $\{\mathcal{D}_{i_1}, ..\mathcal{D}_{i_l}.., \mathcal{D}_{i_L}\}$ represents the subgroups of NACK-devices served by relay users i_1 to i_L, the set $\{m_1, .., m_l, .., m_L\}$ denotes the channels reused by relay users i_1 to i_L, and $\min_{j_l \in \mathcal{D}_{i_l}} (R_{i_l,j_l,m_l})$ means that, for any given relay i_l, the achievable multicast rate is limited by the user whose data rate is the worst one in the corresponding NACK-devices subgroup $\mathcal{D}_{i_l}$. $\min\{\min_{j_1 \in \mathcal{D}_{i_1}} (R_{i_1,j_1,m_1}), \ldots, \min_{j_L \in \mathcal{D}_{i_L}} (R_{i_L,j_L,m_L})\}$ denotes the minimal one of the achievable data rates of all D2D multicast subgroups, which is affected by the results of relay selection, subgroup forming and channel allocation. It is observed that the time consumption for a single frame is determined by the minimal achievable data rate of all D2D multicast subgroups. We assumed that, each NACK-device should be served by one relay user during a single frame transmission, and each cellular channel cannot be shared by multiple relay users, which is shown in Fig. 5(b). Since multiple relay users attempting to access the same channel may aggravate the interference of cellular users, the situation in Fig. 5(a) is unreasonable for relay i_1 and i_2 to share the channel m_1.

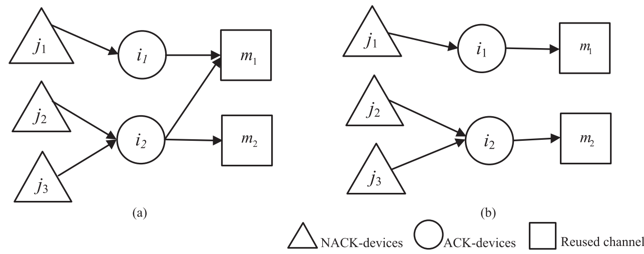


FIGURE 5. Relay selection and channel allocation.

B. THREE-DIMENSIONAL CHANNEL QUALITY MATRIX

In the D2D multicast phase, since the links between ACK-devices and NACK-devices may experience the different fading conditions on different reused channel, the CQM is different from the work in [19], where the two-dimensional matrix is used to reflect the link quality for each D2D link between two nodes. We use the three-dimensional CQM to consider the influence of reusing different channels on D2D link quality. In this paper, the channel quality is measured by the channel quality index (CQI), e.g., $c_{i,j,m}$ denotes the CQI level between ACK-device i and NACK-device j on reused channel m, which is the result of SINR $\gamma_{i,j,m}$ quantification. With the aid of the three-dimensional CQM, the optimization problem of relay selection, subgroup forming and channel allocation can be solved in a joint manner without dividing it into several subproblems, therefore achieving a higher efficiency in terms of maximizing proposed optimization criteria [29].

C. OPTIMIZATION PROBLEM FORMULATION

To transmit as many frames as possible within delay constraints, the time consumption of single frame should be optimized. Meanwhile, there are some constraints about the relay selection, subgroup forming and channel allocation. The optimization problem can be formulated as

$$\begin{aligned} &minimize\ T_h \\ &s.t.\ i_1 \neq i_2 \neq i_3 \neq \ldots \neq i_L, \\ &\quad m_1 \neq m_2 \neq m_3 \neq \ldots \neq m_L 5, \\ &\quad \mathcal{D}_{i_1} \cup \mathcal{D}_{i_2} \cup \ldots \cup \mathcal{D}_{i_L} = \mathcal{U}_{NACK}, \\ &\quad \mathcal{D}_{i_1} \cap \mathcal{D}_{i_2} \cap \ldots \cap \mathcal{D}_{i_L} = \emptyset. \end{aligned} \tag{6}$$

First, in order to simplify the optimization problem, we explore an equivalent optimization related to CQI level feedback. The detailed process is described as follows. The transmission rate of D2D link between user i and j on channel m can be defined as $R_{i,j,m} = e_{i,j,m} \times BW$, where $e_{i,j,m}$ is the spectral efficiency which is associated with the link CQI feedback, and BW is the bandwidth of reused channel [29]. We assume that the uplink channels of cellular users occupy the same bandwidth. It can be observed that the transmission rate R is only related to CQI feedback. According to equation (4), given frame h, the number of bits B_h is a constant, and the time consumption T_h is inversely proportional to $\min R_{d2d}$.

Considering the relationship between transmission rate and CQI, T_h is also inversely proportional to $\min c_{d2d}$, defined as

$$\min c_{d2d} = \min\{\min_{j_1 \in \mathcal{D}_{i_1}} (c_{i_1,j_1,1'}), \ldots, \min_{j_L \in \mathcal{D}_{i_L}} (c_{i_L,j_L,L'})\}, \quad (7)$$

which represents the minimum value of the achievable CQI level of all D2D multicast subgroups. That is, the time consumption is determined by the minimal achievable CQI level of all D2D multicast subgroups. Consequently, the minimization of the time consumption equates to the maximization of $\min c_{d2d}$. The optimization problem can be reformulated as

$$\begin{aligned} & \textit{maximize} \min c_{d2d} \\ & s.t.\ i_1 \neq i_2 \neq i_3 \neq \ldots \neq i_L, \\ & \quad m_1 \neq m_2 \neq m_3 \neq \ldots \neq m_L, \\ & \quad \mathcal{D}_{i_1} \cup \mathcal{D}_{i_2} \cup \ldots \cup \mathcal{D}_{i_L} = \mathcal{U}_{NACK}, \\ & \quad \mathcal{D}_{i_1} \cap \mathcal{D}_{i_2} \cap \ldots \cap \mathcal{D}_{i_L} = \emptyset. \end{aligned} \quad (8)$$

Then, to make the optimization solvable and mathematical, we further find an equivalent expression. Owing to the property of multicast, with a certain result of joint relay selection, subgroup forming, and channel allocation, the minimal achievable CQI level of all D2D multicast subgroups equates to the CQI level of the NACK-device whose link quality is the worst one among all NACK-devices. Our goal is to maximize the CQI level of the worst NACK-device of all NACK-devices through seeking the best one of the combinations of all possible relays, all possible NACK-device subgroups and all possible cellular channels. Thus, the optimization problem is rewritten as

$$\begin{aligned} & \textit{maximize} \ \min_j \sum_i \sum_m c_{i,j,m} \alpha_{i,j,m} \beta_{i,m} \\ & s.t.\ C_1 : \sum_i \sum_m \alpha_{i,j,m} = 1, \quad \forall j \in \mathcal{U}_{NACK} \\ & \quad C_2 : \sum_i \beta_{i,m} \leqslant 1, \quad \forall m \in \mathcal{M} \\ & \quad C_3 : \sum_m \beta_{i,m} \leqslant 1, \quad \forall i \in \mathcal{U}_{ACK} \\ & \quad C_4 : \beta_{i,m} = 1, \quad \textit{if} \sum_j \alpha_{i,j,m} > 0,\ \forall i \in \mathcal{U}_{ACK},\ \forall m \in \mathcal{M} \\ & \quad C_5 : \beta_{i,m} = 0, \quad \textit{if} \sum_j \alpha_{i,j,m} = 0,\ \forall i \in \mathcal{U}_{ACK},\ \forall m \in \mathcal{M} \end{aligned} \quad (9)$$

where $\alpha_{i,j,m}$, and $\beta_{i,m}$ denote association indicators. If $\alpha_{i,j,m} = 1$, we select ACK-device i as the relay of NACK-device j on channel m. Otherwise, it is not. If $\beta_{i,m} = 1$, the channel m will be allocated to relay i, and otherwise, it is not. Constraint C_1 denotes that each NACK-device can be served by only one relay and one channel. Constraint C_2 means that there is only one relay user on channel m. Constraint C_3 expresses that a relay user can only reuse one channel, as shown in Fig. 5(b). Constraints C_4 and C_5 ensure that only if there are NACK-devices who choose the resources of relay i and channel m, the channel m is allocated to relay i.

The optimization is described as how to assign relay users, corresponding D2D multicast subgroups and reused channels to maximize the CQI level of the worst NACK-device among all NACK-devices. After getting the optimal CQI level for retransmitting the target frame, the BS makes a judgment about whether the remaining time is sufficient to transmit the frame. The detailed process is described as follows. The achievable data rate can be obtained according to corresponding spectral efficiency of the CQI. Then, the required time consumption for retransmitting the target frame can be calculated. If the required time is over the remaining time, the retransmission of the remaining frames in the current GOP will be interrupted, and then the BS will multicast the contents of the next GOP. Based on the discussed real-time transmission scheme above, the users with better channel conditions can receive more valuable frame data from the BS within delay constraints, thus obtaining a better perceived video quality and a higher satisfaction. Meanwhile, since the other D2D users can obtain basic-quality video even higher-quality video through D2D communications, the satisfaction among all users can be improved. However, how to solve the optimization proposed in (9) to get the optimal CQI level for retransmitting the target frame? We will next discuss the solution to the optimization problem.

D. COMPLEXITY ANALYSIS

The proposed optimization problem is a mixed integer programming (MIP) problem, which belongs to NP-hard problem. An exhaustive search algorithm can be used to solve this problem, where for all NACK-devices, all possible relays and all possible reused channels are considered under the constraints $C_1 - C_5$. Let I denote the number of ACK-devices, J denote the number of NACK-devices and M denote the number of reused channels. There are $2^I \times 2^M$ possible combinations of relay users and reused channels. Thus, the worst complexity of proposed optimal resource allocation is $O(2^{I+J+M})$. According to section II-A, $I+J = N$, the complexity can be expressed as $O(2^{M+N})$, where N represents the number of D2D users. With the increase of the number of D2D users and reused channels, the complexity of the optimal solution will increase exponentially. Therefore, we should find another solution.

IV. PROPOSED HEURISTIC ALGORITHM

As stated in the previous section, the minimization of time consumption equates to the maximization of the CQI level of the user whose link quality is the worst one among all NACK-devices. Therefore, the formulated optimization problem can be simplified as searching for the optimal CQI level of the worst NACK-device among all NACK-devices. However, since the computational complexity of the optimal solution is prohibitively large, it motivates us to design a new heuristic solution to achieve a near-optimal performance. Thus, we design a heuristic low-complexity algorithm with the aid of three-dimensional CQM to solve the problem jointly.

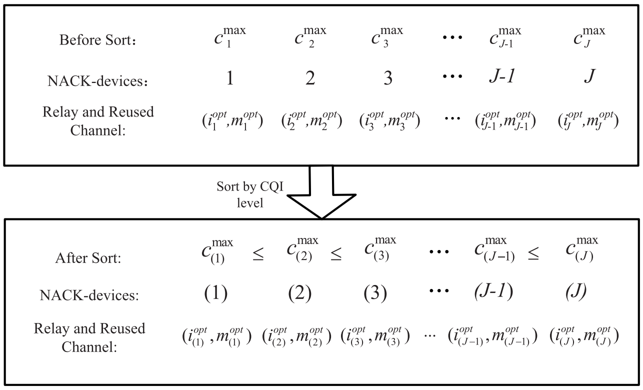


FIGURE 6. The sorting process.

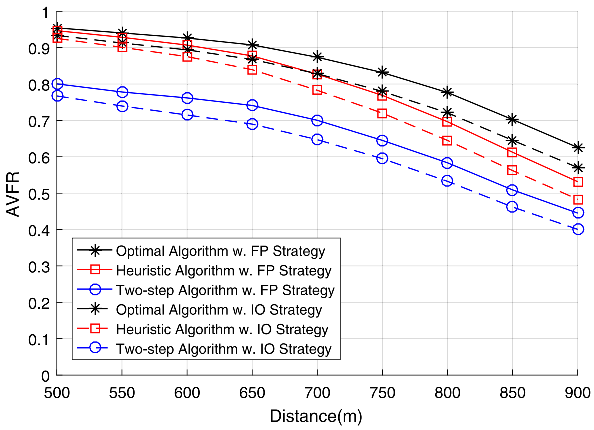


FIGURE 7. The AVFR versus the distance between BS and the multicast cluster center ($a = 0.65$, $M = 15$, $N = 10$).

The basic idea of the proposed heuristic algorithm is to maximize the CQI level of the worst NACK-device. To make the process of proposed algorithm more easier to understand and express, we describe it from the perspective of NACK-devices instead of the BS. First, each NACK-device selects the relay user and reused channel when its D2D link achieves own optimal CQI level. However, it will inevitably result in one-to-many or many-to-one relationship between relay and channel. To satisfy the constraints $C_1 - C_5$, it is necessary to further modify the selected relays and reused channels. Based on the selection result of the worst NACK-device, we adjust the selection results for other NACK-devices. The detailed procedures of the proposed algorithm are presented as follows:

Step 1: For each NACK-device j, we search for its optimal relay user i_j^{opt} and reused channel m_j^{opt} from the proposed three-dimensional CQM denoted by $\mathcal{C}$ so that the CQI level achieves its maximum c_j^{max}.

Step 2: Then the obtained c_j^{max} in step 1 are sorted in an ascending order as $c_{(1)}^{max} \leqslant c_{(2)}^{max} \leqslant \ldots \leqslant c_{(J)}^{max}$, where $c_{(j)}^{max}$ is the CQI level of NACK-device (j). The new index sequence $\{c_{(1)}^{max}, c_{(2)}^{max}, \ldots, c_{(J)}^{max}\}$, distinguished from the original index sequence $\{c_1^{max}, c_2^{max}, \ldots, c_J^{max}\}$, is dynamically adjusted for subsequent operations. The related process is shown in Fig. 7.

Ideally, if the constraints $C_1 - C_5$ are satisfied, $c_{(1)}^{max}$ is the maximum value of the objective function. However, in most cases, the constraints cannot be met without any modification. Hence, we need to further adjust the selected relays and reused channels while keeping the $c_{(1)}^{max}$ as large as possible.

Step 3: Based on the selection result of the worst NACK-device (i.e.$(i_{(1)}^{opt}, m_{(1)}^{opt})$), we make a judgment on whether the selection results of other NACK-devices in Step 1 meet the constraints $C_1 - C_5$. We take NACK-device (2) as an example to describe the process. First, we put the pair $(i_{(1)}^{opt}, m_{(1)}^{opt})$ in a set $\mathcal{S}$, which represents the selected resource pairs (relay users and reused channels). Then each pair in $\mathcal{S}$ is compared with the pair $(i_{(2)}^{opt}, m_{(2)}^{opt})$. The comparison results can be summarized in the following three cases:

Case 1: $i_{(1)}^{opt} = i_{(2)}^{opt}$ and $m_{(1)}^{opt} = m_{(2)}^{opt}$, which means the NACK-device (2) shares the resource pair $(i_{(1)}^{opt}, m_{(1)}^{opt})$ with NACK-device (1).

Case 2: $i_{(1)}^{opt} \neq i_{(2)}^{opt}$ and $m_{(1)}^{opt} \neq m_{(2)}^{opt}$, which means the resource pair selected by NACK-device (2) is completely different from that of NACK-device (1).

Both Case1 and Case2 satisfy the constraints $C_1 - C_5$ while $c_{(1)}^{max}$ remains unchanged and we don't need more adjustments. For the Case2, we put the pair $(i_{(2)}^{opt}, m_{(2)}^{opt})$ in $\mathcal{S}$.

Case 3: $i_{(1)}^{opt} = i_{(2)}^{opt}$ and $m_{(1)}^{opt} \neq m_{(2)}^{opt}$ or $i_{(1)}^{opt} \neq i_{(2)}^{opt}$ and $m_{(1)}^{opt} = m_{(2)}^{opt}$, which means the constraints $C_1 - C_5$ cannot be satisfied simultaneously. It is necessary to make some adjustments for NACK-device (2) while keeping the $c_{(1)}^{max}$ as large as possible. We assume that when NACK-device (2) shares the pair (i_s, m_s) (i.e.$(i_{(1)}^{opt}, m_{(1)}^{opt})$) in $\mathcal{S}$ to maximize its CQI, the corresponding CQI is denoted by c_s and when NACK-device (2) selects a new pair (i_{new}, m_{new}) other than the pair in $\mathcal{S}$ to maximize its CQI, the corresponding CQI is denoted by c_{new}. To save resources, we tell whether the resource pair in $\mathcal{S}$ can be shared by checking whether $c_s > c_{(1)}^{max}$ is satisfied. If not, we select the larger one between c_s and c_{new} as $c_{(2)}^{max}$ and the corresponding resource pair is regarded as $(i_{(2)}^{opt}, m_{(2)}^{opt})$. When the resource pair selected is (i_{new}, m_{new}), we put it in $\mathcal{S}$. In particular, for the case of $c_s < c_{(1)}^{max}$ and $c_{new} < c_{(1)}^{max}$, current $c_{(1)}^{max}$ is no longer the minimum value among all NACK-devices and we should reorder all the users that have been discussed (i.e. NACK-device(1) and (2)) to get a new $c_{(1)}^{max}$.

Step 4: We compare $(i_{(k)}^{opt}, m_{(k)}^{opt})$ obtained in Step 1 with all resource pairs in $\mathcal{S}$ according to the process in Step 3, and then repeat the process.[1]

The pseudo-code of the proposed heuristic algorithm can be summarized as Algorithm 1. The process of Step 1 has the complexity of (IMJ). The complexity of Step 2 for sorting the previously obtained CQI level is $O(Jlog_2J)$. And the

[1]Note that for the process of our algorithm, whether the number of relays is larger than the number of the cellular channels (i.e. $I > M$) or not doesn't matter, because the matching results of relays and channels stored in the set $\mathcal{S}$ are based on the NACK-devices' final selection results of the available relays and available channels, and the direct matching process of channels, relays and their corresponding subgroups is not involved in our algorithm.

Algorithm 1 The Heuristic Algorithm

Input: $\mathcal{U}_{ACK}$, $\mathcal{U}_{NACK}$, $\mathcal{M}$, CQM $\mathcal{C}$
Output: c^{opt}, $\mathcal{S}$
Step1 & Step2:
  for all $j \in \mathcal{U}_{NACK}$, with the given $\mathcal{C}$ **do**
    Find the maximum CQI level as c_j^{max} and corresponding pair (i_j^{opt}, m_j^{opt})
  end for
  Sort c_j^{max} in an ascending order and we have $c_{(1)}^{max} \leqslant c_{(2)}^{max} \leqslant \ldots \leqslant c_{(J)}^{max}$
  Initialize set $\mathcal{S} = \{(i_{(1)}^{opt}, m_{(1)}^{opt})\}$
Step3 & Step4:
  for all $(k) = (2)$ to (J) **do**
    if case1 is true or case2 is true **then**
      continue and put $(i_{(k)}^{opt}, m_{(k)}^{opt})$ into $\mathcal{S}$ when case2 is true
    else
      Find the maximum CQI level of all pairs in $\mathcal{S}$ as c_s, corresponding pair as (i_s, m_s)
      Find the maximum CQI level of all pairs not involved in $\mathcal{S}$ as c_{new}, corresponding pair as (i_{new}, m_{new})
      if $c_s > c_{(1)}^{max}$ **then**
        $(i_{(k)}^{opt}, m_{(k)}^{opt}) = (i_s, m_s)$ and $c_{(k)}^{max} = c_s$
      else
        Choose the larger one between c_s and c_{new} as $c_{(k)}^{max}$, corresponding pair as $(i_{(k)}^{opt}, m_{(k)}^{opt})$, and put the pair in $\mathcal{S}$ when we choose c_{new}
      end if
      Reorder $c_{(1)}^{max}$ to $c_{(k)}^{max}$ when current $c_{(1)}^{max}$ is no longer the minimum and then get a new $c_{(1)}^{max}$
    end if
  end for
  $c^{opt} = c_{(1)}^{max}$

complexity of Step 3 is $O((I-1)(M-1)(J-1))$. Based on the above analysis, the overall complexity of the proposed algorithm is $O(MNJ)$. Obviously, it has lower complexity than the optimal algorithm mentioned in section III-D.

V. PERFORMANCE EVALUATION

In this section, some numerical results are presented to evaluate the performance of proposed real-time video multicast scheme.

A. SIMULATIONS AND PERFORMANCE METRICS

In the simulation, we consider that the video streaming is encoded with H.264/AVC, where the resolution is 352×288, and the video frame rate is 30 frames/s. Here, each COP contains 7 video frames with the sequence of IBBPBBP. According to [31], the multipath fading follows an exponential distribution with unit mean. The rest values of the main parameters are listed in TABLE 2. After 10,000 independent experiments, we obtained the average simulation results.

TABLE 2. Main simulation settings.

Parameter	Value
The radius of the multicast cluster	200m
The time ratio a occupied by BS	0.4 to 0.9
The distance D between BS and the center of multicast cluster	500 to 900m
Maximum transmit power	BS: 43dBm Cellular user: 30dBm D2D user: 23dBm
Channel bandwidth BW	500KHz
Noise spectral density	-174dBm/Hz
Path loss	31.54+37.1log10(d(m))
Interference threshold Q	-40 dBm to -10dBm

The proposed scheme is evaluated in terms of the following performance metrics.

1) AVERAGE VALUABLE FRAME RATIO (AVFR)

First, we use Valuable Frame Ratio (VFR) to describe the ratio between the number of correctly decoded video frames and the total number of transmitted video frames. For user $n (n = 1, 2, \ldots, N)$, the VFR is defined as

$$v_n = \frac{H_n}{H}, \tag{10}$$

where H is the total number of frames transmitted by BS, and H_n represents the decoded frame number of user n. A larger v_n means that user n can decode more video frames correctly, thereby getting a better perceived video quality.

Then, we use AVFR to describe the average valuable frame ratio of all users, which is defined as

$$\mathrm{AVFR} = \frac{1}{N}\sum_{n=1}^{N} v_n. \tag{11}$$

The AVFR represents the valuable video frames performance of all users in the D2D multicast network, thus reflecting the average perceived video quality.

2) AVERAGE SATISFACTION INDEX (ASI)

This metric represents the average satisfaction valuation about the correctly decoded frames for all users. As described in section II-B, the dependency of decoding among video frames indicates that different frames have unequal importance on restoring the video content, thus leading to different experience of perceptual video quality. As a result, we can use Satisfaction Index (SI) to represent the satisfaction valuation for users' experience of perceptual video quality. The I frame has the largest satisfaction valuation for its significant effect on video quality, and it can satisfy the users' basic-quality demands. As the contributions of other frames to perceptual video quality decrease, the valuations of them decrease. The valuations for frame I, P, and B are 0.4, 0.2, and 0.05 respectively. Let SI_n represent the total satisfaction valuation of the decoded frames for user n. The ASI is defined as

$$\mathrm{ASI} = \frac{1}{N}\sum_{n=1}^{N} SI_n. \tag{12}$$

From the perspective of whether considering the characteristics of video streaming combined with users' feedback, we compare the performances of following two strategies: the proposed FP strategy and initial-order (IO) strategy, in which the frames in a GOP are transmitted in the initial order, without considering the frame priority. Moreover, from the perspective of the resource allocation optimization in physical-layer transmission, we compare the performances of following three algorithms:

1). Optimal algorithm: Considering all possible combinations of relay users, NACK-device subgroups and reused channels.

2). Proposed heuristic algorithm: Utilizing the three-dimensional CQM to investigate relay selection, subgroup forming and channel allocation jointly.

3). Two-step algorithm: Dividing the considered problem into two sub-problems and to solve them with the following two steps [30]: The first step is channel allocation, in which the relay users are matched with the reused channels according to the optimal power control as expressed in the equation (1), in order to find the best channel for each relay user to maximize its permitted transmission power. The second step is that the NACK-devices select the relay users based on the two-dimensional CQM to form D2D subgroups.

B. RESULTS ANALYSIS

Fig. 7 shows AVFR performance of different schemes under different distances between BS and the multicast cluster center. For each curve, as the distance increases, AVFR performance degrades for the fact that the increasing of the distance leads to the decreasing of the number of ACK-devices, and therefore, fewer available relay users can be used to serve the increasing number of NACK-devices. First, we prove the advantage of the proposed FP strategy. For the same retransmission algorithm, the FP strategy always outperforms the IO strategy, since the FP strategy enables users to receive as many valuable frames as possible within delay constraints compared to IO strategy and it is observed that the gap between two strategies is gradually increasing. When the distance is small enough, the majority of users can receive all of the frames, and it means that the retransmission order of video frames does not affect AVFR. After that, the number of users who benefit from multicast phase decreases, and then retransmission becomes more important, consequently proving the advantage of the FP strategy. Then, we compare the performances of three algorithms with the same retransmission strategy (FP or IO strategy). The optimal algorithm and the proposed heuristic algorithm have a higher AVFR than the two-step algorithm. Although the gap between the optimal algorithm and the proposed heuristic algorithm becomes apparent when the distance is large enough, the proposed algorithm has lower computational complexity compared with the optimal algorithm whose computational overhead is prohibitively large, especially for plenty of D2D users and cellular channels.

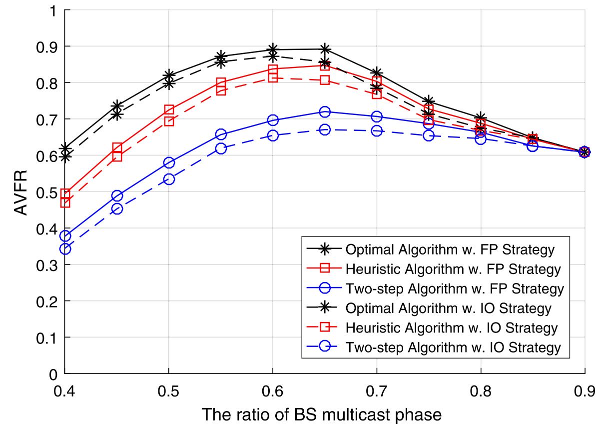


FIGURE 8. The AVFR versus the ratio of BS multicast phase ($D = 700$m, $M = 15$, $N = 10$).

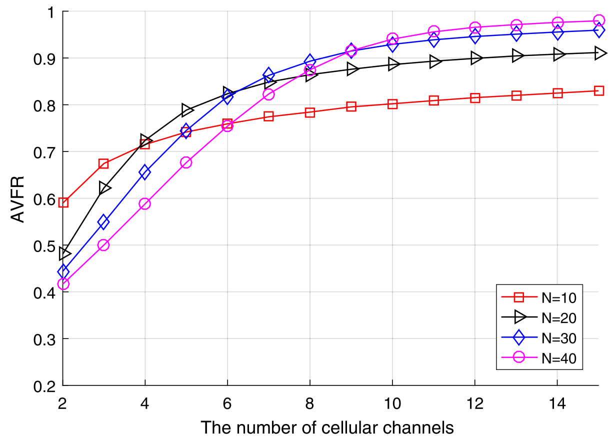


FIGURE 9. The AVFR versus the number of cellular channels under different number of D2D users ($a = 0.65$, $D = 700$m).

Fig. 8 illustrates AVFR performance of different schemes under different ratios of BS multicast phase. We assume that the ratio of BS multicast phase is defined as $a = \frac{T_1}{T}$, where $T = T_1 + T_2$, and T_1 and T_2 represent the time allocated to BS multicast phase and D2D multicast phase, respectively. When the ratio a is small, the multicast rate is large enough to guarantee a complete GOP transmission within time aT, which means only a small number of users can decode the data of a GOP successfully, thus resulting in a poor performance. As the ratio a increases, the AVFR performance is better for the reason that more users can benefit from multicast. Although the increase of ratio a makes the performance during first phase better, it makes the performance of retransmission during second phase worse due to the less time allocated to T_2, so the enhancement of AVFR becomes slowly. The optimal performance is achieved at the ratio $a = 0.65$, which means that there is a trade-off between the effects of first phase and second phase. After that, because of the shorter second retransmission phase, the AVFR is decreasing and the difference among three algorithms is getting smaller. At the ratio $a = 0.9$, even the retransmission of the I frame

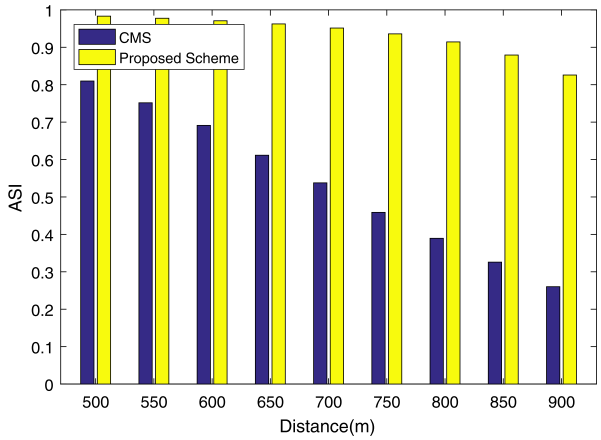


FIGURE 10. The ASI versus the distance between BS and the multicast cluster center ($a = 0.65$, $M = 15$, $N = 30$).

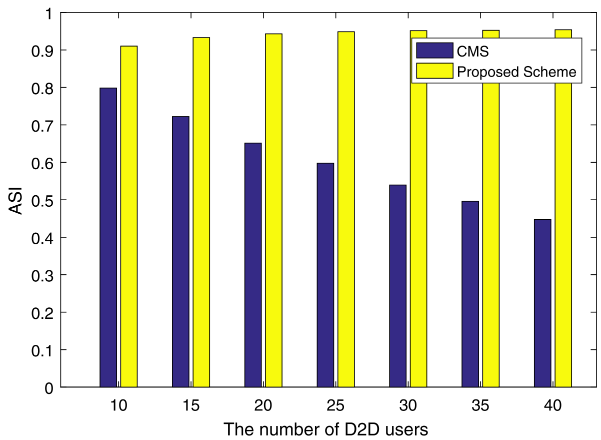


FIGURE 11. The ASI versus the number of D2D users ($a = 0.65$, $M = 15$, $D = 700$m).

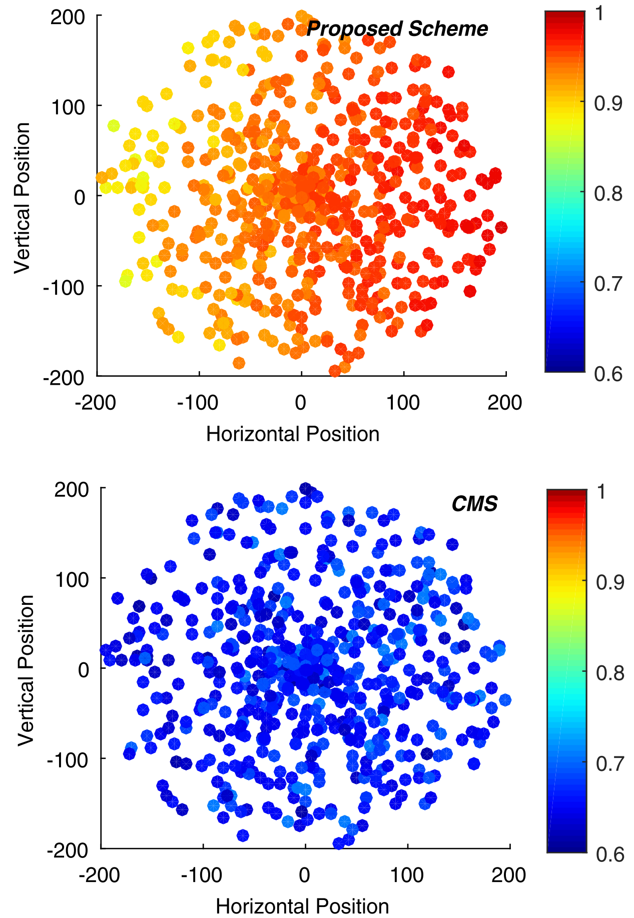


FIGURE 12. The SI versus the users' physical positions ($a = 0.65$, $M = 15$, $D = 700$m, $N = 20$).

cannot be guaranteed, so the performance of these algorithms is the same.

Fig. 9 shows AVFR performance of proposed scheme (proposed heuristic algorithm with FP strategy) under different number of cellular channels and under different number of D2D users. It is observed that, as the increase in the number of cellular channels, AVFR performance is enhanced overall, for the reason that more available reused cellular channels make the achievable retransmission data rate larger, and therefore more valuable frames can be retransmitted within delay constraints. Then, we compare the performance of the proposed scheme under different number of D2D users. When the number of cellular channels is small, the available reused resources unable to bring a well performance for the case of more D2D users. As the number of cellular channels increases, the sufficient available reused resources can afford a large number of D2D users; thereby the performance of overall system benefits from multiuser diversity. The result demonstrates proposed scheme is appropriate to deal with large-scale mobile users scenario in the view of reception of valuable video frames.

Fig. 10 and Fig. 11 respectively illustrate ASI performance of proposed scheme and conventional multicast scheme (CMS) under different distance between BS and the multicast cluster center and under different number of D2D users. In Fig. 10, the proposed scheme always performs better than CMS. For CMS, since the data rate of multicast is limited to the worst link among all users, all users have the same satisfaction valuation; therefore ASI performance of CMS has a fast degradation as the distance increases. Especially for long distance, the value of ASI is less than 0.4, which means basic quality cannot be guaranteed for all users. However, our proposed scheme takes the heterogeneity characteristics of links among all users into consideration, providing good condition users the perceptual video quality matching their channel conditions and providing bad condition users a satisfactory perceptual quality via D2D multicast retransmission, thus improving satisfaction of all users. In Fig. 11, it is obvious that ASI performance of the proposed scheme is enhanced as the number of D2D users increases because of multiuser diversity, while ASI performance of CMS is getting worse. Accordingly, the proposed scheme is well designed in the view of users satisfaction no matter the number of users is large or small.

Fig. 12 demonstrates the SI performance of proposed scheme under different user's physical positions compared with CMS to verify how the proposed scheme improves the satisfaction valuation in different occasions. The origin is the cluster center of the multicast group, and BS is located in the position (700, 0). The result reflects the 30 independent experiments of the SI performance(1000 channel realizations for each D2D user), and the total number of D2D users in each experiment is 20. It is obvious that, on the one hand, all the users in the proposed scheme can achieve a better SI performance compared with CMS. On the other hand, the SI performance in CMS always stays in a low level no matter for good condition users or bad condition users, since that CMS serves all users in the same quality in every single experiment; however, our proposed scheme can provide users the differentiated perceptual video quality to match their channel conditions, which guarantees the better experience of good condition users and at the same time improves the performance of bad condition users as much as possible, thus improving the overall performance of the transmission system.

VI. CONCLUSION

In this paper, we proposed a FP-based real-time video streaming transmission scheme in D2D multicast networks to improve the QoS perceived by users and the satisfaction of all users. Considering the characteristics of video streaming and users' feedback, we obtained the frame priority for the video frames to be retransmitted. We formulated the time consumption of a frame as a optimization problem, where the relay selection, D2D subgroup forming and channel allocation in the physical-layer transmission of D2D multicast networks were jointly investigated. A heuristic algorithm with less computational complexity was designed to solve the optimization problem. Simulation results showed that the proposed scheme had a significant improvement on users' perceived video quality and the satisfaction of all users. In our future work, the problem of trade-off between the spectral efficiency and the energy efficiency for real-time video streaming in wireless communication networks would be considered.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive comments and suggestions.

REFERENCES

[1] J. Wu, R. Tan, and M. Wang, "Streaming high-definition real-time video to mobile devices with partially reliable transfer," *IEEE Trans. Mobile Comput.*, vol. 18, no. 2, pp. 458–472, Feb. 2019.

[2] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2016–2021," Cisco, San Jose, CA, USA, White Paper, 2017. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html

[3] J. Wen, C. Ren, and A. K. Sangaiah, "Energy-efficient device-to-device edge computing network: An approach offloading both traffic and computation," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 96–102, Sep. 2018.

[4] Y. Cao, T. Jiang, X. Chen, and J. Zhang, "Social-aware video multicast based on device-to-device communications," *IEEE Trans. Mobile Comput.*, vol. 15, no. 6, pp. 1528–1539, Jun. 2016.

[5] D. Jurca and P. Frossard, "Media flow rate allocation in multipath networks," *IEEE Trans. Multimedia*, vol. 9, no. 6, pp. 1227–1240, Oct. 2007.

[6] R. Gupta, A. Pulipaka, P. Seeling, L. J. Karam, and M. Reisslein, "H.264 coarse grain scalable (CGS) and medium grain scalable (MGS) encoded video: A trace based traffic and quality evaluation," *IEEE Trans. Broadcast.*, vol. 58, no. 3, pp. 428–439, Sep. 2012.

[7] C. Li, H. Xiong, J. Zou, and D. O. Wu, "Joint dynamic rate control and transmission scheduling for scalable video multirate multicast over wireless networks," *IEEE Trans. Multimedia*, vol. 20, no. 2, pp. 361–378, Feb. 2018.

[8] Y. Zhou, F. R. Yu, J. Chen, and Y. Kuo, "Video transcoding, caching, and multicast for heterogeneous networks over wireless network virtualization," *IEEE Commun. Lett.*, vol. 22, no. 1, pp. 141–144, Jan. 2018.

[9] K. C. Lin, W.-L. Shen, C.-C. Hsu, and C.-F. Chou, "Quality-differentiated video multicast in multirate wireless networks," *IEEE Trans. Mobile Comput.*, vol. 12, no. 1, pp. 21–34, Jan. 2013.

[10] R. O. Afolabi, A. Dadlani, and K. Kim, "Multicast scheduling and resource allocation algorithms for OFDMA-based systems: A survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 240–254, 1st Quart., 2013.

[11] T.-P. Low, M.-O. Pun, Y.-W. P. Hong, and C.-C. J. Kuo, "Optimized opportunistic multicast scheduling (OMS) over wireless cellular networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 791–801, Feb. 2010.

[12] L. Lv, J. Chen, Q. Ni, and Z. Ding, "Design of cooperative non-orthogonal multicast cognitive multiple access for 5G systems: User scheduling and performance analysis," *IEEE Trans. Commun.*, vol. 65, no. 6, pp. 2641–2656, Jun. 2017.

[13] S.-C. Wang and W. Liao, "Cooperative multicasting for wireless scalable video transmissions," *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 3980–3989, Sep. 2013.

[14] P. Li, H. Zhang, B. Zhao, and S. Rangarajan, "Scalable video multicast with adaptive modulation and coding in broadband wireless data systems," *IEEE/ACM Trans. Netw.*, vol. 20, no. 1, pp. 57–68, Feb. 2012.

[15] L. Yang, J. Chen, H. Zhang, H. Jiang, S. A. Vorobyov, and D. T. Ngo, "Cooperative wireless multicast: Performance analysis and time allocation," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5810–5819, Jul. 2016.

[16] M. Ji, G. Caire, and A. F. Molisch, "Wireless device-to-device caching networks: Basic principles and system performance," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 1, pp. 176–189, Jan. 2016.

[17] B. He, J. Chen, Y. Kuo, and L. Yang, "Cooperative jamming for energy harvesting multicast networks with an untrusted relay," *IET Commun.*, vol. 11, no. 13, pp. 2058–2065, Oct. 2017.

[18] D. Wu, J. Yan, H. Wang, D. Wu, and R. Wang, "Social attribute aware incentive mechanism for device-to-device video distribution," *IEEE Trans. Multimedia*, vol. 19, no. 8, pp. 1908–1920, Aug. 2017.

[19] L. Militano, M. Condoluci, G. Araniti, A. Molinaro, A. Iera, and G. M. Muntean, "Single frequency-based device-to-device-enhanced video delivery for evolved multimedia broadcast and multicast services," *IEEE Trans. Broadcast.*, vol. 61, no. 2, pp. 263–278, Jun. 2015.

[20] X. Zhang, M. Yang, Y. Zhao, J. Zhang, and J. Ge, "An SDN-based video multicast orchestration scheme for 5G ultra-dense networks," *IEEE Commun. Mag.*, vol. 55, no. 12, pp. 77–83, Dec. 2017.

[21] Y. Xu, "A device-to-device multicast scheme for delay-constraint content delivery," *Mobile Inf. Syst.*, vol. 2017, Feb. 2017, Art. no. 8012517.

[22] Y. Wang, L. Chau, and K.-H. Yap, "Bit-rate allocation for broadcasting of scalable video over wireless networks," *IEEE Trans. Broadcast.*, vol. 56, no. 3, pp. 288–295, Sep. 2010.

[23] N. Ansari, H. Liu, Y. Q. Shi, and H. Zhao, "On modeling MPEG video traffics," *IEEE Trans. Broadcast.*, vol. 48, no. 4, pp. 337–347, Dec. 2002.

[24] 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.

[25] H. Meshgi, D. Zhao, and R. Zheng, "Joint channel and power allocation in underlay multicast device-to-device communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., Jun. 2015, pp. 2937–2942.

[26] G. Ozcan and M. C. Gursoy, "Optimal power control for underlay cognitive radio systems with arbitrary input distributions," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4219–4233, Aug. 2015.

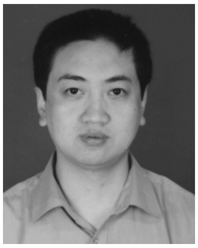
[27] P. Seeling and M. Reisslein, "Video transport evaluation with H.264 video traces," *IEEE Commun. Surveys Tuts.*, vol. 14, no. 4, pp. 1142–1165, 4th Quart., 2012.

[28] A. K. A. Tamimi, C. So-In, and R. Jain, "Modeling and resource allocation for mobile video over WiMAX broadband wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 3, pp. 354–365, Apr. 2010.

[29] A. Asheralieva and Y. Miyanaga, "QoS-oriented mode, spectrum, and power allocation for D2D communication underlaying LTE-A network," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 9787–9800, Dec. 2016.

[30] C. Ren, J. Chen, Y. Kuo, D. Wu, and M. Yang, "Recommender system for mobile users: Enjoy Internet of Things socially with wireless device-to-device physical links," *Multimedia Tools Appl.*, vol. 77, no. 4, pp. 4133–4153, Feb. 2018.
[31] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

QIQI REN received the B.Eng. degree from the School of Information and Communication Engineering, North University of China, China, in 2016. She is currently pursuing the Ph.D. degree with the School of Telecommunications Engineering, Xidian University, China. Her research interest includes wireless multimedia communications.



JIAN CHEN (M'14) received the B.Eng. degree from Xi'an Jiaotong University, China, in 1989, the M.Eng. degree from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, in 1992, and the Ph.D. degree in telecommunications engineering from Xidian University, China, in 2005. He was a Visiting Scholar with The University of Manchester, from 2007 to 2008. He is currently a Full Professor with the School of Telecommunications Engineering, Xidian University. His research interests include cognitive radio, physical layer security, wireless sensor networks, compress sensing, and signal processing.

BO CHEN received the B.Eng. and M.Eng. degrees from the School of Telecommunications Engineering, Xidian University, China, in 2013 and 2017, respectively. Her research interest includes wireless device-to-device multicast communications.

LI JIN received the B.Eng. degree from the Chengdu University of Information Technology, China, in 2014, and the M.Eng. degree from the School of Telecommunications Engineering, Xidian University, China, in 2018. Her research interest includes wireless device-to-device multicast communications.