Optimal transmission for scalable video coded streaming in cellular wireless networks with the cooperation of local peer-to-peer network

Guangsheng Feng¹, Junyu Lin², Qian Zhao³, Huiqiang Wang¹, Hongwu Lv¹, Bingyang Li¹ and DongDong Su¹

Abstract
We study the problem of transmitting scalable video coded streams in a hybrid scenario including a global cellular network and a local peer-to-peer network. Given desirable channel condition in cellular downlink, some cellular user equipments can fetch the video chunks via the cellular base station directly, and also provide relay service via the local peer-to-peer network for other user equipments unable to meet their basic video quality. To take the advantage of channel diversity gains, the method of random linear coding is adopted to generate linear combinations of the video chunks for transmission in both the cellular and peer-to-peer networks. To optimize the cellular and peer-to-peer transmission arrangement, we first formulate it as a mixed integer nonlinear programming problem, which becomes more complicated when the number of user equipments increases. Then, we convert it to a quasi-convex optimization problem by approximating the indicator function with a continuous step function, and it can be solved by a centralized approach. Furthermore, a primal-dual decomposition approach is developed and a distributed algorithm is proposed accordingly. Simulation results show that the proposed approach achieves a near-optimal solution.

Keywords
Cellular networks, peer-to-peer network, scalable video coding, random linear coding, multicast

Introduction
According to a recent statistics, global Internet video traffic accounted for 57% of all consumer traffic in 2012 and will even rise to 69% in 2017.¹ More than a half of watching online video activities are via various wireless devices, for example, smartphone and tablet. Supporting video streaming in wireless cellular networks becomes a challenging issue, given the large bandwidth consumption, time-varying wireless channels, and more higher video quality requirements from various wireless terminals.²⁻⁴ Recently, the scalable video coding (SVC) technology has been proposed.⁵⁻⁶ It is quite attractive since the video streaming under wireless networks can be adaptive according to the network.
conditions, user equipment (UE) constraints, and so on. However, streaming video content in wireless networks is still one of the most bandwidth-hungry applications compared with other Internet applications.

In a loss-prone wireless network, cooperative transmission of nearby users is an effective way to achieve the desirable video quality. User cooperation can be implemented via low-cost local connections, for example, local peer-to-peer (P2P) connections and device-to-device (D2D) connections, which is a promising technology in cellular networks. Considering the varying wireless channel conditions, traditional cooperative transmission requires frequent feedbacks among the cooperative users, and thus it can consume much communication resources. To conquer this problem, network coding is often used in the broadcast/multicast scenarios. In our work, we also adopt one of network coding approaches, namely, random linear coding (RLC), for encoding each video chunk by the base station (BS) or server. Different from the existing work where the BS unicasts videos to users, we employ broadcast in cellular transmission, where the transmission rate does not cater to the lowest receiving rate of UEs for broadcast efficiency. Correspondingly, some UEs with much worse-downlink channel conditions, that is, such UEs’ receiving rates are lower than the cellular transmission rate, cannot receive the cellular broadcasted chunks at all. Those incapable of obtaining video chunks from the cellular network can purchase relaying service from other UEs with better cellular downlink conditions at the monetary cost. Another difference is that some existing works utilize the D2D connections in the local transmissions without considering transmission delay, for only one-hop transmission is allowed. In contrast, there is no one-hop limitation when we multicast the video streams in the local P2P network, but the transmission duration has to be addressed. As a result, both the BS and UEs can cooperate under the broadcast/multicast manner, which will be more efficient when a large number of terminals request the same video.

Although we enable all UEs to multicast their peer nodes, the achieved video rate will be affected by both the cellular network and local P2P network, which plays a significant role on video playback quality. For multi-layer SVC video streaming, different layers have different effectiveness in users’ perceived video quality, and one high layer usually relies on all its lower layers to decode. Intuitonally, the lower layers should be allocated with more transmission resources, which guarantees the higher layers can be decoded. Otherwise, the higher layers become invalid. Furthermore, only the UEs, that received sufficient chunks from both the BS and their peers in advance, are allowed to multicast. The UEs, that received sufficient RLC-encoded chunks to reconstruct one entire layer, are called “matured” nodes. After reconstructing a layer successfully, the matured UE can be re-encoded for multicasting its unmatured peers. However, unmatured to matured state transfer occurs in the midway of transmission, the precise time is hardly known, which results in an intractable constraint in the system modeling. Those difficulties are the main concerns that we need to deal with in this work. Our objective is to tradeoff the achieved video playback quality and transmission cost, and design an optimal transmission scheduling algorithm among video layers considering the cooperation of cellular and local P2P transmissions. Our main contributions are summarized as follows:

- We formulate the problem of cooperative transmission in cellular network and local P2P network for multi-layer SVC streaming as a mixed integer nonlinear programming (MINLP) problem, in which the cellular transmission duration and the local P2P transmission duration are coupled.
- By introducing a continuous step function, we convert the problem of cooperative transmission to a quasi-convex problem, where the UE’s unmatured to matured state transfer can be approximated by the quasi-convex step function. To decouple the transmission durations of cellular and local P2P networks, we design a pipeline work mode at the cost of postponed start-up video playback.
- A near-optimal solution can be obtained by employing the dual-decomposition approach, and we also propose a distributed algorithm such that the BS and each UE can calculate its own transmission duration for each layer independently.
- We conduct extensive experiments to verify our solution and evaluate the performance of our distributed approach.

The rest of this article is organized as follows. Section “Related work” summarizes the state-of-art of the video streaming in wireless networks. Section “System overview” presents the system model and formulates the optimization problem with respect to the reward and traffic cost in section “Problem formulation.” Section “Approximation of the primal optimization problem” decouples the proposed optimization problem by a quasi-convex step function and relaxing the start-up playback time, and solve it in a centralized way. In section “Distributed algorithm design,” we design a distributed algorithm for each node to calculate its own broadcast traffic duration independently. Section “Performance evaluation” evaluates the proposed approach with extensive simulations, followed by the concluding remarks in section “Conclusion.”
Related work

To improve the perceived video quality in the loss-prone wireless networks, much more attentions have been focused on transmission cooperation.\textsuperscript{8,10–12,19} Jointly considering the cellular transmission and local UEs’ transmission, Abedini et al.\textsuperscript{10} proposed a provable minimum-cost algorithm to guarantee each user’s quality of service (QoS), which took the BS stopping time under unicast and local UEs’ broadcast scheduling into account. Bethanabhotla et al.\textsuperscript{11} designed a dynamic scheduling policy for transmitting video streaming in wireless networks, where multiple nearby nodes can help each to fulfill their video quality requirement. They formulated it as a network utility maximization (NUM) problem, and then solved it using the Lyapunov drift plus penalty approach. Le and colleagues\textsuperscript{8} proposed a network utility framework for maximizing video quality using cellular network and local network communication resources, in which was formed by the D2D links. Xing and colleagues\textsuperscript{12,19} investigated the problem of improving video streaming quality while reducing the wireless service cost by utilizing multiple links simultaneously. Most of the existing approaches assume that all nearby terminals are completely cooperative and can also communicate with each other within one hop. Since the transmission cost in wireless networks is a significant issue that greatly affects the video quality, lots of efforts have been paid on the energy cost arising from streaming video in wireless scenarios.\textsuperscript{13,20–22} Duong et al.\textsuperscript{20} proposed an energy-aware approach for bit-level rate allocation in cellular networks with the assistance of D2D communications, in which they considered how to save communication resources for video streaming with higher transmission rates and larger sizes. Sickkinen et al.\textsuperscript{21} investigated the tradeoff in energy waste between prefetching small and large chunks of video content, and proposed an algorithm called eSchedule that used viewing statistics to predict viewer behavior and designed an energy optimal download strategy. In order to minimize the energy consumption of mobile devices and the load on cellular networks, Almowuena et al.\textsuperscript{13} proposed a strategy to concurrently utilize unicast and multicast for streaming video. Atawia et al.\textsuperscript{22} addressed the problem of predictive resource allocation for energy-efficient video streaming. In their work, by offering a mechanism to control the probability constraint satisfaction, operators may control the tradeoff between energy savings and the risks associated with erroneous predictions.

In order to take advantage of the diversity of wireless channels, some network coding approaches, for example, RLC, are usually adopted while streaming video content in wireless networks.\textsuperscript{8,10,11,14} As discussed in section “Video model,” one video content consists of a sequence of a group of pictures/frames (GoPs). Accordingly, one common way is to perform RLC across all frames within one GoP, and then the resulting coded chunks are delivered to all users via the wireless networks. All the video frames contained by one GoP can be decoded at terminals once they have received a certain number of coded chunks, no matter what the arriving order is.\textsuperscript{13,15,16} Thanks to the flexibility and adaptability of RLC, the users incapable of receiving sufficient coded chunks can be made up by the neighbors.

The previous works indicated that using both the cellular and local network resources can improve video transmission efficiency substantially. However, they often used unicast in the cellular downlink and broadcast in one-hop D2D network separately. In this case, the BS has to separately consider each UE’s requirement and prepare a unique coded-chunk sequence. Consequently, there are no duplicated chunks on any UE even if they are received from multiple sources.\textsuperscript{17,23} Such an approach is not scalable with the increasing number of UEs requesting the same video content simultaneously. Different from the previous approaches, in this work, both the BS and UEs can broadcast/multicast coded chunks to the neighbors and there is no need for an exclusive video sequence preparation in advance and no one-hop constraint as well. Thus, our approach can be more scalable.

System overview

Network model

Considering a cellular network scenario, multiple UEs are in a cell with one BS, where all UEs, denoted by $N = \{1, 2, \ldots, N\}$, are equipped with cellular and local network interfaces. Through the cellular network, the video chunks fetched from the remote video server are broadcasted to all UEs by the BS at a rate of $r_{bs}$. Note that if the broadcast rate of BS is adopted catering to all UEs, that is, extremely low transmission rate can be acceptable by some UEs with much worse-downlink condition, a lot of communication resources will be wasted, and some UEs even with better channel condition have to experience degraded video playback quality (such cases are quite common in reality due to the fact that receiving rates of UEs are possibly affected by their geographical remote location in the cell or sheltered by the buildings). In this model, the BS broadcast rate is only selected to cater to partial UEs with better-downlink channel condition, regardless of the worse-downlink UEs. As a result, the video chunks broadcasted to the worse-downlink UEs via cellular network are totally lost.

Note that some video chunks broadcasted to the better-downlink UEs are also possibly lost due to the
dynamic cellular network scenario. Although the BS can be used to identify the cellular downlink condition, the receiving probabilities of UEs could be different, which are independent and identically distributed (i.i.d.). Correspondingly, we assume that the UEs’ receiving probabilities from the cellular network, denoted by $p_j > 0$ ($j \in N$), that is, the loss probability, can be calculated by $1 - p_j$. Denote $P_{\text{thr}}$ as the receiving probability threshold, that is, its value is equal to 1 minus loss probability threshold. If the receiving probabilities of some UEs are lower than $P_{\text{thr}}$, such that they hardly recover the received video chunks, the transmissions between BS and such UEs are failed. Accordingly, if $p_j \leq P_{\text{thr}}$, the video chunks transmitted to UE $j$ by the BS are lost.

Through the local network interface, all the UEs can connect with each other directly or via local network access points indirectly, which forms a P2P local network. Therefore, each UE can receive video streams from not only the BS via the cellular downlink but also from peer UEs via the P2P network. In the local P2P network, the receiving probabilities, denoted by $p_{ij}$ from UE $i$ to UE $j$, are assumed to be i.i.d. Considering different cellular downlink conditions, those UEs unable to obtain desirable video quality from the cellular network can possibly purchase service from their peer UEs (with better cellular downlink condition) via the local P2P network. In other words, the UEs failed to meet the video requirement, which led to waiting for others’ help via the local P2P network at the monetary cost.

**Video model**

In a multi-layer SVC system, the GoP is the basic unit for video playback, which specifies the order of intra- and inter-frames being arranged.\(^5\) Generally, one video stream consists of a sequence of GoPs, and each GoP has a unique time for playback. This time is called playback deadline, meaning that the referred GoP has to be delivered to the terminals before this playback deadline, and it becomes invalid otherwise. Assume that each GoP has an equal playback duration, denoted by $T_{\text{GoP}}$, and it includes multiple quality layers, one base layer and several enhancement layers, denoted by the set $L = \{0, 1, \ldots, L\}$. The base layer, that is, Layer 0, guarantees a basic playback quality for video users, and the enhancement layers can provide an incremental video quality given all the previous layers being received. Note that in the following sections, we use the single term *layer* to represent the *layer in one GoP* for the purpose of briefness.

We can take H.264 SVC stream as an example.\(^24\) As shown in Figure 1, it illustrates a two-layer structure with temporal and spatial scalability, which includes one base layer (lower layer) and one enhancement layer (upper layer). The temporal scalability means that the frame rate of stream can be scalable according to the network conditions or the terminal constraint. It can be found that the frame rate is higher in the upper layer than that in the lower layer. Furthermore, each frame in the upper layer owns a larger size since it enhances the video quality by increasing the spatial resolution, and thus such scalability can be suitable for UEs with larger screens.

As discussed in the introduction, the RLC is adopted in our model, where each layer is partitioned into a number of equal-size chunks, and correspondingly, an RLC-encoded chunk is a linear combination of the partitioned chunks. With such a technique, the UE can recover one layer once it has received sufficient RLC-encoded chunks.
with respect to the specific layer. The number required to recover one layer is called decoding threshold. Given the decoding threshold $C(l)$ of Layer $l \in L$, this layer will be partitioned into $C(l)$ original chunks, and each of the RLC-encoded chunks is a linear combination of the $C(l)$ original chunks. Considering the layer dependency that the $l$th ($L \geq L > 0$) video quality requires all the lower layers including layer $l$ should be received. Different from the decoding threshold which is the minimum number to recover one layer, the layer dependency determines the video playback quality that the users can perceive. According to the work, we use the reward model to represent the UEs’ perceived video quality, which is given by

$$U(r) = e^{-\phi(r_{\text{min}})} - \theta$$

where $r$ is the video rate that is being played back, $R_{\text{max}}$ is the maximum video rate, and $\phi$ and $\theta$ are the video-specific parameters of the quality model. However, this reward model cannot be directly applied in our work, because the layer dependency should be guaranteed. Taking this into account, we specify that the lower layers should be delivered first, and then the higher layers both in cellular network and the local P2P network, but how to allocate the transmission time for each layer will affect the UEs’ playback quality.

### System process and problem description

Based on the network model and video model, the system process can be described as follows:

- The RLC-encoded video chunks fetched from the video server are first broadcasted to all the UEs under the cellular network. The BS needs to optimize the transmission duration for each layer considering the playback deadline.
- When one UE has received sufficient number of UEs that exceeds the decoding threshold, they can recover this specific layer correctly. In the following, such UE that can recover Layer $l$ ($l \in L$) is called $l$-matured UE.
- For an $l$-matured UE, it can then re-encode Layer $l$ with RLC for multicasting its peers in the local P2P network (to avoid duplicated RLC-encoded chunks, each node including the BS and UEs will be pre-allocated with an exclusive linear coefficient field, such that each RLC-encoded chunk is unique. Thus, once the number of received encoded chunks for one layer reaches the decoding threshold, that is, $C(l)$, the corresponding layer can be recovered, no matter which nodes they are generated from).

Note that there is no feedback for packet loss during both the cellular and P2P transmissions, and the BS or UEs simply generate new RLC-encoded chunks for transmission.

The addressed problem in this work is described as follows. Naturally, to maximize the overall video quality of all UEs is the main objective. Each UE’s video stream can come from two sources, one is from the cellular downlink and the other is from the peer UEs via P2P network. However, maximizing the video quality on the whole would sacrifice some UEs’ video quality for their experienced poor cellular downlink or P2P link conditions. There are several constraints that should be taken into account when we maximize the overall video quality. The first constraint is the transmission rate of one multicast source, which should be no larger than the minimum receiving rate of all its peer UEs. Considering the transmission power limitation, the multicast source rate should take the minimum value from the minimum receiving rate and its supportable maximum transmission rate. The second constraint is the minimum transmission rate for base layer delivery, which can guarantee the smooth playback of all the UEs with the basic video quality. The third constraint is the bound on the traffic volume, which means that the multicast source has a preferred traffic volume due to its transmission resources limitation. The fourth constraint is the transmission duration that each GoP should be delivered to the UEs before its playback deadline, and it becomes invalid otherwise. The last one is the budget constraint, which means that the peer UE pays an equivalent price for its receiving traffic from the multicast source, and the the total pay has to meet with the budget constraint. Considering these mentioned constraint, how to jointly optimize the BS and local P2P transmissions to maximize the video playback quality is the main objective of this work.

### Problem formulation

#### Receiving rate and traffic volume

Since all the UEs can work cooperatively under the cellular network and the local P2P network, the total received chunks of Layer $l$ by UE $i$ ($i \in \mathcal{N}$), denoted by $R_i^{(l)}$, come from two parts: the BS and its matured peers, which can be calculated by

$$R_i^{(l)} = r_{\text{bs}}t_{\text{bs}}^{(l)}p_i + \sum_{j \in \mathcal{N}_i} r_j^{(l)}p_{ji}$$
where \( r_j^{\text{max}} \) is the maximum supportable transmission rate of UE \( j \), and \( a_i \) is the maximum acceptable rate of UE \( i \). Here, we simply use \( r_j \) as the multicast rate without considering the link capacity constraint, but our work can easily be extended to this case where the link and path information of the local P2P network is known in advance.

Correspondingly, the achieved video rate at Layer \( l \) can be calculated by

\[
g_i^{(l)} = \frac{R_i^{(l)}}{T_{\text{GoP}}}\tag{4}
\]

which can also be achieved by aggregating all the incoming rates of Layer \( l \).

### The constraint for base layer delivery

If one UE cannot recover the base layer before its playback time, one stall event occurs and the UE has to stop the playback and wait for the chunks of base layer, even when the higher layers have been delivered. To guarantee no stall event happening in the video playback, it is necessary to guarantee that the base layer rates of all UEs are not less than the required minimum rate, which is given by

\[
g_i^{(0)} \geq g_{i,\text{min}}^{(0)} \forall i \in \mathcal{N}\tag{5}
\]

where \( g_{i,\text{min}}^{(0)} \) is the minimum rate that guarantees that the base layer can be recovered before its playback deadline.

### The constraint to be a multicast source

Note that UE \( j \) serving as a multicast source of Layer \( l \) has to be \( l \)-matured in advance, that is, it has received sufficient number of RLC-encoded chunks exceeding the decoding threshold, such that it can recover Layer \( l \) and re-encode it for transmission. We introduce an indicator function, \( B_j^{(l)} \), to denote whether UE \( j \) is \( l \)-matured

\[
B_j^{(l)} = \begin{cases} 1, & \text{if } R_j^{(l)} \geq C_j^{(l)} \\ 0, & \text{otherwise} \end{cases}
\]

### The bounds on the traffic volume

For an \( l \)-matured UE \( j \), it can serve as the multicast source of its peer UEs \( i \in \mathcal{M}_j \), given that some of them are \( l \)-unmatured. Each \( l \)-matured UE \( j \) has its own transmission cost concern, for example, transmission energy consumption and bandwidth consumption, and thus it has its own preferred transmission traffic limit, which can be represented by

\[
f_j^{(l)} = r_j t_j^{(l)} \leq \eta_j^{(l)} B_j^{(l)}, \forall j \in \mathcal{N}, \forall l \in \mathcal{L}\tag{7}
\]

where \( \eta_j^{(l)} \) is the maximum traffic of Layer \( l \) that UE \( j \) can transmit, and \( t_j^{(l)} \) is the total transmission duration of Layer \( l \), which can be given by

\[
t_j^{(l)} = \max_{i \in \mathcal{M}_j} \{ t_i^{(l)} \}\tag{8}
\]

Note that the upper bound of \( f_j^{(l)} \), that is, \( \eta_j^{(l)} B_j^{(l)} \), will be changed as UE \( j \)'s state transfers from the immatured to the matured one. Meanwhile, the lower bound of the traffic volume should be nonnegative, and thus we have

\[
f_j^{(l)} \geq 0, \forall j \in \mathcal{N}, \forall l \in \mathcal{L}\tag{9}
\]

### Transmission duration constraint

Due to the fact that each GoP has its own playback deadline, that is, it has to be delivered to all the UEs before this moment, otherwise, it becomes invalid for missing its playback deadline. According to the proposed system model, each GoP may experience two periods, namely, transmission in the cellular network and the local P2P network. The transmission durations in two parts are denoted by \( t_{bs}^{(l)} \) and \( t_{j}^{(l)} \), respectively. This constraint can be represented by

\[
\sum_{i \in \mathcal{L}} \left( t_{bs}^{(l)} + \sum_{j \in \mathcal{N}_i} t_j^{(l)} \right) \leq \alpha T_{\text{GoP}}, \forall i \in \mathcal{N}\tag{10}
\]

where \( \alpha \) is a tolerance coefficient, and \( \alpha T_{\text{GoP}} \) denotes the longest postponement for the video’s start-up playback. It is rational that the value of \( \alpha \) is not less than one \( T_{\text{GoP}} \), which can guarantee one GoP to be delivered during the last one’s playback duration. For instance, if \( \alpha = 1 \), all UEs can playback this video after one \( T_{\text{GoP}} \), that is, one GoP’s playback can be synchronized with the next GoP’s transmission.

### The budget constraint

According to the proposed system model, the UEs incapable of being matured from the cellular network can get help from the matured UEs at the monetary cost. The total received \( i \)th traffic of UE \( i \) from the matured UEs is calculated by

\[
R_{p2p,i}^{(l)} = \sum_{j \in \mathcal{N}_i} t_j^{(l)} p_j \tag{11}
\]

and thus the total payoff of UE \( i \) is \( R_{p2p,i}^{(l)} y \), where \( y \) is the unit price of data traffic. Given the UE \( i \)'s budget \( G_i^{(l)} \), the budget constraint should be satisfied, which is given by

\[
R_{p2p,i}^{(l)} y \leq G_i^{(l)} \quad \forall i \in \mathcal{N}\tag{12}
\]
Generally, each UE will have a higher budget on the lower layer, and a smaller one on the higher layer for the lower layers, playing a more significant role in video playback. Therefore, we have $G_i^{[0]} \geq G_i^{[1]} \geq \cdots \geq G_i^{[L]}$. The total income of UEs providing relay service is equal to the total payoff of unmatured UEs, and we need not take the payoff and income into the objective function design other than the constraint set.

**Objective function**

Our objective is to maximize the video quality reward of all UEs considering the traffic cost and other constraints. The primal optimization problem can be formulated as

\[
P_0 : \max \sum_{i \in \mathcal{N} \setminus \{0\}} \sum_{l \in \mathcal{L}} \left( U_i (g_i^{(l)}) - \omega_i C_i (f_i^{(l)}) \right) \tag{13}
\]

s.t. (5), (7), (9), (10) and (12)

where $U_0 = 0$ and $C_0 \geq 0$, which are the BS’s reward and cost, respectively. Because both $g_i^{(l)}$ and $f_i^{(l)}$ are the linear function with respect to cellular and P2P transmission durations, and their coefficients are positive, both functions have no affect on the convexity of the objective function.

**Remark 1.** The weight of cost function $C_i$, $\omega_i$, can be used to adjust video playback quality and the traffic cost.

**Remark 2.** In the primal optimization problem $P_0$, it can be found that the reward function $U_i$ is concave and twice differentiable. Similar to the work\(^8\) we also assume the cost function to be convex and twice differentiable. Therefore, the objective function of $P_0$ is concave.

In the primal optimization problem $P_0$, the constraints (5), (9), and (12) are linear. In constraint (7), the value of function $R_i^{(l)}$ will experience a 0–1 transformation when UE $i$’s state transfers from the unmatured to the l-matured one, which will hinder the solution of the primal problem. Meanwhile, in constraint (10), the cellular transmission duration and P2P transmission duration are couple together, which further increases the difficulty to obtain the optimal solution. To solve the optimization problem, first, we need to deal with the two difficulties.

**Approximation of the primal optimization problem**

**Decouple the cellular and P2P transmission durations**

In constraint (10), the tolerance-efficient $\alpha T_{GoP}$ shows the largest postponement for video’s start-up playback. The basic idea to decouple the cellular transmission and P2P transmission is to relax the time constraint (10), that is, to increase the value of $\alpha$ such that this constraint will become too relaxing, but other constraints can play more roles in finding an optimal solution. With this relaxation, the cellular transmission can be stopped before the first half of $\alpha T_{GoP}$, and the P2P transmission can be stopped before the second half. Considering an extreme case, if all the UEs allow a sufficient long start-up delay, the whole video content is delivered to terminals via the cellular and P2P network transmissions. As a result, all the UEs can achieve the highest playback quality.

We use an example to illustrate our basic idea ($\alpha = 2$). Figure 2 shows that the first GoP is playing back currently. Simultaneously, the transmissions of the second GoP in cellular network and the third GoP in P2P network are ongoing concurrently. Therefore, the P2P transmission, the cellular transmission, and the video playback can work in a pipeline manner.

As a start point, we use $\alpha = 2$ in this work, and thus both the cellular transmission and P2P transmission are constrained in a single GoP’s playback duration, that is, each single transmission is not larger than $T_{GoP}$. Therefore, the constraint (10) is transformed as follows

\[
\sum_{i \in \mathcal{L}} t_{i}^{(l)} \leq T_{GoP} \tag{14}
\]

\[
\sum_{i \in \mathcal{L}} \sum_{j \in \mathcal{N}_i} t_j^{(l)} \leq T_{GoP}, \forall i \in \mathcal{N} \tag{15}
\]

**Approximation of the indicator function**

The objective function and all constraints to be differentiable is the necessary condition as gradient-based solution techniques for convex/concave programming. Because constraint (7) is discontinuous, Problem $P_0$
cannot be solved by the gradient-based method directly. Inspired by the work of Gaudette et al., we introduce a continuous step function to approximate the binary function in constraint (7), The continuous step function is expressed by \( s(x^{(l)}_i) = 1/(1 + e^{-2kx^{(l)}_i}) \), in which \( x^{(l)}_i \) is the difference of receiving chunks and decoding threshold, that is, \( x^{(l)}_i = R^{(l)}_i - C^l \). The schematic diagram of the step function is shown in Figure 3.27.

The approximation of Problem P0

Since \( \eta^{(l)}_i \) is a constant and constraint (16) is quasi-convex, Problem P0 can be approximated to Problem P1, which is given by

\[
P1: \max \sum_{i \in N_i(0)} \sum_{l \in L} \left( U_i(g^{(l)}_i) - \omega_i C_i(f^{(l)}_i) \right)
\]

s.t. (5), (9), (12), (14), (15) and (16)

Since the step function is quasi-convex and all the variables are continuous, the above formulation is a concave optimization problem, which can be solved by the gradient-based method.28,29 In the following section, we develop a distributed traffic allocation algorithm for cellular and P2P transmissions.

Distributed algorithm design

In order to simplify the development of the distributed algorithm, we first neglect the budget constraint (12). If the solution cannot satisfy the constraint, we can deduct the purchasing traffic from the higher layer to the lower one correspondingly. Second, the transmission duration constraints (14) and (15) can also be neglected, because the \( T_{GoP} \) in practice is 0.5 s, which can be sufficient for cellular downlink and P2P transmissions. However, if the solution cannot satisfy this constraint, we can deduct the higher layer transmission time correspondingly. Therefore, the transmission traffic including cellular and P2P networks can be obtained, based on which the transmission durations can be inferred.

Let \( \lambda \) and \( \beta \) be the Lagrange multipliers with respect to constraints (5) and (9), respectively, and then the Lagrangian function of constraint (17) is given by

\[
L(\lambda, \beta) = \max \sum_{i \in N_i(0)} \sum_{l \in L} \left( U_i(g^{(l)}_i) + \omega_i C_i(f^{(l)}_i) \right)
+ \lambda^{(l)}_i [g^{(l)}_i - g^{(0)}] + \beta^{(l)}_i f^{(l)}_i
\]

s.t. (16)

where \( g^{(l)}_i = 0 \) if \( l \geq 1 \) and \( g^{(0)} = g^{\min} \). Correspondingly, the dual problem of constraint (18) is

\[
\min L(\lambda, \beta)
\]

s.t. \( \lambda^{(l)}_i \geq 0, \beta^{(l)}_i \geq 0, \quad \forall i \in N \cup \{0\}, \forall l \in L
\]

To attain a distributed algorithm, we can decompose Problem P1 to be \( L(\lambda^{(l)}_i, \beta^{(l)}_i) \) such that each node, including the BS and UEs, can figure out its own traffic volume. For any UE \( i \), the decomposed Problem P1 could be, given that

\[
L(\lambda^{(l)}_i, \beta^{(l)}_i) = \max \sum_{l \in L} \left( U_i(g^{(l)}_i) + \omega_i C_i(f^{(l)}_i) \right)
+ \lambda^{(l)}_i [g^{(l)}_i - g^{(\min)}] + \beta^{(l)}_i f^{(l)}_i
\]

s.t. (16)

Based on the designed model, the basic video quality should be guaranteed, that is, inequality (5) holds for base layer. Other layers \( (l \geq 1) \) have no such constraint, that is, \( g^{\min}_l = 0/l = 1,2,…,L \). According to constraints (2), (4), and (7), the receiving rate \( g^{(l)}_i \) can be considered as the linear function of \( f^{(l)}_i \) and \( j \in N_i \). Based on the Karush–Kuhn–Tucker (KKT) conditions or other optimization tools, for example, fmincon, the
optimal traffic volume of each UE can be obtained, denoted as \( \text{opt}(f_{l}^{(i)}) \). Accordingly, the optimal receiving traffic volume can be obtained, denoted as \( \text{opt}(g_{l}^{(i)}) \).

Such optimal traffic volume can be used in the calculation of the objective value if a sub-gradient method is employed.

Accordingly, the sub-gradient method is employed to update the Lagrangian multipliers \( \lambda \) and \( \beta \) iteratively:

\[
\lambda_{i}^{(l)}(m + 1) = \left[ \lambda_{i}^{(l)}(m) + \varrho (g_{i}^{(l)} - g_{\text{min}}^{(l)}) \right]^{+} \quad (21)
\]

\[
\beta_{i}^{(l)}(m + 1) = \left[ \beta_{i}^{(l)}(m) + \varrho f_{i}^{(l)} \right]^{+} \quad (22)
\]

where \( g_{\min}^{(l)} = 0(l = 1, 2, \ldots, L); m \) is the iteration number; \( \varrho > 0 \), is a small-enough constant step size; and \([\cdot]^{+} = \max(0, \cdot)\). The pseudo-code is shown in Algorithm 1, in which each node updates its own \( \lambda, \beta \) and calculate the quantity of chunks to be broadcasted locally. With a small-enough step size \( \varrho \), Algorithm 1 can converge after a limited iterations.

### Performance evaluation

#### Simulation settings

In this section, numerical simulation results are provided to demonstrate the performance of the proposed approach. In our simulations, there are 10 UEs in one cellular network with one BS. Considering the dynamic local P2P network, the P2P connections are generated randomly, and each UE is served by at most two peers. The topology is shown in Figure 4, where the UEs under the dotted red line cannot be satisfied for their low receiving probabilities. The transmission rates and receiving probabilities are set as shown in Table 1. Note that in the cellular receiving probability settings, we only show the receiving probabilities of UEs 1–5, other UEs under the dotted red line in Figure 4 with receiving probabilities lower than the minimum receiving probability (\( P_{\text{thr}} = 0.8 \)) cannot receive the video chunks correctly. Note that the topology is manually configured in the last simulation (Figure 4). In the P2P

### Table 1. Transmission/receiving rate/probability settings.

| Rate/probability          | Cellular network | P2P network     |
|---------------------------|------------------|-----------------|
| Receiving probability     | Probability 1: 0.95 0.95 0.95 0.90 0.90 | Probability 1: (0.80 0.85) |
|                           | Probability 2: 0.99 0.99 0.99 0.95 0.95 | Probability 2: (0.85 0.90) |
|                           | Probability 3: 1.00 1.00 1.00 0.95 0.95 | Probability 3: (0.90 0.95) |
|                           | Probability 4: 0.99 0.99 0.99 0.99 0.99 | Probability 4: (0.95 1.00) |
| Transmission rate (bytes/s) | Rate 1: 3 * 10^4 | Rate 1: 3 * 10^4 |
|                           | Rate 2: 1 * 10^4 | Rate 2: 1 * 10^4 |
|                           | Rate 3: 1 * 10^4 | Rate 3: 1 * 10^4 |

---

**Algorithm 1. Distributed algorithm for cellular and P2P networks transmission durations (DATD).**

Initialize \( \alpha, \lambda \) and \( \beta \);
Flag = true;
while flag do
  for \( i = 0; i \leq N \) do
    for \( l = 0; l \leq L \) do
      Node \( i \) calculates its traffic \( f_{l}^{(i)} \) and \( g_{l}^{(i)} \), denoted by \( \text{opt}(f_{l}^{(i)}) \) and \( \text{opt}(g_{l}^{(i)}) \), for each layer based on (19);
    end
    Place \( \text{opt}(f_{l}^{(i)}) \) and \( \text{opt}(g_{l}^{(i)}) \) into set \( F^{(\text{last})} \) and \( G^{(\text{last})} \);
    Node \( i \) updates the Lagrange Multipliers \( \lambda_{i}^{(l)} \) and \( \beta_{i}^{(l)} \) using (21) and (22);
  end
  if \( (F^{(\text{last})} = F^{*}) \) \( (G^{(\text{last})} = G^{*}) \) then
    Flag = false;
    Break;
  end
  \( \text{set} \) \( F^{*} = F^{(\text{last})}; \)
  \( \text{set} \) \( G^{*} = G^{(\text{last})}; \)
end
return \( F^{*} \) and \( G^{*} \);

---

 Performance evaluation

**Simulation settings**

In this section, numerical simulation results are provided to demonstrate the performance of the proposed approach. In our simulations, there are 10 UEs in one cellular network with one BS. Considering the dynamic local P2P network, the P2P connections are generated randomly, and each UE is served by at most two peers. The topology is shown in Figure 4, where the UEs under the dotted red line cannot be satisfied for their low receiving probabilities. The transmission rates and receiving probabilities are set as shown in Table 1. Note that in the cellular receiving probability settings, we only show the receiving probabilities of UEs 1–5, other UEs under the dotted red line in Figure 4 with receiving probabilities lower than the minimum receiving probability (\( P_{\text{thr}} = 0.8 \)) cannot receive the video chunks correctly. Note that the topology is manually configured in the last simulation (Figure 4). In the P2P...
transmission, the UEs’ receiving probabilities are uniformly distributed in the specific intervals.

Considering load balance of the network, we use $C_i(f_i^{(l)}) = (f_i^{(l)})^2$ as the objective function. It means that, for each UE, the cost of multicast traffic per chunk will be increased much faster with the increasing number of broadcasted chunks. The weights of P2P network transmissions ($w_i, i \in N$) are identically set to $10^{-8}$ (the weight of P2P network transmission is given very small, because that the UE’s reward is a normalization measurement from 0 to 1, but the number of transmission video chunks is high to tens or even hundreds of thousands. That is to say, the objective function would not be affected effectively, given a comparable higher weight). The vide sequence “Sony Demo” is used in our simulation, which is encoded with the GoP format of “G16B15,” the solution of “352 × 288,” and the frame rate of 30 frames/s using the JSVM 9.14 tools. The playback duration $T_{GoP}$ of one GoP is 0.5 s. In the reward model, the parameter $\phi = 0.16$ and $\theta = 0.66$. The simulation tool is MATLAB.

**Simulation results and analysis**

Note that all the receiving probabilities of UEs are generated randomly, and each UE receives the video chunks according to the generated probability. For instance, the receiving probability, 0.95, of one UE from the BS means that this probability is randomly distributed in interval (0.95, 1). In the P2P network, the receiving probabilities of all UEs are uniformly distributed in the specified interval, such as (0.80, 0.85). Due to the RLC technology employed in our work, one GoP can be recovered or not only depends on the number of encoded chunks received by the UE. That is, the order of lost encoded chunks in one GoP has no effect on its recovery.

Figure 5 shows the achieved rewards by the UEs under different receiving probabilities, where $x$ axis denotes the UE’s subscript, and $y$ axis denotes the rewards of different UEs. Note that the value of $y$ is a scalar according to constraint (1), whose minimum value is 0 and maximum is 1. In the simulation, the transmission rates in both the cellular network and the P2P network are set to Rate 2, respectively, and the receiving probabilities of UEs in the local P2P network are uniformly distributed in the interval of Probability 2. It can be found that with a higher receiving probability in the cellular network, the UE can achieve a higher reward, that is, the video playback quality is comparatively higher than others with lower receiving probabilities. From Table 1, it can be found that UE4’s
receiving probability is lower than UE1, UE2, and UE3, and thus the UE4’s reward is decreased accordingly. Correspondingly, the received traffic from the cellular network is shown in Figure 6, where \( x \) is the UE’s subscript and \( y \) is the number of chunks received by the UEs. Obviously, the number of received chunks is consistent with the achieved reward for the same UE.

The traffic costs including UEs and BS are shown in Figure 7, where \( x \) is the UE’s subscript and \( y \) is the quadratic of number of chunks broadcasted by one of the UEs or BS. In the simulation, the parameter settings are same with that in Figure 5. The UEs (1–5) undertake more traffic transmission task due to their higher cellular receiving probabilities. Correspondingly, given a higher receiving probability in the P2P network, the multicasting traffic volume will be higher than that with a lower receiving probability, as shown in Figure 8, where \( x \) is the UE’s subscript and \( y \) is the number of chunks broadcasted by one of the UEs. In contrast, the traffic costs of UEs (6–10) are zero, because they cannot receive video chunks from the BS other than from their matured peers. The traffic costs of UEs (1–5) have a small difference for their similar receiving probabilities from the cellular network. The BS’s traffic cost is little lower than the UEs (1–5) for there is no loss consideration from the video server to the cellular BS.

The following simulation is conducted with different cellular transmission rates, as shown in Figure 9, where \( x \) is the UE’s subscript and \( y \) is the UE’s reward. In the simulation, the receiving probabilities in the cellular and P2P network are set to Probability 1 and Probability 3, respectively. It can be found that the achieved rewards of UEs (6–10), incapable of receiving data from the cellular network, are affected by the cellular transmission rate. With a higher cellular transmission rate, the UEs (6–10) can be greatly benefitted from the matured UEs (1–5). In contrast, the rewards of those UEs will be decreased rapidly. The reason is that the UE matured from cellular network with a smaller
cellular receiving probability will pay more efforts to achieve the desirable video playback quality, which hinders it to multicast more video chunks to its unmatured peers over the local P2P network.

In this simulation, we use a manually configured P2P network topology with different P2P transmission rates, where the P2P connections are as follows: UE 1 → UE 6–10, UE 2 → UE 7–10, UE 3 → UE 8–10, UE 4 → UE 9–10, and UE 5 → UE 10, such that the unmatured UEs can be served by multiple peers. The simulation result is shown in Figure 10, where $x$ is the UE's subscript and $y$ is the UE's reward. It can be found that the rewards of UEs (6–10), that are incapable of receiving data from the cellular network, increase with both the P2P transmission rate and the number of peers that provide relay service.

To verify the convergence of the proposed distributed approach, we conduct the last simulation with Probability 4 in both P2P and cellular networks. The simulation result is shown in Figure 11, where the line with legend Convergence rate 1 stands for the proposed approach, and the line with legend Convergence rate 2 stands for the approach in Feng et al. It can be found that the proposed approach can converge to the optimal value obtained by the centralized approach. Since more factors are considered in our model, for example, budget limit and dynamic P2P connections, the convergence rate of the proposed approach is a little slower than that in Feng et al., but it is still rational considering the more complicated network scenario.

**Conclusion**

In this article, we have studied the problem of transmitting multi-layered SVC video in the cellular network. Our objective is to tradeoff the achieved video playback quality and traffic cost considering the local P2P connections formed by the UEs. We first construct an MINLP model, in which only the matured UEs can serve as multicast source in the local P2P network. Then, we convert it to a convex model by introducing a step function and decouple the transmission durations between cellular and P2P networks by postponing the UEs' start-up playback. Thereafter, we developed a distributed algorithm to obtain its near-optimal value. Extensive simulations are performed to demonstrate the efficiency of our approach. For the future work, we will focus on investigating incentive mechanism for UEs' cooperative transmission in wireless network.

**Acknowledgements**

The authors thank Prof. Jizhong Han’s efforts on smoothing the whole paper language, who is a full stuff in the Institute of Information Engineering, Chinese Academy of Sciences.

**Declaration of conflicting interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Natural Science Foundation of China (nos 61502118, 61402127, and 61370212), the Fundamental Research Fund for the Central Universities (no. HEUCF160602), and the Natural Science Foundation of Heilongjiang Province in China (nos F2016028, F2016009, and F2015029).

**References**

1. Tavakoli S, Egger S, Seufert M, et al. Perceptual quality of http adaptive streaming strategies: cross-experimental analysis of multi-laboratory and crowdsourced subjective studies. *IEEE J Sel Area Comm* 2016; 34(8): 2141–2153.
2. Zhou H, Ji Y, Wang X, et al. Joint resource allocation and user association for SVC multicast over heterogeneous cellular networks. *IEEE T Wirel Commun* 2015; 14(7): 3673–3684.

3. Bakhshali A, Chan WY, Blostein SD, et al. QoE optimization of video multicast with heterogeneous channels and playback requirements. *EURASIP J Wirel Comm 2015*; 2015(1): 260.

4. He Z, Cai Z, Cheng S, et al. Approximate aggregation for tracking quantiles and range countings in wireless sensor networks. *Theor Comput Sci* 2015; 607: 381–390.

5. Van der Auwera G, David PT and Reisslein M. Traffic and quality characterization of single-layer video streams encoded with the H.264/mpeg-4 advanced video coding standard and scalable video coding extension. *IEEE T Broadcast* 2008; 54(3): 695–718.

6. Condoluci M, Araniti G, Molinaro A, et al. Multicast resource allocation enhanced by channel state feedbacks for multiple scalable video coding streams in LTE networks. *IEEE T Veh Technol* 2016; 65(5): 2907–2921.

7. Wang X, Chen J, Dutta A, et al. Adaptive video streaming over whitespace: SVC for 3-tiered spectrum sharing. In: *Proceedings of the 2015 IEEE conference on computer communications (INFOCOM)*, Hong Kong, 26 April–1 May 2015, pp.28–36. New York: IEEE.

8. Keller L, Le A, Cici B, et al. Microcast: cooperative video streaming on smartphones. In: *Proceedings of the 10th international conference on mobile systems, applications, and services*, Ambleside, 25–29 June 2012, pp.57–70. New York: ACM.

9. Cheng S, Cai Z and Li J. Curve query processing in wireless sensor networks. *IEEE T Veh Technol* 2015; 64(11): 5198–5209.

10. Abedini N, Sampa S, Bhattacharyya R, et al. Realtime streaming with guaranteed QoS over wireless D2D networks. In: *Proceedings of the fourteenth ACM international symposium on mobile ad hoc networking and computing*, Bangalore, India, 29 July–1 August 2013, pp.197–206. New York: ACM.

11. Bethanabhotla D, Caire G and Neely MJ. Adaptive video streaming for wireless networks with multiple users and helpers. *IEEE T Commun* 2015; 63(1): 268–285.

12. Xing M, Xiang S and Cai L. A real-time adaptive algorithm for video streaming over multiple wireless access networks. *IEEE J Sel Area Comm* 2014; 32(4): 795–805.

13. Almowuena S, Rahman MM, Hsu CH, et al. Energy-aware and bandwidth-efficient hybrid video streaming over mobile networks. *IEEE T Multimedia* 2016; 18(1): 102–115.

14. Vukobratovic D and Stankovic V. Unequal error protection random linear coding strategies for erasure channels. *IEEE T Commun* 2012; 60(5): 1243–1252.

15. Ostovari P, Wu J, Khreishah A, et al. Scalable video streaming with helper nodes using random linear network coding. *IEEE/ACM T Network* 2016; 24(3): 1574–1587.

16. Thomos N, Kurdoglu E, Frossard P, et al. Adaptive prioritized random linear coding and scheduling for layered data delivery from multiple servers. *IEEE T Multimedia* 2015; 17(6): 893–906.

17. Wu J, Cheng B, Yuen C, et al. Distortion-aware concurrent multipath transfer for mobile video streaming in heterogeneous wireless networks. *IEEE T Mobile Comput* 2015; 14: 688–701.

18. He Z, Cai Z and Wang X. Modeling propagation dynamics and developing optimized countermeasures for rumor spreading in online social networks. In: *Proceedings of the 2015 IEEE 33th international conference on distributed computing systems (ICDCS)*, Columbus, OH, 29 June–2 July 2015, pp.205–214. New York: IEEE.

19. Xiang S, Xing M, Cai L, et al. Dynamic rate adaptation for adaptive video streaming in wireless networks. *Signal Process: Image* 2015; 39: 305–315.

20. Duong TQ, Vo NS, Nguyen TH, et al. Energy-aware rate and description allocation optimized video streaming for mobile D2D communications. In: *Proceedings of the 2015 IEEE international conference on communications (ICC)*, London, 8–12 June 2015, pp.6791–6796. New York: IEEE.

21. Siekkinen M, Hoque MA and Nurminen JK. Using view statistics to control energy and traffic overhead in mobile video streaming. *IEEE/ACM T Network* 2016; 24(3): 1489–1503.

22. Atawia R, Abou-Zeid H, Hassanein HS, et al. Joint chance-constrained predictive resource allocation for energy-efficient video streaming. *IEEE J Sel Area Comm* 2016; 34(5): 1389–1404.

23. Zhou R, Khemmarat S, Gao L, et al. Boosting video popularity through keyword suggestion and recommendation systems. *Neurocomputing* 2016; 205: 529–541.

24. Schwarz H, Marpe D and Wiegand T. Overview of the scalable video coding extension of the H.264/AVC standard. *IEEE T Circ Syst Vid* 2007; 17(9): 1103–1120.

25. Hu H, Zhu X, Wang Y, et al. QoE-based multi-stream scalable video adaptation over wireless networks with proxy. In: *Proceedings of the 2012 IEEE international conference on communications (ICC)*, Ottawa, ON, Canada, 10–15 June 2012, pp.7088–7092. New York: IEEE.

26. Gaudette B, Hanumaiah V, Vrudhula S, et al. Optimal range assignment in solar powered active wireless sensor networks. In: *Proceedings of the 2012 IEEE conference on computer communications (INFOCOM)*, Orlando, FL, 25–30 March 2012, pp.2354–2362. New York: IEEE.

27. Feng G, Lin J, Zhang Y, et al. Joint optimization of downlink and D2D transmissions for SVC streaming in cooperative cellular networks. In: *Proceedings of the international conference on wireless algorithms, systems, and applications*, Bozeman, MT, 8–10 August 2016, pp.149–161. Cham: Springer.

28. Zhang Y, He S, Chen J, et al. Distributed sampling rate control for rechargeable sensor nodes with limited battery capacity. *IEEE T Wirel Commun* 2013; 12(6): 3096–3106.

29. Zhang Y, He S and Chen J. Data gathering optimization by dynamic sensing and routing in rechargeable sensor networks. *IEEE/ACM T Network* 2016; 24(3): 1632–1646.