Notice of Violation of IEEE Publication Principles

"Group-Aware Delay-Constrained Video Transmission Over Multihomed Device-to-Device Networks"
by Junfei Huang and Zhaowen Lin,
in IEEE Access, Volume 5, 2017, pp 2651-2664

After careful and considered review of the content and authorship of this paper by a duly constituted expert committee, this paper has been found to be in violation of IEEE's Publication Principles.

This paper is a duplication of the original text from the paper cited below. The original text was copied without attribution (including appropriate references to the original author(s) and/or paper title) and without permission.

"Group-Aware Delay-Constrained Streaming in Wireless Device-to-Device Networks"
by Xiaoyi Zhang, Jiawei Liang, Jiyan Wu, and Lin Zhang,
in the ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM). Submitted December 26, 2016

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 20, 2017; accepted February 11, 2017, date of publication February 20, 2017, date of current version March 15, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2671900

Group-Aware Delay-Constrained Video Transmission Over Multihomed Device-to-Device Networks

JUNFEI HUANG AND ZHAOWEN LIN
Network and Information Center, Institute of Network Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China
National Engineering Laboratory for Mobile Network Security, Beijing 100876, China

Corresponding author: J. Huang (huangjunfei@bupt.edu.cn)

This work was supported in part by the National High Technology Research and Development Program of China (863 Program) under Grant 2013AA014702, in part by the Fundamental Research Funds for the Central Universities under Grant BUPT2016RC48 and Grant 2014ZD03-03, and in part by the National Natural Science Foundation of China under Grant 61601041.

ABSTRACT The technological advancements in wireless communication systems enable mobile users to leverage different radio interfaces (e.g., cellular and WiFi) for concurrent data transmission. However, the existing transmission schemes do not seriously consider the problem of real-time video multicast to a cluster of co-located multihomed mobile devices. Conventionally, each client fetches the video streaming to the best of its capability, and this results in competing resources that degrade user-perceived video quality. Several literatures investigated the problem of using cellular to obtain video contents from the remote server and sharing them through WiFi. However, the stringent delay constraint of real-time video is not addressed in these solutions. In this paper, a cooperative transmission scheme is proposed to tackle the problem. First, a mathematical framework dubbed (Group-Aware Delay-COnstraint) is developed to formulate the delay-constrained goodput maximization problem of real-time video transmission to a group of multihomed mobiles. Second, a dataflow distribution mechanism is presented to conserve the cellular bandwidth and maximize user experience. Then, a prototype is implemented on the Android platform involving real-time video encoded with H.264 codec. Experimental results show that the proposed scheme achieves appreciable improvements over the reference schemes in video peak signal-to-noise ratio, end-to-end delay, and goodput.

INDEX TERMS Video transmission, delay constraint, multihoming, device-to-device communications.

I. INTRODUCTION

Driven by the technological advancements in wireless communication systems, recent years have witnessed the tremendous growth of mobile video traffic over the Internet. The popularity of powerful mobile terminals promotes the surprising proliferation of mobile video services. As released in the recent report of US IT research firm Gartner [21], global computing device shipments has reached 2.4 billion in 2016, of which 82% are smartphones. Video streaming accounted for 55% percent of the mobile traffic usage over the Internet in 2015 and will reach 75% percent by the end of 2020 [12]. An important ongoing trend accompanying this tremendous growth is the popularity of high definition (HD) video services (e.g., cloud gaming, live sports program, etc.).

In spite of the rapid advancement in mobile communication technologies, the network resources of existing 3G/4G systems are still limited compared to the ever-growing video traffic.

Credit Suisse reported that in 2012, 23% of base stations globally have utilization rates of more than 80-85% in peak hours, up from 20% in 2011 [24]. In addition to traffic consumption, mobile video services have a certain timeliness and are sensitive to delay and jitter.

In delay-sensitive video services [22], goodput [11] is key performance metric for real-time traffic over wireless networks and overdue packets cannot contribute to the decoding process. Goodput differs from throughput as it indicates the amount of data successfully received by the destination within a requested deadline. The video applications demand on strict requirement of instantaneity but comfy toleration on video distortion. The fluctuation and unreliability of the communication paths in wireless networks [25], in concert with the long delay, pose crucial challenges to cope with the stringent Quality of Experience (QoE) requirements. Due to competion for limited radio resources, the service quality of mobile networks is even worse in crowded places (e.g., subway stations), whereas in such places online video requests, for instance live videos, are frequently generated. Therefore, it is important to consider efficient utilization of the limited radio resources to satisfy the ever-growing demand for video services, as well as providing better QoE level (for example lower delay, etc.).

In this paper, we consider the following scenario: a group of mobile users are geographically located close to each other and watch the same real-time video content. Each user within the group will download same video data via the cellular network interface, which results in the wasting of access network bandwidth. Due to the limitation of cellular throughput, the quality of video services will be degraded as well. To make things worse, since the users are competing with each other on the limited cellular network resources, the video quality will be further reduced.

The technology progress in wireless communications and network infrastructures enables mobile users with ubiquitous access to the Internet, e.g., cellular networks (HSDPA, LTE), and wireless local area networks (802.11 family). And the state-of-the-art smartphones are able to support concurrent multi-path transfer via heterogeneous networks [43]. Consequently, in this paper, we leverage multi-interface data transfer on each smartphone for video sharing among smartphones within the group. Each mobile device is able to simultaneously access two networks through the cellular interface and the WiFi interface. The cellular interface is used to perform Long Range Transmission (LRT), i.e., to connect the video server and download video data. The WiFi interface is adopted for the Short Range Transmission (SRT), i.e., to connect other devices within the group and share the data downloaded from the cellular network with other group members via a local WiFi network.

Thus, the downloading rate of a specific smartphone can be reached to the aggregation of all cellular links within the group.

To achieve this objective, this research develops Group-Aware Delay-COnstraint (GADCO) mechanism. The proposed solution establishes a device-to-device (D2D) content sharing framework using heterogeneous wireless networks as shown in Fig. 1. The video server is used to allocate video data rate into different sub-flows according to the path status and delay constraint of the cellular communication connections. After fetching video content from the remote server, the group exchange specific video datas via local D2D networks. The proposed system aims at maximizing the goodput of entire group while satisfying the video quality requirement.

The contributions of the paper are summarized as follows:

- We formulate a mathematical framework dubbed GADCO to state the delay-constrained goodput maximization problem of a group watching real-time video streaming using different radio interfaces concurrently.

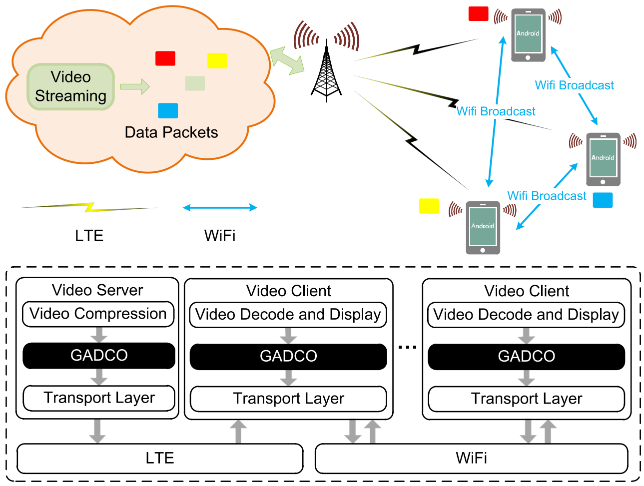


FIGURE 1. Scenario of the Proposed GADCO.

- We develop a data flow distribution algorithm utilizing the hierarchical structure using heterogeneous wireless networks to conserve the cellular bandwidth and maintain the user experience while satisfying the delay constraint.
- The proposed prototype on Android platform enables the proposed mechanism and actualizes D2D pseudo-broadcast scheme to evaluate the system performance in the real-world environment.
- Experimental results involving real-time video encoded with H.264 codec demonstrates that:
  - GADCO reduces the average end-to-end delay by up to 99.5 and 75.9 ms compared to GDASH [48] and EDPF [10], respectively.
  - GADCO increases the average goodput by up to 395 and 360 Kbps compared to GDASH and EDPF, respectively.
  - GADCO improves the average video Peak Signal-to-Noise Ratio (PSNR) by up to 3.1 and 2.8 dB compared to GDASH and EDPF, respectively.

The remainder of this paper is organized as follows. In Section II, we briefly survey and discuss the related work to this research. Section III presents the system model design and problem statement. The implementation of the proposed GADCO mechanism is described in Section IV. Section V provides the performance evaluation on real-world testbed and conclusion are given in Section VI.

II. RELATED WORKS

In this section, we discuss the related works on the problems of: 1) multi-interface transmission; 2) D2D cooperation; and 3) real-time video transmission. As far as we can understand, there are two typical ways for multi-path data transfer:

- Multiple Access Point solutions (MAP), under such situation there are more than one access points (for instance, a base station or an eNodeB acts as one access point, a WiFi AP act as another access point)

for each smartphone [46]. A smartphone shall access the cellular network via the cellular interface, and at the same time utilize other interfaces (for example, WiFi interface) to connect the other access points respectively [43]. The introduction of multiple access points increases the transmission bandwidth. An effective approach [8], [25], [29] is to enable a single smartphone using multiple interfaces (e.g. the cellular and the WiFi interface) to implement multi-path data transfer simultaneously and thus reduce the pressure of mobile networks and in addition provide better user experiences for video services. However, the smartphones are still competing transmission resources allocated by these access points.
- Single Access Point solutions (SAP), in such scenario there is only a single access point, which is usually the cellular base station. Each smartphone connects to this access point to compete limited transmission resources using the cellular interface, while at the same time using another interface (Blue Tooth or WiFi) to form a local network and perform data sharing within this network [48]. Unlike the former type, besides competing resources allocated by the access point, smartphones also cooperates with each other within the local sharing network using another interface simultaneously [24].

A. SAP VIDEO SYSTEM

Several solutions are proposed in previous works for video streaming and cooperative sharing in adaptive streaming systems. The scenario of these works are similar with that in this thesis. The authors in [34] proposed a system that allows smartphones to utilize the cellular connections and the D2D links (Bluetooth or WiFi) in order to maximize the video quality. Xi et al. [45] proposed a multi-path sharing framework with a cloud-based live stream server. By computing and transcoding the original stream segment according to the accumulated bandwidth of the devices' group, the server will improve the DASH service, then divide the transcoded stream segment into fragments for each device to fetch. Zhang et al. [49] designed a system called DECOMOD, which is a social-aware cooperative streaming framework. In this system, local friends serve as peers such that an aggregated bandwidth can be achieved to ensure a high streaming quality.

The solutions mentioned above enable the reliable transmission strategy and since the video streaming is completely delivered, the integrity of the file transmission is guaranteed. However, the non-timely video packet transmission schemes only consider successful delivery and the delay-constraint (which is the special feature in real-time applications) is completely ignored. The mobile terminal may suffer the halt phenomena caused by depletion of the video buffer, which will significantly affect the user experiences, especially for online live video services. In the system proposed in [47], smartphones were grouped into collaborative clusters according to a low-complexity clustering algorithm. In each cluster, there is a cluster head multicasting the content to other cluster members using a short-range wireless communication technology. The system in [1] allowed smartphones to use multiple interfaces for data transmission while at the same time minimizing the usage of their base-station-to-device (B2D) interfaces (3G/4G) to reduce cost, and try maintaining synchronous reception and play-out of content. However, the goodput and video distortion between frames is overlooked which impact the QoE of users. However, the advantages of the aggregate bandwidth is dissipated and may lead to the degradation of video quality and due to delay and/or jitter problems.

B. MAP VIDEO SYSTEM

The idea of using multiple interfaces of mobile devices has been explored before but not in the same way as in this work. Chebrolu et al. [10] presented a network layer architecture that enables diverse multiaccess services. And an algorithm called Earliest Delivery Path First (EDPF) is proposed, which ensures packets meet their playback deadlines by scheduling packets based on the estimated delivery time of the packets. The authors in [50] developed an analytical framework which optimizes rate allocation based on observed available bit rate (ABR) and round-trip time (RTT) over each access network and video distortion-rate (DR) characteristics. Alpcan et al. [3] investigated a novel robust flow control framework for heterogeneous network access by devices with multi-homing capabilities. In addition, the authors also proposed an H-infinity-optimal control formulation, which was used for allocating rates to devices on multiple access networks with heterogeneous time-varying characteristics. A framework is proposed in [13] to consider the problem of scalable video streaming from a server to multi-network clients over heterogeneous access networks while minimizing the distortion of the received videos. The authors in [42] believed that the end-to-end video frame delay was a severely challenging problem for high definition online video services, and proposed SFL, which is a novel scheduling approach, and deliberately splits large-size video frames into sub-frames and dispatches each of them onto a different wireless network to the multi-homed client. However, the devices equipped with multi-interfaces in the solutions above strengthen their streaming capacity on their own and ignore the potential cooperative opportunity, which may lead to the resource competition on the limited wireless networks.

C. MULTIPLE INTERFACES FILE SYSTEM

Several works discussed the usage of multi-interfaces in file delivery systems. Taking the social ties and geographical proximity into account, [18] considered a scenario in which device-to-device and cellular connections are used to disseminate the content. The authors [32] proposed an infrastructure that exploits wireless diversity (channel diversity, network diversity, and technology diversity) to provide improved data performance for wireless data users. Boldrini el al. [7] proposed a context-aware framework, which is used for routing and forwarding in opportunistic

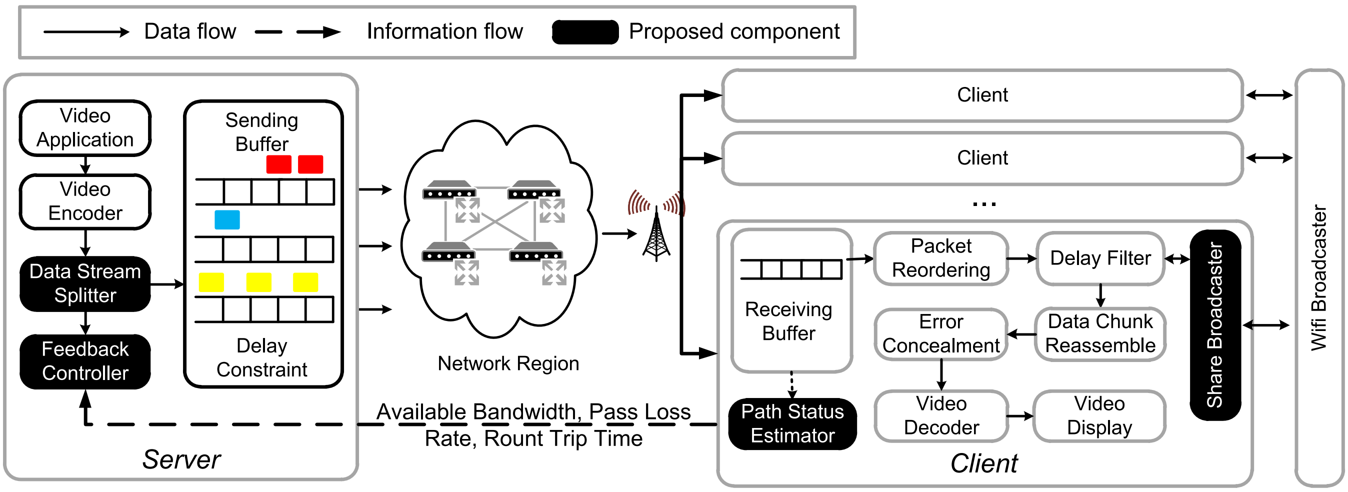


FIGURE 2. Overview of Group-Aware Delay-Constraint (GADCO) Model.

networks. Prabhavat et al. [29] exploited cellular and WiFi interfaces simultaneously to create multiple paths to mobile devices. Soroush et al. [38] investigated creating concurrent connections to multiple WiFi APs from highly mobile clients, and presented a system called Spider. The authors in [17] proposed BUBBLE, which is a social based forwarding algorithm to improve the forwarding efficiency significantly compared to oblivious forwarding schemes. However, the current schemes always schedule the file data traffic in a content-agnostic fashion without considering the complex video streaming characteristics such as the certain timeliness and sensitivity to delay and jitter requested by mobile video services.

III. SYSTEM MODEL AND PROBLEM STATEMENT

In this section, we firstly describe the overview of the proposed system architecture. The proposed solution is an end-to-end scheduling and sharing framework and the implementation requires the modifications on the sender and receiver sides.

Considering n clients composing a group $I = \{0, 1, \ldots, n\}$ cooperated with each other, as shown in Fig. 2. The key component on the server side is used to split the data stream and distribute the packets onto different paths. The Feedback Controller collects the information from the clients such as the status of the communication paths and outputs the feedback to the data stream splitter. The Data Stream Splitter takes the parameters to split video data into small pieces. Each piece contains several entire Group of Picture (GoP) and will be assigned into different paths. Then, the data packets in the sending buffer will be delivered to the clients via cellular links. And different devices will receive different pieces, respectively.

The client side can be regard as two parts since the device takes the responsibility to handle cellular downloading network and D2D sharing network concurrently. For each device i in the group, the data is downloaded from the cellular path and then share through the short-range D2D networks (e.g., WiFi). After receiving from the cellular networks, the packets will be re-ordered and out-of-date packets will be dropped due to the delay constraint. To reduce wireless collision and transmission overhead, broadcast scheme is established on D2D network and successful packets will be shared to other group mates attached with the limit of delay.

TABLE 1. Basic notations.

I, n	the group $\{0, 1, \ldots, n\}$ and group capacity.
D	the consumption of time.
b, o	the basic and optional parts of the data
$\chi(t)$	the state of Markov chain, G (Good) or B (Bad).
ξ_G, ξ_B	the transition probability from B to G and G to B.
$P(\cdot), E[\cdot]$	the probability value, expectation value.
μ, υ	the bandwidth of the cellular and WiFi paths.
η, φ	the average packet loss rate of the cellular and WiFi paths.
ψ, ψ'	the effective loss rate of the cellular and WiFi paths.
λ, λ'	the arriving rate on the cellular and WiFi paths.
RTT_C, RTT_W	the round trip time of the cellular and WiFi paths.
d, d'	the assigned data rate on the cellular and WiFi paths.
T_C, T_W	the delay constraint on the cellular and WiFi paths.
Φ_C, Φ_W	the goodput on the cellular and WiFi paths.
$\Phi, \overline{\Phi}$	the summary and average of the group goodput.
$R, \Delta R$	the assigned rate of sub-flows and the iterative increment value.
$F(\cdot), \nabla F(\cdot)$	the utility function and the total differential equation.
γ	the learning rate of the gradient descent method.

The problem of traffic distribution with the models of transmission path and goodput will be described in the rest of this section. The basic notations used throughout this paper are listed in Table 1.

A. PATH PROPERTIES

The end-to-end connection can be described by a tuple including a pair of IP addresses from the source and destination nodes, respectively. It is well-known that wireless access link is most likely to be the bottleneck for end-to-end transmission due to the limited bandwidth and time-varying channel status. Each communication path is considered to be an independent transport link uncorrelated with others [43] and is characterized by the following properties.

- The available bandwidth μ. This property does not represent the original capacity of the path, but the available time-varying rate utilized by the end-to-end flow.
- The round trip time RTT, which indicates the delay between a packet is sent and the acknowledge of the packet is received. This value consists of packet delivery time, processing latency, and path propagation delay.
- The average packet loss rate η, assumed to be an independent and identically distributed (i.i.d.) process, which is uncorrelated with the input traffic rate.

Similar to the previous works [15], the data traffic load we push on a path does not affect its loss statistics as the background traffic is recognized much larger than our own.

B. EFFECTIVE LOSS RATE

In the context of heterogeneous overlay networks, the packet losses can be classified into three categories [41]:

- The losses caused by congestions due to the bandwidth limitation or buffer overflow;
- The errors caused by noise or interference in the wireless networks;
- The path failure loss or handover loss.

In wireless networks, most packet losses are due to the wireless channel fluctuations or path failure and not caused by the link congestion [43]. And we model burst loss on each path using the Gilbert loss model [14], which can be expressed as a two-state stationary continuous time Markov chain. The state $\chi(t)$ at time t assumes one of two values: G (Good) or B (Bad). If a packet is sent at time t and $\chi(t) = G$ and then the packet can be successful delivered; if $\chi(t) = B$ then the packet is lost. Let ξ_G and ξ_B represent the transition probability from B to G and G to B, respectively. Thus, the average burst loss length is $\frac{1}{\xi_B}$ and we can have:

$$\eta = \frac{\xi_B}{\xi_B + \xi_G} \tag{1}$$

Considering a stable system, the assigned data rate d should not exceed the bandwidth μ of the path. And the receiving buffer at the client side is used to maintain the fluctuating delay of the arriving packet due to the time-varying latency of the path. Thus, the packet loss caused by the transmission error can be summarized as $d \cdot \eta$.

For IP-based protocols, a packet is the smallest unit of data that transmits over packet switching networks. However, to increase in a network with a high degree of parallelism, the scheduled packets may arrive at the client out-of-order. Thus, the early-arrival packets have to wait for late packets in the receiving buffer. For real-time applications, the received data is also request by a time deadline T, such as the live video streaming services. If late packets arrive within T, the transmission is successful. Otherwise, the waiting time causes additional packet delay and the overdue packets are lost.

Thus, without loss of generality η, a packet may be dropped at the client since it is out of date. And the effective loss rate ψ includes both transmission and overdue loss.

C. CELLULAR REAL-TIME TRAFFIC GOODPUT

Focused on the cellular network, clients in the group downloading data from the server independently along with the time deadline T_C. In a capacity-limited transport link, the overdue loss rate can be approximated by the $M/G/1$ queueing model [33]. Hence, the packet delay over a single cellular link follows an exponential distribution [50] and can be modeled as

$$P(D > T_C) = e^{-\lambda \cdot T_C} \tag{2}$$

where λ represents the arriving rate and is determined by the average delay

$$\lambda = \frac{1}{E[D]} \tag{3}$$

where $E[\cdot]$ represents the expectation value. The arriving rate λ can be calculated from the end-to-end delay statistics, nevertheless, a great number of the sample is needed. In order to perform efficient online operation, we employ a fractional function to approximate the delay combined with assigned data transmission and network congestion.

$$\lambda = \frac{1}{\frac{d}{\mu} + \frac{RTT_C^i}{2}} = \frac{2 \cdot \mu}{2 \cdot d + \mu \cdot RTT_C^i} \tag{4}$$

In conjunction with η, which is the average packet loss rate of the cellular path, the effective loss rate ψ can be interpreted as

$$\psi = \eta + (1 - \eta) \cdot P(D > T_C) \tag{5}$$

and the received packets expected by the client i (i.e. goodput) can be estimated with

$$\begin{aligned} \Phi_C^i &= d_i \cdot (1 - \psi_i) \\ &= d_i \cdot (1 - \eta_i) \cdot (1 - e^{-\lambda_i \cdot T_C}) \\ &= d_i \cdot (1 - \eta_i) \cdot (1 - e^{-\frac{2 \cdot \mu_i \cdot T_C}{2 \cdot d_i + \mu_i \cdot RTT_C^i}}) \end{aligned} \tag{6}$$

D. D2D REAL-TIME TRAFFIC GOODPUT

On the D2D network, we establish the data sharing operation based on WiFi technology. However, all data will relay through the access point (AP) in infrastructure mode, which will cause more capacity of the bandwidth to solve the collision. In order to reduce the relay cost, we enable the broadcast scheme to share the data received via cellular path. Since the clients share the same D2D path, the average packet loss rate φ and round trip time RTT_W of WiFi is treated as same. Similar with cellular path, the arriving rate λ' according to the bandwidth υ

$$\lambda' = \frac{1}{\frac{d'}{\upsilon} + \frac{RTT_W}{2}} = \frac{2 \cdot \upsilon}{2 \cdot d' + \upsilon \cdot RTT_W} \tag{7}$$

Then, we can have the effective loss rate ψ' of device i from j according to the average packet loss rate φ as:

$$\psi'_{i,j} = \varphi_j + (1-\varphi_j)\cdot e^{-\lambda'_j \cdot T_W} \tag{8}$$

To establish efficient online operation, we approximately regard the average packet loss rate as same, i.e. $\varphi_i = \varphi_j (i \neq j)$. And the goodput can be modeled as

$$\Phi_W^{i,j} = d'_j \cdot (1-\varphi)\cdot (1-e^{-\lambda'_j \cdot T_W}) \tag{9}$$

And the aggregate goodput from the group is estimated as follows

$$\begin{aligned}\Phi_W^i &= (1-\varphi)\cdot \sum_{m\neq i}^{m\in I}\{d'_m \cdot (1-e^{-\lambda'_m \cdot T_W})\}\\ &= (1-\varphi)\cdot \sum_{m\neq i}^{m\in I}\{d'_m \cdot (1-e^{-\frac{2\cdot\upsilon\cdot T_W}{2\cdot d'_m+\upsilon\cdot RTT_W}})\}\end{aligned} \tag{10}$$

E. OPTIMIZATION FORMULATION

As mentioned above in Eq. 6 and Eq. 10, the all goodput of the group can be summarized as:

$$\Phi = \sum_{i\in I}(\Phi_C^i + \Phi_W^i) \tag{11}$$

and the average goodput is:

$$\begin{aligned}\overline{\Phi} &= \frac{1}{n}\cdot\sum_{i\in I}(\Phi_C^i + \Phi_W^i)\\ &= \frac{1}{n}\cdot\sum_{i\in I}\{d_i\cdot(1-\eta_i)\cdot(1-e^{-\lambda_i\cdot T_C})\}\\ &\quad + \frac{(n-1)\cdot(1-\varphi)}{n}\cdot\sum_{i\in I}\{d'_i\cdot(1-e^{-\lambda'_i\cdot T_W})\}\end{aligned} \tag{12}$$

It is obvious that the scheduling algorithm will only transmit packets within the time constraint based on the feedback collecting from the clients and the assigned data rate will always under the available bandwidth. Furthermore, as is already proved in existing studies [25], [29] that the optimal flow allocation to minimize end-to-end delay is to eliminate the delay differences of available communication paths. Now, we are ready to formulate the following linear constraint optimization problem for given transmission paths of cellular and WiFi on maximizing the average goodput of the group while satisfying the path capacity, delay constraint, etc.

For each data distribution interval,
determine the values of d, d'
to maximize $:\overline{\Phi} = \frac{1}{n}\cdot\sum_{i\in I}(\Phi_C^i + \Phi_W^i)$

$$s.t. \begin{cases}\Phi_C^i = \text{Equation}(6),\\ \Phi_W^i = \text{Equation}(10),\\ d_i \leq \mu_i (i\in I),\\ \sum_{i\in I} d'_i \leq \upsilon,\\ \dfrac{2\cdot\mu_i\cdot T_C}{2\cdot d_i + \mu_i\cdot RTT_C^i} > 1 (i\in I),\\ \dfrac{2\cdot\upsilon\cdot T_W}{2\cdot d'_i + \upsilon\cdot RTT_W} > 1 (i\in I),\\ \lambda_i = \lambda_j (i,j\in I, i\neq j).\end{cases} \tag{13}$$

The distribution interval is correlated with the imposed deadline (e.g., 250 milliseconds, which is the duration of a GoP). According to the recent works [16], this problem is a kind of cost-minimization flow-splitting problem over multiple paths (the effective loss rate can be regarded as the cost). The multi-path data rate scheduling problem can be converted to the precedence constrained multiple knapsack problem and such problems prove to be NP-hard [27]. Therefore, it is computationally prohibitive to obtain the optimal solution. In this paper, we propose the heuristic algorithms to approximate the above goal function and obtain the close-to-optimal result with fast convergence for efficient online operation.

IV. PROPOSED MECHANISM IMPLEMENTATION

In this section, the proposed mechanism can be divided into two parts. The server side firstly takes responsibility of data repartition and delay distribution. As a result, the data flow is allocated into different paths to achieve the goal function mentioned above. Along with the packets transmission, the per-path status is collected by the client to improve the performance of the data splitting algorithm. Passing the delay filter, the client shares the seasonable data using the broadcast scheme vie D2D network.

A. DATA REPARTITION AND DELAY CONSTRAINT

In this paper, considering the sharing through the heterogeneous network, the video data rate can be simply regarded as two sub-flows, b (Basic) and o (Optional), i.e. $b + o = 1$. The basic part for each device is same and only the optional part will be shared. In every distribution interval, each client will download the entire basic sub-flow and partial optional sub-flow of the streaming data. Thus, the average data rate on the cellular paths can be summarized as:

$$b + \frac{o}{n} = \frac{1+(n-1)\cdot b}{n} \tag{14}$$

If the ratio of the basic part is close to 1, almost data is transferred via cellular network and only few chunks is pushed onto the D2D sharing network. In contrary, different devices will get different pieces respectively and all clients are tend to achieve entire video streaming based on sharing. Thus, to utilize the network resource of both cellular and D2D paths, the repartition of the data should be carefully designed.

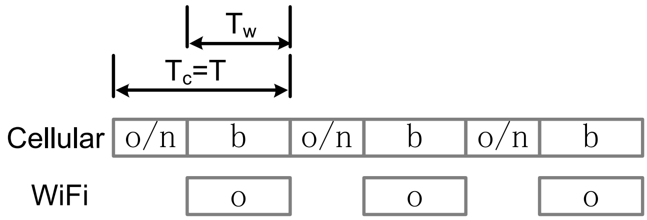


FIGURE 3. Concurrent Strategy of Data Splitting and Constraint Distribution.

Since the received data is request by a time deadline T for real-time applications, we proposed a concurrent strategy to determine the delay distribution on cellular T_C and D2D T_W as shown in Fig. 3. The concurrent strategy lets the devices download the basic and optional parts at the same time. The client firstly obtains the optional part rather than the basic one. When the basic part is downloading, the received optional data will be shared. According to the bandwidth of both cellular and WiFi paths, we can have:

$$b/(n \cdot \frac{o}{n}) = \frac{E[\mu]}{\upsilon} \tag{15}$$

Since the deadline on cellular path is equal to the total constraint, $T_C = T$:

$$b/\frac{1+(n-1)\cdot b}{n} = \frac{b \cdot n}{1+(n-1)\cdot b} = \frac{T_W}{T} \tag{16}$$

Thus,

$$\begin{cases} b = \dfrac{E[\mu]}{E[\mu]+\upsilon}, \\ o = \dfrac{\upsilon}{E[\mu]+\upsilon}, \\ T_C = T, \\ T_W = \dfrac{b \cdot n \cdot T}{1+(n-1)\cdot b} = \dfrac{n \cdot E[\mu] \cdot T}{n \cdot E[\mu]+\upsilon} \end{cases} \tag{17}$$

B. COMPONENT ON SERVER

The Data Stream Splitter in Fig. 2 is the key component on server. This module is responsible for splitting the video data rate into different sub-flows based on the feedback collecting from the client side.

According to Eq. 13, we have a data rate constraint that:

$$\frac{2 \cdot \mu_i \cdot T_C}{2 \cdot d_i + \mu_i \cdot RTT_C^i} > 1 (i \in I) \tag{18}$$

$$d_i < \mu_i \cdot \frac{2 \cdot T_C - RTT_C^i}{2} \quad (d_i \leq \mu_i, i \in I) \tag{19}$$

However, considering Eq. 6, it is a challenging issue to determine the appropriate data rate among several sub-flows to maximizing the goodput of the group. Fig. 4 shows the goodput Φ_C of one path when $\eta = 0.01$, $\mu = 6Mbps$, $T_C = 0.25s$, $RTT_C = 0.02s$. From Fig. 4, we can observed that the increase is rapid for slight rate assignment but becomes very slow in the heavy one. To obtain the close-to-optimal result $R = \{d_0, d_1, \ldots, d_n\}$ with fast convergence for efficient online operation, we proposed a heuristic algorithm based on the gradient descent method [5].

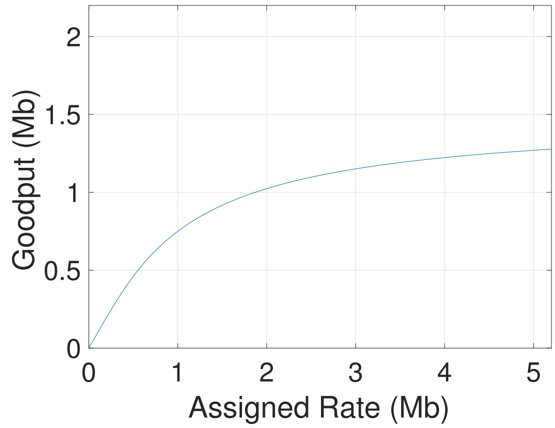


FIGURE 4. Goodput with data rate assignment.

According to Eq. 6, we develop a utility function based on the good put of cellular path Φ_C:

$$\begin{aligned} F(R) &= \left\{\Phi_C^0, \Phi_C^1, \ldots, \Phi_C^n\right\} \\ &= \left\{d_i \cdot (1-\eta_i) \cdot (1-e^{-\frac{2 \cdot \mu_i \cdot T_C}{2 \cdot d_i + \mu_i \cdot RTT_C^i}})\right\} \end{aligned} \tag{20}$$

and we can have the total differential equation that:

$$\begin{aligned} \nabla F(R) &= \left\{\frac{\partial \Phi_C^0}{\partial d_0}, \frac{\partial \Phi_C^1}{\partial d_1}, \ldots, \frac{\partial \Phi_C^n}{\partial d_n}\right\} \\ &= \Big\{(1-\eta_i) \cdot [1-e^{-\frac{2 \cdot \mu_i \cdot T_C}{2 \cdot d_i + \mu_i \cdot RTT_C^i}} \\ &\quad \cdot (1+\frac{4 \cdot d_i \cdot \mu_i \cdot T_C}{(2 \cdot d_i + \mu_i \cdot RTT_C^i)^2})]\Big\} \end{aligned} \tag{21}$$

With initial value setting to the basic part as $\Delta R_0 = \{b\}$, we can iteratively approach the expected data rate assignment where γ represents the learning rate of the gradient descent method.

$$\Delta R_{k+1} = \Delta R_k + \gamma \cdot \nabla F(\Delta R_k) \tag{22}$$

The iteration is regarded as convergence when the increment $\gamma \cdot \nabla F(\Delta R_k)$ is less than a threshold limit value (TLV) [48]. The converged cellular path will be filtered and the iteration will terminate until no more source data or path to be assigned (the TLV is small enough). Fig. 5 shows the gradient descent process of two cellular paths when splitting a video streaming whose bitrate is $3Mbps$, with parameters $\eta = \{0.01, 0.02\}$, $\mu = \{6Mbps, 8Mbps\}$, $T_C = 0.25s$, $RTT_C = \{0.02s, 0.025s\}$.

The proposed data rate assignment algorithm is shown in Algorithm 1.

C. COMPONENT ON CLIENT

To effectively utilize the available bandwidth resources in heterogeneous overlay networks, it is significant to accurately

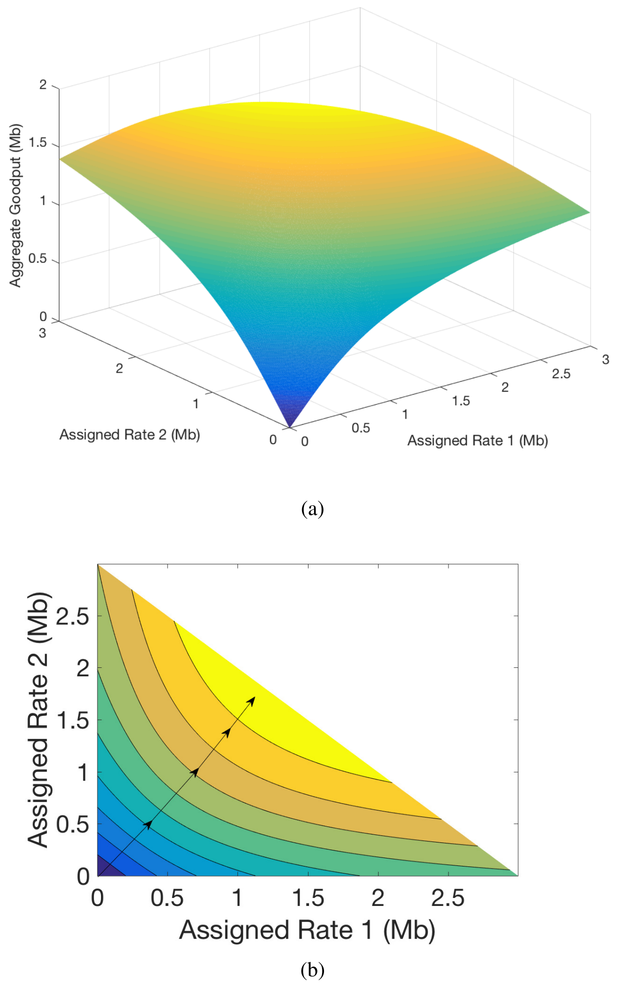


FIGURE 5. **Gradient descent method of two cellular paths. (a) Summation of the utility function. (b) Iterative descent process.**

Algorithm 1 Data Rate Splitting Assignment

Input: Data Rate d, Initial Value $\Delta R_0 = \{b\}$, Deadline T_C, Path Status $\eta = \{\eta_i\}$, $\mu = \{\mu_i\}$, $RTT_C = \{RTT_C^i\}$, TLV, Learning Rate γ

Output: Expected Data Rate Assignment $R = \{d_0, d_1, \ldots, d_n\}$

$k = 0$;
$\nabla F(R)$ =Equation(21);
repeat
  $sum = \gamma \cdot \sum \nabla F(\Delta R_k)$;
  for $i \in I$ **do**
    if $\gamma \cdot \nabla F(\Delta R_k^i) < TLV$ *or* $\gamma \cdot \nabla F(\Delta R_k^i) > \mu_i \cdot \frac{2 \cdot T_C - RTT_C^i}{2} - d_i$ **then**
      knockout path i from iteration and lock the data rate d_i;
      $sum = sum - \gamma \cdot \nabla F(\Delta R_k^i)$
    end
  end
  if $sum < d$ **then**
    $\Delta R_{k+1} = \Delta R_k + \gamma \cdot \nabla F(\Delta R_k)$;
    $d = d - sum$;
  else
    $\Delta R_{k+1} = \Delta R_k + \frac{d}{sum} \cdot \gamma \cdot \nabla F(\Delta R_k)$;
    $d = 0$;
  end
  k++;
  Update R;
until $d = 0$ *or* $\Delta R_k = \emptyset$;

estimate the status of each communication channel. In this paper, the path estimating algorithm dubbed pathChirp [31] is used which modify the methodology of Self-Loading Periodic Streams (SLoPS) [20], using different probing packet stream patterns. The main objective in pathChirp is to achieve accuracy with shorter measurement latency [30]. The probe packets travel one-way from clients to server, and the bandwidth estimation is processed at the server side. And no echo information is sent back to the clients to avoid the problem of echo probe traffic interfering. Since the pathChirp generally performs better with larger packet sizes, we thus use 1000-byte packets containing the timestamp, packet loss rate η and RTT to assist the Feedback Controller.

After checking in the Delay Filter, the timely data will be shared through the D2D network. If a video frame experiences transmission or overdue loss, it is considered to be lost and will be concealed by copying from the last received frame before it. To perform data sharing properly and thus avoid potential problems, smartphones that are within the sharing group are guided to use a specific method called Token Ring Method (TRM). To best of our knowledge, without implementing TRM, there are serious packet loss problems when smartphones try sharing data with each other using the local sharing network. Packets collide with each other when more than one smartphones try to broadcast data using the local sharing network at a same specific time period. Packets will be lost when collisions happen, and it is well-known that packet collisions are the dominant reason for packet losses in the system proposed in this thesis. Without TRM, smartphones just try broadcasting data whenever there are data to share without considering whether there are or there will be other smartphones try broadcasting data using the local sharing network as well. To make the local sharing system efficient and reliable, packet losses that due to packet collisions need to be reduced.

The TRM is used for avoiding packet losses due to packet collisions. TRM provides a Round-Robin (RR) way of data sharing and thus avoid packet collisions due to multiple smartphones broadcasting data at the same time. To realize TRM, besides the WiFi based broadcasting network, smartphones shall establish minimum number of TCP connections with each other and form a ring topology.

Examples of these ring topologies are shown as Fig. 6: Fig. 6(a) shows a ring topology of a three-device group. These

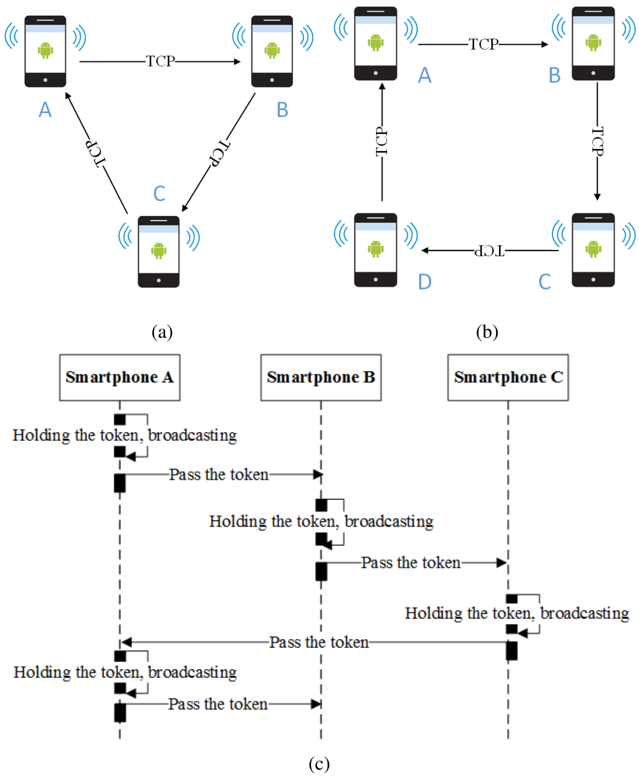


FIGURE 6. **Token ring topology. (a) 3 devices. (b) 4 devices. (c) Token ring with round-robin.**

three devices use only 3 TCP connections and form a ring. For each smartphone, there are only two connections that are used for data exchanging with two different smartphones. Fig. 6(b) gives an example of a ring topology of a four-device group. Like in Fig. 6(a), these four smartphones also use the minimum number of TCP connections to form a ring topology. Each smartphone can transmit/relay data to any other devices using these connections. An alternative way of establishing TCP connections is to create connections between any two devices directly within the sharing group. The reason why we don't do so is that such type of solutions requires more connections to be established. Creating and maintaining extra connections requires extra resources of smartphones and will only affect the battery life, but also degrades the system performance.

In the proposed TRM, the TCP ring topology is used for transmitting a specific token message. There is only one token in the topology, and only the smartphone that holds the token has the right to broadcast data using the local sharing network. An instance of token passing is shown as Fig. 6(c). At the beginning, smartphone A has the token and performs data sharing. Like in Eq. 19, data to be shared is constraint by Eq. 23.

$$d_i' < \upsilon \cdot \frac{2 \cdot T_W - RTT_W}{2} \; (d_i' \leq \upsilon, i \in I) \tag{23}$$

After finishing sharing by implementing broadcasting using the local sharing network, smartphone A will pass the token to smartphone B. And then smartphone B possesses the token, which means that it has the right to perform data sharing under constraint using the local broadcasting network. After finishing data sharing, the token will be passed to smartphone C via another connection, and smartphone C will perform data sharing and then pass the token to smartphone A. The token will be passed on as described above in Fig. 6, and smartphones broadcasts data one by one. By implementing TRM, smartphones broadcasts data in a RR way, and thus avoid packet collisions due to multiple smartphones try performing data sharing at the same time period in the local sharing network.

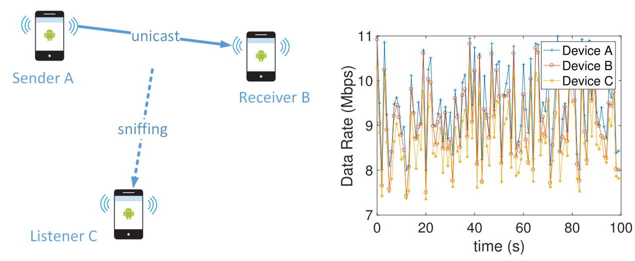
V. PERFORMANCE EVALUATIONS

A. EXPERIMENTAL SETUPS

H.264/AVC reference software JM 19.0 is used as the video encoder. The generated video streaming is encoded at 60 frames per second and a GoP consists of 15 frames with $IP \ldots P$ structure. The test video sequences [4] are *Akiyo*, *Coastguard*, *Container* and *Mobile* in Common Interchange Format (CIF) with 300 frames. These test sequences feature different patterns of temporal motion and spatial characteristics that reflected in their corresponding video quality versus encoding rates. We concatenate the video sequences 20 times to be 6000 frame-long in order to obtain statistically meaningful results. The video encoding rate is set to be 3500, 4500, 5500, and 6500 Kbps in different trails. The delay constraint T for GoP is set to be equal to the play-out duration, i.e., 250 ms.

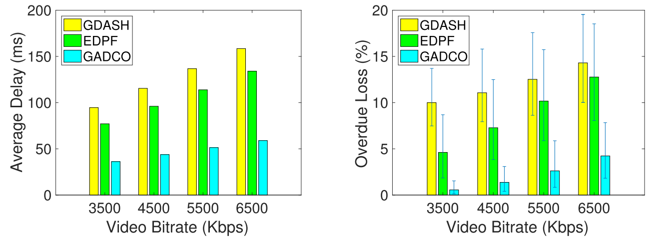
In this paper, Nexus 5 (hammerhead) on the Android platform is used in the experiments. All smartphones have a 2.26 GHz quad-core CPU and 2 GB RAM and use Cyanogen Mod 11 (4.4.4 KitKat) as their operating systems and the kernel version of the linux is 3.4.0. The operator of the cellular path using LTE is China Unicom. We placed them in an indoor environment and fixed the position to eliminate any possible bias due to the mobility. The server is deployed on the Elastic Compute Service (ECS) on Aliyun [2] equipped with an eight-core 2.6 GHz CPU and 8GB RAM. The operating system is Centos 6.5 and the version of the linux version is 2.6.32.

Although devices within proximity of each other can overhear all wireless transmissions [23], the existing modes do not support the high-rate broadcast in particular. The unicast mode of 802.11 does not exploit broadcast and transmits the same packets to each receiver separately. When a device transmits a packet to another device, the packet has to be relayed by the AP, which results in double amount of traffic. The broadcast mode of 802.11 limits [35] the transmission rate to the base rate. Thus, in this paper, the Pseudo-Broadcast (PB) scheme [24] is used which combines the benefits of unicast and broadcast. Unicast is used as the transmission mode, but the devices sniff all transmissions in their neighborhood as shown in Fig. 7(a). Within sniffing by device C, the capture data rate of device A unicast to



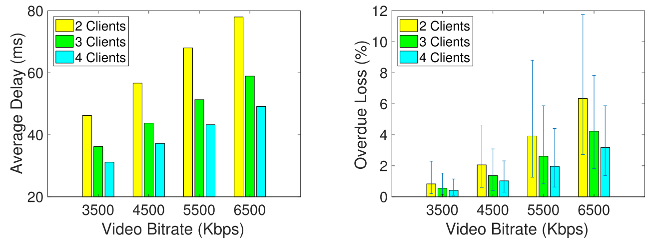
(a) (b)

FIGURE 7. Pseudo broadcast. (a) Network topology. (b) Data rate of devices.



(a) (b)

FIGURE 8. End-to-end delay and overdue loss compared with reference schemes. (a) End-to-end delay. (b) Overdue loss.



(a) (b)

FIGURE 9. End-to-end delay and overdue loss compared with different group size. (a) End-to-end delay. (b) Overdue loss.

B can be discovered in Fig. 7(b). Tcpdump [28] is used to trace the packets from the Network Interface Card (NIC) driver module. To reduce the double transmission, the AP device will not forward packets since the other devices should already have received via overhearing [9]. In this manner, we are able to enable pseudo-broadcast, while ensuring only one transmission per packet. In order to maximize the sharing efficiency, we modified the *cfg80211* module of the *bcmdhd* driver with *__wl_band_2ghz* and avoid the influence by the rate control algorithm in *mac80211* (e.g., Minstrel [6]).

B. REFERENCE SCHEMES AND PERFORMANCE METRICS

In this paper, the reference schemes are listed below:

- GDASH [48]. The proposed green and cooperative scheme meets the same scenario with this paper. A group of users cooperating with each other to experience video service but the mechanism is based on the DASH protocol. Since reliable connection is used, unabridged content is obtained but no delay constraint is taken into consider. The segment length is set to the duration of the GoP (250ms) and bitrate selection algorithm is disabled.
- EDPF [10]. The Earliest Delivery Path First scheme is proposed to ensure packets meet their playback deadlines by scheduling packets based on the estimated delivery time of the packets. The proposed algorithm is designed for only one individual user to watch the video streaming while considering the delay constraint. In order to approach the network condition of the group-based schemes, EDPF is deployed on 3 devices respectively at the same time.

For performance comparison with regard to the fairness and representativeness, performance metrics are chose as:

- End-to-End Delay. The end-to-end video frame delay are measured every data distribution interval to reflect the delay performance of the competing schemes.
- Goodput. Goodput is an application-level throughput, i.e., the number of useful information bits successfully received by the destination within the imposed deadline.
- PSNR. PSNR is the standard metric of objective video quality and is a function of the mean square error between the original and the received video frames.

To obtain the confidence results, we repeat each set of experiments more than 10 runs and present the averaged results. For the microscopic results and time series analysis, we present the data of a single run with finer granularity.

C. END-TO-END DELAY

The result of the average end-to-end delay when 3 devices are watching at the same time is shown in Fig. 8(a). It can be discovered that the proposed GADCO mechanism achieves significantly lower delay than the reference schemes. GADCO reduces the average end-to-end delay up to 99.5 and 75.9 ms when watching 6500 Kbps video compared to GDASH and EDPF, respectively. The end-to-end delay should be treated carefully in the real-time multimedia applications, since the delay performance gains can significantly prevent playback buffer starvation, which is critical for the user perceived media quality. And it should also be noticed that more devices cooperating with each other in Fig. 9(a) bring lower delay, which reduce video stalls/glitches during the play-back process and result in graceful user experience.

In real-time applications, larger end-to-end packet delay always incurs more overdue packets, which in turn results in lower goodput. Although some packets can be successfully delivered to the clients, out-of-date packets are dropped by the upper applications and regarded as loss indeed. For GDASH mechanism, we use the same method to calculate the overdue loss rate, but this constraint is not required by the DASH applications in particular. DASH based protocols will first initialize several chunks of the video data until the player is start. The play-out progress of GDASH is later than other schemes. Thus, the probability of jitter can be absorbed by the buffer. The overdue loss rates of the competing models are depicted in Fig. 8(b). The proposed GADCO

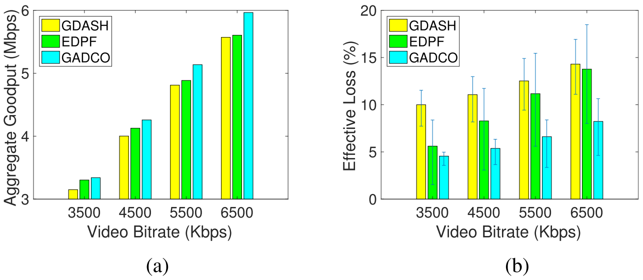


FIGURE 10. **Aggregate goodput and effective loss compared with reference schemes. (a) Aggregate goodput. (b) Effective loss.**

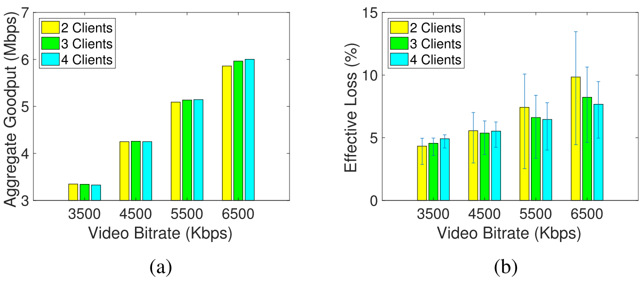


FIGURE 11. **Aggregate goodput and effective loss compared with different group size. (a) Aggregate goodput. (b) Effective loss.**

achieves competitive overdue loss rate with a smaller fluctuation range. The GDASH also fluctuates more slightly than the EDPF because the group-based cooperation has stronger tolerance to meet the dramatic changes of the network condition. Fig. 9(b) shows that a larger group is more elastic and moderate.

D. GOODPUT

It is well-know that goodput differs from throughput with regard to the imposed deadline. Although all received video data is handed over to GDASH application layer based on the reliable connections, the goodput can be count at the same way as the overdue loss rate. Fig. 10(a) plots the average goodput obtained by different models in different video bitrate and it can be observed that GADCO achieves higher goodput. The pattern is nearly opposite to that in Fig. 8(a) since larger end-to-end delay will introduce more overdue packets and results in lower goodput. GADCO increases the average goodput up to 395 and 360 Kbps when watching 6500 Kbps video compared to GDASH and EDPF, respectively, as the data splitting distribution algorithm is able to effectively leverage the path diversity to optimize the aggregate goodput. The instantaneous goodput values for video streaming compared with the reference schemes and different group size are shown in Fig. 12 and Fig. 13 to have a microscopic view.

As introduced in Section III-C, the effective loss rate includes both the transmission and overdue losses over the available paths. With regard to the delay constraint, the overdue packets arrived at the destination are considered to be lost. Fig. 10(b) shows that GADCO significantly reduces the

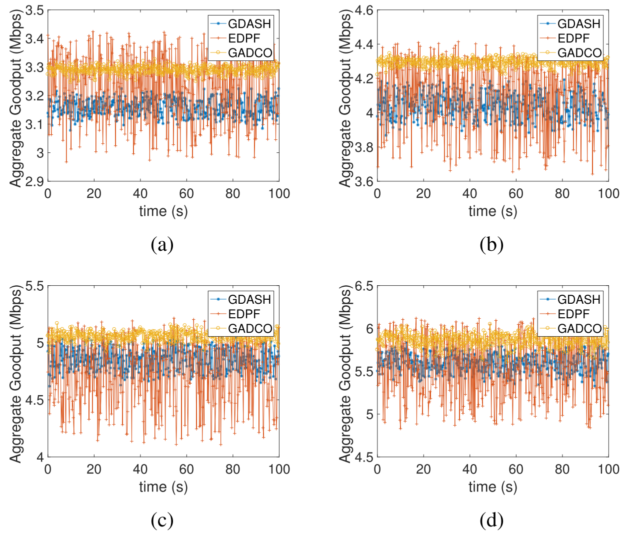


FIGURE 12. **Aggregate goodput of different video bitrate compared with reference schemes. (a) 3500Kbps. (b) 4500Kbps. (c) 5500Kbps. (d) 6500Kbps.**

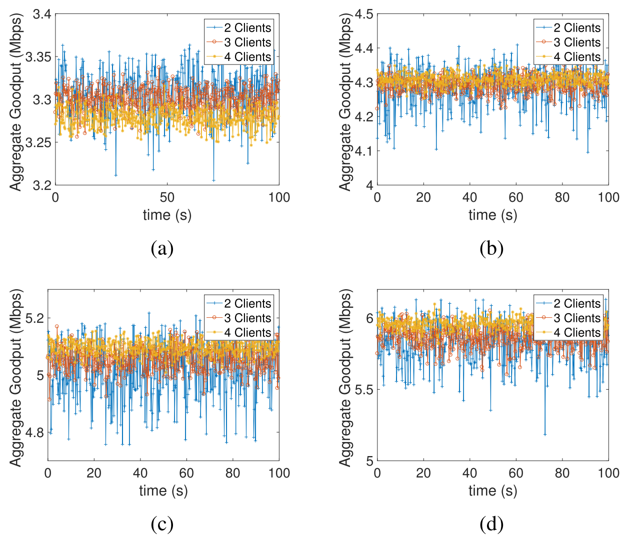


FIGURE 13. **Aggregate goodput of different video bitrate compared with different group size. (a) 3500Kbps. (b) 4500Kbps. (c) 5500Kbps. (d) 6500Kbps.**

overall packet losses with lower variations than the reference schemes as it reduces the end-to-end delay as well as mitigates the consecutive losses. It can be also discovered that the group-based schemes has a stronger elasticity facing the fluctuation of the network condition especially with a heavy transmission task.

The other interesting fact we notice is that bigger group performs worse with low quality video (3500 Kbps) as shown in Fig. 11(b). It is well-known that cellular networks exhibit better performance in sustaining user mobility than WiFi but provide a lower peak data rate [36], [37]. The slight task introduces a inconspicuous overdue loss with a non-negligible WiFi transmission error. And the pressure of the overdue loss

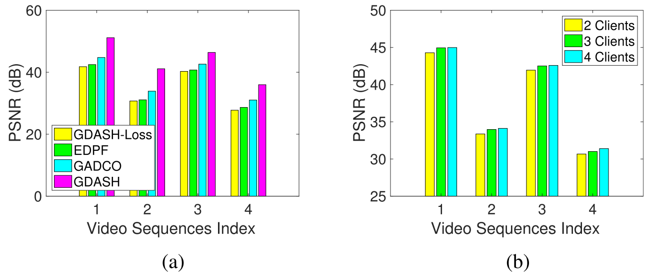


FIGURE 14. **PSNR of 5500 Kbps Video (1 *Akiyo* 2 *Coastguard* 3 *Container* 4 *Mobile*). (a) Compared of schemes. (b) Compared of group sizes.**

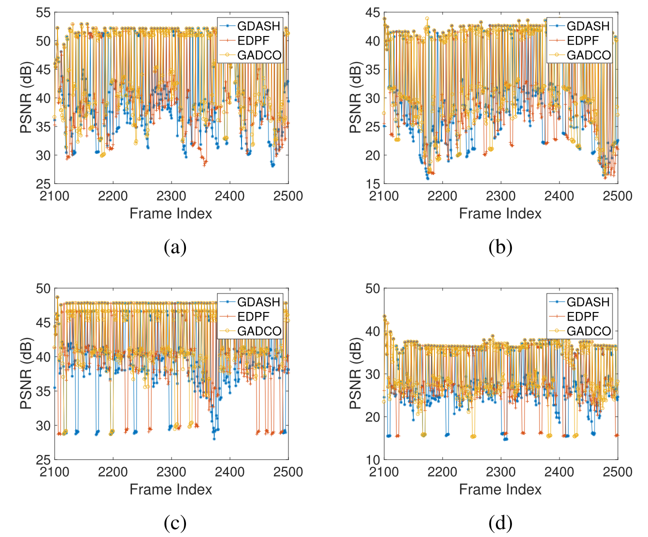


FIGURE 15. **PSNR of 5500 Kbps video compared of reference schemes. (a) Akiyo. (b) Coastguard. (c) Container. (d) Mobile.**

when watching a high quality video can be amortized by a bigger group and the smaller group will achieve a higher effective loss.

E. VIDEO PSNR

In this paper, PSNR is used to expose the video quality. We plot the average PSNR compared of the reference schemes in Fig. 14(a) and 2 versions of the GDASH is provided. Based on the reliable links, the original GDASH hand over the complete video content and the PSNR value can be regard as the upper bound limited by the H.264/AVC file itself. And the loss version of the GDASH is calculated by the same principles which follow the same pattern in Fig. 10(a). It can be found that the GADCO improves the average PSNR value up to 3.1 and 2.8 dB when watching the *Coastguard* test sequence with 5500 Kbps bitrate compared to GDASH and EDPF, respectively. GADCO achieves higher PSNR values, under various video sequences with different patterns of temporal motion and spatial characteristics, than the reference schemes due to its excellent performance in goodput. Higher goodput performance generally leads to lower content distortion of streaming video, which guarantees more video frames deliver within the imposed deadline and thus

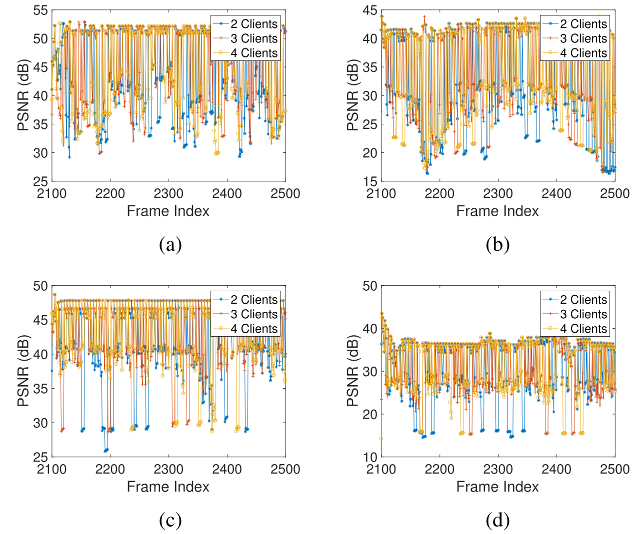


FIGURE 16. **PSNR of 5500 Kbps video compared of different group size. (a) Akiyo. (b) Coastguard. (c) Container. (d) Mobile.**

results in the effective gains in PSNR. Generally, the PSNR values achieved by the competing groups upgrade with the increase in group size because the aggregated bandwidth is sufficient compared to the input traffic rate which is plotted in Fig. 14(b). The per-frame PSNR values indexed from 2100 to 2500 for video streaming with 5500 Kbps bitrate compared of the reference schemes (GDASH with loss) and different group size are shown in Fig. 15 and Fig. 16 to have a microscopic view.

VI. CONCLUSIONS AND FUTURE WORKS

In this paper, we present GADCO, a group-aware delay-constraint mechanism to enhance real-time video streaming experience of a cluster of multihomed mobile users. The proposed scheme is different from the existing transmission approaches as it uses a D2D sharing structure to aggregate the bandwidth resources with different radio interfaces in order to download and share simultaneously. Through modelling and analysis, we develop solutions for per-path status estimation, video data rate splitting and assignment while satisfying the fluctuation of the network condition and requirement raised by the video application. A prototype adopting D2D broadcast scheme is implemented on Android platform involving real-time video encoded with H.264 codec. The experimental results demonstrate appreciable improvements in terms of increasing video PSNR, goodput and reducing packet delay of the entire group.

As the future work, we will consider the following directions: 1) employing the video rate adaption for dynamically adjusting the transmission rate, and 2) study the video distortion model to design quality-based flow rate splitting mechanism, and 3) applying energy-minimized protocol to improve performance in group sharing.

REFERENCES

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

[2] Aliyun. (2016). *Elastic Compute Service*. [Online]. Available: http://www.aliyun.com

[3] T. Alpcan, J. P. Singh, and T. Başar, "Robust rate control for heterogeneous network access in multihomed environments," *IEEE Trans. Mobile Comput.*, vol. 8, no. 1, pp. 41–51, Jan. 2009.

[4] Arizona State University. (2016). *YUV Video Sequences*. [Online]. Available: http://trace.eas.asu.edu/yuv/index.html

[5] M. Avriel, *Nonlinear Programming: Analysis and Methods*. Honolulu, HI, USA: Courier Corporation, 2003.

[6] Johannes Berg, Linux Wireless (wiki content). (2016). *Minstrel*. [Online]. Available: https://wireless.wiki.kernel.org/en/developers/documentation/mac80211/ratecontrol/minstrel

[7] C. Boldrini, M. Conti, and A. Passarella, "Exploiting users' social relations to forward data in opportunistic networks: The HiBOp solution," *Pervasive Mobile Comput.*, vol. 4, no. 5, pp. 633–657, Oct. 2008.

[8] V. Bui, W. Zhu, A. Botta, and A. Pescape, "A Markovian approach to multipath data transfer in overlay networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 21, no. 10, pp. 1398–1411, Oct. 2010.

[9] R. Chandra *et al.*, "DirCast: A practical and efficient Wi-Fi multicast system," in *Proc. ICNP*, 2009, pp. 161–170.

[10] K. Chebrolu and R. R. Rao, "Bandwidth aggregation for real-time applications in heterogeneous wireless networks," *IEEE Trans. Mobile Comput.*, vol. 5, no. 4, pp. 388–403, Apr. 2006.

[11] S. Cheshire. (2005). *TCP Performance Problems Caused by Interaction Between Nagle's Algorithm and Delayed ACK*. [Online]. Available: http://www. stuartcheshire. org/papers/NagleDelayedAck

[12] Cisco. (2016). *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015–2020*. White Paper ID: 1454457600805266. [Online]. Available: http://goo.gl/iGOJMS

[13] N. M. Freris, C.-H. Hsu, J. P. Singh, and X. Zhu, "Distortion-aware scalable video streaming to multinetwork clients," *IEEE/ACM Trans. Netw.*, vol. 21, no. 2, pp. 469–481, Apr. 2013.

[14] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell Syst. Tech. J.*, vol. 39, no. 5, pp. 1253–1265, 1960.

[15] L. Golubchik *et al.*, "Multi-path continuous media streaming: What are the benefits?" *Perform. Eval.*, vol. 49, nos. 1–4, pp. 429–449, Sep. 2002.

[16] T. Hartman, A. Hassidim, H. Kaplan, D. Raz, and M. Segalov, "How to split a flow?" in *Proc. INFOCOM*, Mar. 2012, pp. 828–836.

[17] P. Hui, J. Crowcroft, and E. Yoneki, "BUBBLE rap: Social-based forwarding in delay-tolerant networks," *IEEE Trans. Mobile Comput.*, vol. 10, no. 11, pp. 1576–1589, Nov. 2011.

[18] S. Ioannidis, A. Chaintreau, and L. Massoulie, "Optimal and scalable distribution of content updates over a mobile social network," in *Proc. INFOCOM*, Apr. 2009, pp. 1422–1430.

[19] M. Ismail and W. Zhuang, "Mobile terminal energy management for sustainable multi-homing video transmission," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4616–4627, Aug. 2014.

[20] M. Jain and C. Dovrolis, "End-to-end available bandwidth: Measurement methodology, dynamics, and relation with TCP throughput," *IEEE/ACM Trans. Netw.*, vol. 11, no. 4, pp. 537–549, Aug. 2003.

[21] N. Jones, "Top 10 mobile technologies and capabilities for 2015 and 2016," Gartner, Stamford, CT, USA, Gartner Rep. G00260239, 2014.

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

[23] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Médard, and J. Crowcroft, "XORs in the air: Practical wireless network coding," *IEEE/ACM Trans. Netw.*, vol. 16, no. 3, pp. 497–510, Jun. 2008.

[24] A. Le, L. Keller, H. Seferoglu, B. Cici, C. Fragouli, and A. Markopoulou, "MicroCast: Cooperative video streaming using cellular and local connections," *IEEE/ACM Trans. Netw.*, vol. 24, no. 5, pp. 2983–2999, Oct. 2016.

[25] D. Ma and M. Ma, "A QoS oriented vertical handoff scheme for WiMAX/WLAN overlay networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 4, pp. 598–606, Apr. 2012.

[26] S. Mao, S. S. Panwar, and Y. T. Hou, "On optimal partitioning of realtime traffic over multiple paths," in *Proc. IEEE 24th Annu. Joint Conf. IEEE Comput. Commun. Soc.*, vol. 4. Mar. 2005, pp. 2325–2336.

[27] S. Martello and P. Toth, "Heuristische Algorithmen zur Packung von mehreren Rucksäcken," *Computing*, vol. 27, no. 2, pp. 93–112, Jun. 1981.

[28] Tcpdump and Libpcap website. (2016). [Online]. Available: http://www.tcpdump.org/

[29] S. Prabhavat, H. Nishiyama, N. Ansari, and N. Kato, "Effective delay-controlled load distribution over multipath networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 22, no. 10, pp. 1730–1741, Oct. 2011.

[30] R. Prasad, C. Dovrolis, M. Murray, and K. Claffy, "Bandwidth estimation: Metrics, measurement techniques, and tools," *IEEE Netw.*, vol. 17, no. 6, pp. 27–35, Nov./Dec. 2003.

[31] V. J. Ribeiro, R. H. Riedi, R. G. Baraniuk, J. Navratil, and L. Cottrell, "pathChirp: Efficient available bandwidth estimation for network paths," in *Proc. Passive Active Meas. Workshop*, 2003, pp. 1–11.

[32] P. Rodriguez, R. Chakravorty, J. Chesterfield, I. Pratt, and S. Banerjee, "MAR: A commuter router infrastructure for the mobile Internet," in *Proc. 2nd Int. Conf. Mobile Syst., Appl., Services*, 2004, pp. 217–230.

[33] U. K. Sarkar, S. Ramakrishnan, and D. Sarkar, "Modeling full-length video using Markov-modulated gamma-based framework," *IEEE/ACM Trans. Netw.*, vol. 11, no. 4, pp. 638–649, Aug. 2003.

[34] H. Seferoglu, L. Keller, B. Cici, A. Le, and A. Markopoulou, "Cooperative video streaming on smartphones," in *Proc. 49th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep. 2011, pp. 220–227.

[35] S. Sen, N. K. Madabhushi, and S. Banerjee, "Scalable WiFi media delivery through adaptive broadcasts," in *Proc. NSDI*, 2010, pp. 191–204.

[36] P. Si, H. Ji, and F. R. Yu, "Optimal network selection in heterogeneous wireless multimedia networks," *Wireless Netw.*, vol. 16, no. 5, pp. 1277–1288, Jul. 2010.

[37] W. Song and W. Zhuang, "Performance analysis of probabilistic multipath transmission of video streaming traffic over multi-radio wireless devices," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1554–1564, Apr. 2012.

[38] H. Soroush, P. Gilbert, N. Banerjee, M. D. Corner, B. N Levine, and L. Cox, "Spider: Improving mobile networking with concurrent Wi-Fi connections," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 4, pp. 402–403, 2011.

[39] C.-L. Tsao and R. Sivakumar, "On effectively exploiting multiple wireless interfaces in mobile hosts," in *Proc. 5th Int. Conf. Emerg. Netw. Experim. Technol.*, 2009, pp. 337–348.

[40] ITU-T VCEG. (Jun. 2015). *H.264/14496-10 AVC Reference Software*. [Online]. Available: http://iphome.hhi.de/suehring/tml/

[41] J. Wu, B. Cheng, C. Yuen, Y. Shang, and J. Chen, "Distortion-aware concurrent multipath transfer for mobile video streaming in heterogeneous wireless networks," *IEEE Trans. Mobile Comput.*, vol. 14, no. 4, pp. 688–701, Apr. 2015.

[42] J. Wu, J. Yang, X. Wu, and J. Chen, "A low latency scheduling approach for high definition video streaming over heterogeneous wireless networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2013, pp. 1723–1729.

[43] J. Wu, C. Yuen, B. Cheng, Y. Shang, and J. Chen, "Goodput-aware load distribution for real-time traffic over multipath networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 8, pp. 2286–2299, Aug. 2015.

[44] J. Wu, C. Yuen, B. Cheng, Y. Yang, M. Wang, and J. Chen, "Bandwidth-efficient multipath transport protocol for quality-guaranteed real-time video over heterogeneous wireless networks," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2477–2493, Jun. 2016.

[45] G. Xi, X. Zhang, and K. Qu, "On adaptive live streaming in mobile cloud computing environments with D2D cooperation," in *Proc. 21st Int. Conf. Telecommun. (ICT)*, May 2014, pp. 405–409.

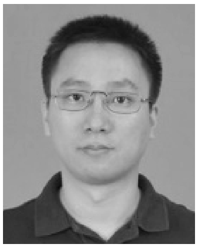
[46] C. Xu, T. Liu, J. Guan, H. Zhang, and G.-M. Muntean, "CMT-QA: Quality-aware adaptive concurrent multipath data transfer in heterogeneous wireless networks," *IEEE Trans. Mobile Comput.*, vol. 12, no. 11, pp. 2193–2205, Nov. 2013.

[47] E. Yaacoub, Z. Dawy, and A. Abu-Dayya, "On real-time video streaming over LTE networks with mobile-to-mobile cooperation," in *Proc. 19th Int. Conf. Telecommun. (ICT)*, Apr. 2012, pp. 1–6.

[48] L. Zhang, X. Zhang, K. Qu, L. Ren, J. Deng, and K. Zhu, "Green and cooperative DASH in wireless D2D networks," *Wireless Pers. Commun.*, vol. 84, no. 3, pp. 1797–1816, Oct. 2015.

[49] Y. Zhang, C. Li, and L. Sun, "DECOMOD: Collaborative DASH with download enhancing based on multiple mobile devices cooperation," in *Proc. 5th ACM Multimedia Syst. Conf.*, 2014, pp. 160–163.

[50] X. Zhu, P. Agrawal, J. P. Singh, T. Alpcan, and B. Girod, "Distributed rate allocation policies for multihomed video streaming over heterogeneous access networks," *IEEE Trans. Multimedia*, vol. 11, no. 4, pp. 752–764, Jun. 2009.



JUNFEI HUANG received the Ph.D. degree in communications and information systems from Beijing University of Posts and Telecommunications (BUPT), China, in 2004. He was a Member of the Technical Staff at Lucent Technologies, Bell Labs. He is currently a Lecturer with the Research Institute of Network Technology, BUPT. His research interests include mobile Internet applications, software testing, and blockchain technology.

ZHAOWEN LIN received the Ph.D. degree in computer science and technology from Beijing University of Posts and Telecommunications (BUPT), China, in 2008. He is currently an Associate Professor with the Research Institute of Network Technology, BUPT. His research interests include mobile Internet security and new generation Internet technology.