

Received February 18, 2021, accepted February 20, 2021, date of publication February 24, 2021, date of current version March 5, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3061762

Constrained Maximum-Utility Rate Optimization for Unicast-Based Streaming Applications Using Differentiated Multipaths

JIACHEN WANG¹, HONGLIN XIE¹, YONGHUI HE¹, FENG ZONG¹, AND ZHIHONG WANG²

¹School of Information Engineering, Shandong Yingcai University, Jinan 250104, China

²School of Information Science and Engineering, Shandong Agriculture and Engineering University, Jinan 251100, China

Corresponding author: Honglin Xie (sd_xie@aliyun.com)

This work was supported by the Shandong provincial Natural Science Foundation of China under Grant ZR2013FM010 and Grant ZR2020MF057.

ABSTRACT The multipath-based method can effectively improve the video streaming quality. However, it suffers a limitation on the realistic deployment and needs global optimization to improve the network accommodation capability. In this article, we study a novel multipath-based streaming rate optimization solution for unicast-based video applications. The proposed solution employs a differentiated multipath delivery model, which uses a main path built at the network layer and multiple auxiliary paths built at the application layer. Because of the application layer-based implementation, auxiliary paths overcome the aforementioned limitation on deployment. Based on the proposed multipath delivery model, we study the global multipath-based streaming rate optimization problem, which strives for an effective tradeoff among video streaming quality, network accommodation capability and video delivery stability. The studied optimization problem is complicated, and we solve it approximately using an approximate algorithm and a convex optimization-based technique.


INDEX TERMS Multipath, congestion control, utility, network accommodation capability.

I. INTRODUCTION

In recent years, we have witnessed a rapid development of various video services, such as video on demand, live streaming and video conferencing. These video services have accounted for most of the data transmitted over the Internet, and video traffic continues to increase steadily every year. A high video streaming rate undoubtedly improves the quality of user experience. However, the networks cannot keep up with the growing bandwidth demand for video streaming despite the continually increasing network capacity [1]. Because of variability and limitation of available network bandwidth, it is necessary to dynamically adjust the video streaming rate according to current network conditions, which has stimulated the development of adaptive streaming techniques such as DASH (Dynamic Adaptive Streaming over HTTP) [2].

In adaptive streaming techniques, videos are encoded at multiple resolutions/quality levels. Then, an appropriate

quality level is selected according to current network conditions. As summarized in [3], the common approaches to selecting the quality level include client-side, server-side and network-assisted solutions. Client-side solutions are very flexible and scalable. However, they underperform in a multi-client scenario because of the competition for shared network resources [4], [5]. Server-side solutions determine quality levels for multiple clients accessing the same server and can alleviate the abovementioned problem to some extent. However, they cannot globally optimize the rate allocation for video streams from different servers. The emerging software-defined networking (SDN) techniques can conveniently ascertain network conditions [6], including network topology and link bandwidth, and monitor the traffic over links [7], [8], thereby making network-assisted rate allocation feasible. To date, several SDN-based solutions (e.g., [9] and [10]) for determining video streaming rates have been proposed. In these solutions, the video streaming rates are computed centrally according to current network conditions and link traffic, resulting in better optimization of rate allocation compared with client-side and server-side solutions.

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

Multipath-based streaming can further improve the performance of video delivery by using the network congestion avoidance capability and increasing the total video streaming rate based on completely or partially disjoint delivery paths [11], [12]. Despite the above advantages, multipath-based streaming suffers from a limitation in practical deployment. For a designated sender-receiver pair, video delivery performance improves with the increase in the number of the paths associated with that pair. However, in the current network architecture, the number of the associated paths is limited by the number of network interfaces of the receiver because the route is selected in terms of the destination address. In addition, it is known that the Internet faces the routing scalability problem [13], which further restricts multipath deployment. The number of multipaths also must be limited in software-defined networks because of the limited ternary content-addressable memory size [14]. In addition to being subject to the above multipath number restriction, multipath-based streaming reduces the network accommodation capability compared with single-path streaming because the completely or partially disjoint paths for the same sender-receiver pair inevitably increase the average delivery path length. A low network accommodation capability might cause more severe network congestion in wide area networks [15]. A feasible approach to coping with the above problem is to use a network-assisted solution to globally determine the streaming rates along various paths according to current network and application conditions. Unfortunately, to the best of our knowledge, no studies have explored global multipath rate allocation.

In this article, we propose a novel network-assisted and multipath-based streaming rate optimization solution, named elastic streaming rate orchestration (ESRO), for unicast-based and cache-supported video applications such as video on demand. ESRO globally and dynamically optimizes streaming rates according to current network and application conditions based on a differentiated multipath delivery (DMD) model. In comparison to existing multipath delivery models, the DMD model uses a main delivery path (MDP) and multiple auxiliary delivery paths (ADPs). MDPs and ADPs are used by different strategies. The MDP is created at the network layer, while the ADPs are established at the application layer with the assistance of some proxies. Using the DMD model, we study a global multipath-based streaming rate optimization problem called the constrained maximum-utility multipath rate orchestration (CMUMRO). CMUMRO fully considers the tradeoffs among video streaming quality, network accommodation capability and video delivery stability. We solve the CMUMRO problem approximately based on an approximate algorithm and a convex optimization solution. The global rate optimization in ESRO requires that the network be able to conveniently assess network conditions, including network topology and link capacity, and monitor link traffic. As explained earlier, the SDN technique can satisfy the above requirement.

Different from existing multipath delivery solutions, our solution attempts to use the shortest path as the MDP. As a result, our solution can better improve the network accommodation capability. Because of the application layer-based implementation, ADPs in our solution can be flexibly deployed, thereby overcoming the aforementioned multipath count restriction. Our solution aims to solve a multi-objective multipath streaming rate optimization problem, which fully considers the whole video streaming quality, network accommodation capability and the whole video delivery stability. Compared with the optimization problems of existing multipath delivery solutions, the above multi-objective optimization can simultaneously improve the whole video delivery performance and network service capability.

The remainder of this article is organized as follows. We introduce the related work in Section II. The system architecture and the DMD model are introduced in Section III. In Section IV, we describe the CMUMRO problem and present the algorithm that solves this problem approximately. A performance evaluation is provided in Section V. Finally, we conclude our work in Section VI.

II. RELATED WORK

To date, video streaming rate optimization has been explored by numerous studies. These research efforts mainly focus on video coding, adaptive video streaming and multipath-based streaming enhancement.

Adaptive coding is a precondition of adaptive video streaming techniques. Two popular coding techniques are H.264/MPEG-4 advanced video coding (AVC) and scalable video coding (SVC). AVC encodes a video into L independent versions of different quality levels. In SVC, a chunk is encoded into ordered layers: one base layer (layer 0) with the lowest playable quality, and multiple enhancement layers (layers with $i > 0$) that further improve the chunk quality based on layer $i - 1$. Compared with AVC, SVC is more scalable and can better adapt to network congestion. With adaptive coding, adaptive video streaming techniques can select video quality levels according to current network conditions. The video quality level is usually selected by clients or servers. HTTP adaptive streaming (HAS) is a typical client-side adaptive streaming technique. Several adaptive streaming multicast techniques (e.g., [16]) also determine streaming rates at the client side. The main disadvantage of client-side adaptive streaming is that the heuristics used by clients cannot select the appropriate quality level if multiple users request the same video from a server [4], [5]. To address the above problem, several server-side adaptive streaming techniques have been proposed [17]. In addition to the client-side and server-side adaptive streaming techniques, several other solutions based on proxies or particular network devices [18] have been devised.

Network-assisted streaming rate optimization can set streaming rates according to whole network and application requirements. As mentioned above, SDN techniques make network-assisted adaptive streaming feasible. In recent

years, several SDN-based streaming rate optimization solutions (e.g., [9], [10], and [19]) have been proposed. Georgopoulos *et al.* [9] presented an OpenFlow-enabled system where an orchestrating module explicitly informed the DASH players of the representation they had to select to achieve user-level fairness in an adaptive video streaming environment. In [19] Kleinrouweler *et al.* proposed a DASH-aware networking architecture based on SDN. In that architecture, the controller assisted the players in selecting optimal rates. Cofano *et al.* [10] explored several network-assisted streaming strategies that relied on active cooperation between video streaming applications and the network, and they solved a max-min fairness optimization problem. In contrast to the existing SDN-based streaming rate optimization solutions, our solution strives to arrange multipath-based streaming rates using a particular model. In addition, our approach aims to solve a different optimization problem.

Because multipath-based method can effectively improve video delivery performance, it has been extensively researched [11], [12], [20]–[22]. Chen *et al.* [21] performed experimental measurements on various applications using single-path TCP, two-path MPTCP and four-path MPTCP, and they showed that MPTCP could be reasonably used for video streaming. In [12], Han *et al.* proposed a multipath framework for DASH streaming with awareness of users’ network interface preferences. In [20], Wu *et al.* investigated the problems of using multihomed terminals to stream video on mobile devices in a heterogeneous wireless network. In [11], the authors studied a reliable adaptive multipath provisioning method that allowed a steady flow of a significant portion of the traffic along multiple paths. In [15], the authors proposed an SDN-based dynamic multipath routing technique. Furthermore, Doshi *et al.* [22] proposed a solution to compute k -max min disjoint paths in SDN. Note that it is easier to implement multipath-based streaming in the SDN environment because different paths can be distinguished by the combination of various types of identifiers (including IP address and port). In the existing multipath-based streaming solutions, the multiple paths for a sender-receiver pair do not differ in delivering video data and are deployed at the network layer. In contrast, our solution performs differentiated multipath delivery and implements multipaths at the network and application layers. Additionally, the existing multipath-based streaming solutions do not study the global streaming rate orchestration, which is one of the main concerns in this article.

III. ARCHITECTURE AND MODEL

A. SYSTEM ARCHITECTURE

Fig. 1 describes our proposed network-assisted video streaming rate optimization solution, called ESRO. In ESRO, the video data are delivered to a receiver based on a particular multipath delivery model, i.e., the DMD model. The delivery paths in this model can be classified into two types, i.e., MDPs and ADPs. The latter are built at the application layer with the assistance of special servers named rate

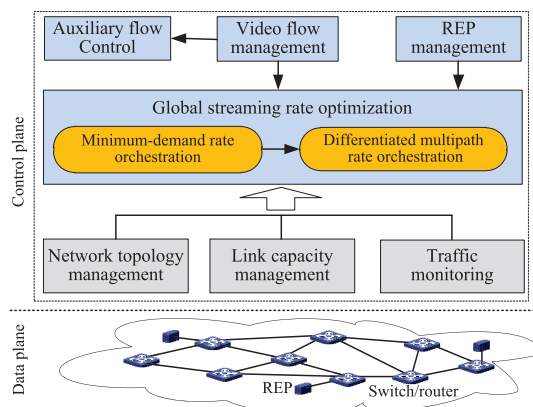


FIGURE 1. System architecture of our proposed solution.

enhancement proxies (REPs) that are deployed at dispersed locations in the network. The video flow along the MDP is called the main video flow, and the video flow along an ADP is called an auxiliary video flow. We further introduce the DMD model in detail in the next section. Similar to many content distribution systems (e.g., [23]), each video file is divided into many chunks. These chunks are identified by their sequences. ESRO uses the SVC technique to encode the video data. We reasonably assume that the clients can cache the received video data.

The ESRO system includes the following seven components: network topology management, link capacity management, traffic monitoring, streaming rate optimization, video flow management, REP management and auxiliary flow control. ESRO obtains network topology using the network topology management component, determines the link capacity using the link capacity management component, and assesses the link traffic conditions using the traffic monitoring component. The video flow management component is responsible for managing video sessions’ delivery paths and bandwidth requirements corresponding to layers of different quality; the REP management component is responsible for recording REPs and periodically verifying that REPs remain alive using heartbeat messages. The auxiliary flow control component is responsible for reducing the rates of auxiliary video flows by using the even reduction rule so that the sum of bandwidth consumed by auxiliary video flows and other flows at each link does not exceed a designated threshold. Because ADPs usually consume more bandwidth resources than do other paths, the above process helps improve the network accommodation capability.

Our proposed solution is not designed to set initial rates for video sessions; instead, it globally optimizes the rates of existing video sessions on demand. Thus, it overcomes the common scalability problem of the centralized method. The initial video streaming rate can be determined by the existing solutions (e.g., the DASH method [2]). The streaming rate optimization component periodically assesses the video streaming rate optimization conditions. If some video

sessions do not attain the expected high quality levels, this component globally orchestrates video streaming rates according to current network and application conditions, as described in Section IV. The streaming rate optimization scheme computed in the above procedure is implemented by notifying the relevant video senders to make the corresponding rate adjustments.

In the following part, we discuss the implementation of the ESRO system. The streaming rate optimization, REP management and auxiliary flow control can be easily implemented because they are performed at the application layer. Using the knowledge of the network topology and routing policy, video flow management can also be implemented. In SDN, a logically centralized controller manipulates network behaviors through a communication interface called a southbound interface [24]. Therefore, video flow management can be implemented more easily. In traditional networks, it is not easy to obtain necessary network conditions, including network topology and link capacity, and monitor the traffic at links. However, the three related components can be easily implemented in the SDN environment. Existing SDN controllers (e.g., OpenDaylight [25]) have implemented topology management. Link capacity management can be easily implemented based on OpenFlow, a popular communication interface of SDN. In the OpenFlow switch specification [26], the port description request OFPMP_PORT_DESCRIPTION enables the controller to obtain a description of all the standard ports of the OpenFlow switch. The reply body includes a field *max_speed*, which denotes maximum bitrate of the port. Thus, the link bandwidths can be obtained based on the OFPMP_PORT_DESCRIPTION request and the reply to it. The link traffic monitoring methods have been explored by many studies [7], [8]. These methods can be used in ESRO. Though the ESRO system can be easily implemented in the SDN environment, it is not exclusively limited to that environment. In fact, the ESRO system can be conveniently deployed in any network if the necessary network conditions can be obtained by some method.

B. THE DMD MODEL

As introduced previously, the DMD model employs two types of video delivery paths, i.e., MDPs and ADPs. Each main or ADP is associated with a designated sender-receiver pair, as shown in Fig. 2. The MDP can be built by an existing unicast routing protocol that aims to minimize the path length. Note that the shortest unicast routing protocols (e.g., open shortest path first [27]) have been widely studied. Video delivery based on an ADP uses either TCP or UDP. In this article, we only consider TCP-based delivery because of reliability. Note that the maximum rate of a TCP connection is limited by an existing solution (e.g., [28]) to ensure that a rate allocation scheme operates correctly.

Functionally, the ADPs can be built at the network layer by a revised multipath routing solution, in which ADPs should be selected to avoid the links used in the corresponding MDP.

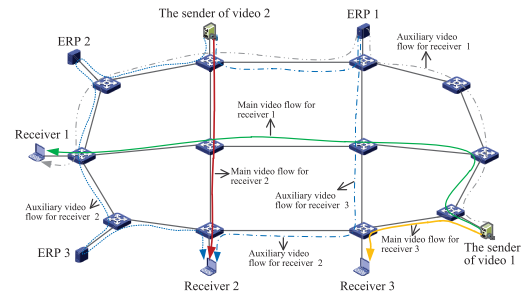


FIGURE 2. Multipath-based video delivery.

However, to solve the multipath number restriction, we only consider building ADPs at the application layer in this article. For an ADP, we use a forwarding sequence to denote the sequence of addresses of the application-layer nodes (REPs and the receiver) in the ADP. When the video sender delivers data along an ADP, it sends data packets, each of which contains the corresponding forwarding sequence, to the first REP in the ADP. When a REP receives a data packet, it retrieves the second address of the forwarding sequence in the data packet and deletes the first address (i.e., the address of that REP). Then, it forwards modified data packets to the next node, which is identified by the second address mentioned above, in the ADP. The above procedure continues until the data packet arrives at the receiver. To reduce the delay of data delivery based on an ADP and limit the extra load caused by the forwarding sequences embedded in data packets, the maximum number of the REPs in an ADP is limited by parameter Th_{REP} . This parameter is set to a low value. An example value could be $Th_{REP} = 3$. To avoid an excessively high reordering overhead, the maximum number of ADPs for a sender-receiver pair is limited by the parameter Th_{adp} , which can be configured based on a practical test.

For simplicity, the router or the switch to which a video sender or receiver is connected is called video access node. We use the REP-based relay path to denote a path from a video access node to another video access node that contain at least one REP and at most Th_{REP} REPs. Because the network, apart from the host nodes, is relatively stable, the REP-based relay paths for video access node pairs can be computed in advance. Thus, the ADPs for a sender-receiver pair can be immediately obtained according to the corresponding REP-based relay paths, if necessary. Because Th_{REP} is set to a low value, all REP-based relay paths from a video access node to another video access node can be computed within an acceptable time. To further select at most Th_{adp} REP-based relay paths with good congestion avoidance capability, we define a link overlap avoidance degree. Such a degree for an REP-based relay path ζ associated with another REP-based relay path ζ' is denoted by $\rho(\zeta, \zeta')$ and is defined as

$$\rho(\zeta, \zeta') = |\mathcal{L}(\zeta') - \mathcal{L}(\zeta) \cap \mathcal{L}(\zeta')|, \quad (1)$$

where $\mathcal{L}(X)$ indicates the set of links in path X . Assume that ξ_1 denotes the MDP that contains two video access nodes A and B ; that Ξ represents the set of REP-based relay paths from A to B ; that ξ_{i+1} ($1 \leq i \leq Th_{adp}$) indicates the i th selected REP-based relay path from A to B ; and that $Ed(\zeta, i) = \min\{\rho(\zeta, \xi_j) | 1 \leq j \leq i\}$. Then, we can select the i th REP-based relay path according to the following steps: find a path ζ among Ξ_i ($\Xi_i = \Xi - \{\xi_2, \xi_3, \dots, \xi_i\}$) such that $Ed(\zeta, i) = \max\{Ed(\zeta', i) | \zeta' \in \Xi_i\}$; if $Ed(\zeta, i) > 0$; then, ζ becomes the i th selected REP-based relay path.

In the DMD model, MDPs and ADPs are used by different strategies. ADPs are used to improve the quality of video streams only if the expected streaming rates cannot be obtained over MDPs. Because the latter are constructed using the shortest unicast routing, video data distribution based on MDPs obviously improves the network accommodation capability. To avoid potential network congestion, an ADP and the corresponding MDP should share as few links as possible. The above feature and the application layer-based implementation clearly cause the ADP to be longer than the corresponding MDP. As a result, an auxiliary video flow should be controlled when it competes for the bandwidth resources with a main video flow. The above description implies that a global streaming rate optimization is necessary for optimizing the tradeoffs among video streaming quality, network accommodation capability and video delivery stability. We will further describe the global streaming rate optimization in Section IV-A.

A potential weakness of multipath-based video delivery entails reordering overhead. Several previous studies (e.g., [29]) have shown that reordering overhead remains acceptable. Another concern is the difference in delays between different video delivery paths. This problem can be solved using content caching, as most video delivery solutions (e.g., [30]) do. In this article, we only consider unicast-based video streaming. However, the DMD model can also be applied to optimize multicast-based video streaming rates when the main delivery path is replaced by the multicast tree.

IV. CONSTRAINED MAX-UTILITY RATE ORCHESTRATION

A. PROBLEM DESCRIPTION

We first introduce the utility function used for video streaming rate optimization. Several previous studies (e.g., [31]–[33]) use a weighted logarithmic function to express the utility, hereafter called the rate utility, corresponding to a designated rate of a video session (i.e., the video delivery from the sender to the receiver). Specifically, the utility of rate r_i associated with session i is denoted by $U_i(r_i)$ and computed using Eq. (2). Note that ω_i denotes the weight of session i . The weighted logarithmic function ensures that the marginal rate utility of a flow decreases as the throughput increases, and recent advancements in end-to-end adaptation of application flows (e.g., adaptive video streaming) allow modeling most application traffic as elastic [32]. Note that the rate utility can be considered as to be continuous when

TABLE 1. Notation Used in Streaming Rate Orchestration

Notation	Description
n_s	Total number of video sessions
n_l	Total number of communication links
$U_i(r_i)$	Utility of rate r_i associated with video session i
ω_i	Weight of video session i
LR_i	Minimum-demand rate of video session i
\hat{B}_j	Available bandwidth of link j
$K(i, j)$	MDP predicate function
$\mu(i, r_i)$	Minimum-demand rate predicate function
$\lambda(i, j, k)$	Delivery path predicate function
r_i	Basic rate, allocated by Algorithm 1, of video session i
$x_{i,k}$	Extended rate of the k delivery path for video session i
\mathbf{r}, \mathbf{x}	Rate column vectors
δ_i	Number of delivery paths for video session i
α	Weight of the main video delivery
UB_i	Maximum rate for video session i

the content delivered in advance can be cached though video quality levels require discrete rates.

$$U_i(r_i) = \omega_i \times \log r_i. \tag{2}$$

As described above, our proposed solution adopts the SVC encoding technique. In SVC, the base layer is characterized by the lowest playable quality. In this article, the rate corresponding to the base layer is called the minimum-demand rate. We use LR_i to denote the minimum-demand rate of session i . For convenience, we list the notation used in streaming rate orchestration in Table 1. Note that video sessions and links are numbered in this article. Sometimes, we use a number to refer to a video session or a link. The logarithmic function has a feature whereby a small difference in the rate value near 0 leads to a very large difference in the rate utility, which is not appropriate for practical utility evaluation. To cope with this problem, we assume that $\log(LR_i) \geq 1$ for any video session i . This assumption can be easily satisfied by adjusting the unit of measurement of streaming rates so that $\log(LR_i) \geq 1$.

Two most popular optimization objectives of video streaming rate optimization are max-min fairness [9], [10] and utility-maximization [33]–[35]. In this article, we present a novel optimization problem called CMUMRO, specified in terms of the DMD model and the rate utility defined by Eq. (2). The CMUMRO problem includes the following two subproblems: minimum-demand rate orchestration (MDRO) and differentiated multipath rate orchestration (DMRO).

The MDRO subproblem entails maximizing the rate utility corresponding to the allocated minimum-demand rates based on MDPs and all available bandwidth resources. Let r_i denote the main delivery rate (i.e., the rate of the main video flow) of session i and \mathbf{r} ($\mathbf{r} = [r_1, r_2, \dots, r_{n_s}]^T$) denote the main delivery rate vector. Formally, the MDRO subproblem can be described as

$$\begin{aligned} & \underset{\mathbf{r}}{\text{maximize}} \sum_{i=1}^{n_s} \mu(i, r_i) \cdot U_i(r_i) \\ & \text{s.t.} \sum_{i=1}^{n_s} K(i, j)r_i \leq \hat{B}_j, \quad \forall j \ (1 \leq j \leq n_l) \end{aligned} \tag{3}$$

$$r_i \geq 0, \quad \forall i \ (1 \leq i \leq n_s) \quad (4)$$

$$r_i \leq LR_i, \quad \forall i \ (1 \leq i \leq n_s) \quad (5)$$

where $K(i, j)$ is defined as

$$K(i, j) = \begin{cases} 1 & \text{if } i\text{'s MDP includes link } j, \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

The function $\mu(i, r_i)$ is defined as

$$\mu(i, r_i) = \begin{cases} 1 & \text{if } r_i \geq LR_i, \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

According to the definition, MDRO attempts to achieve a weighted fairness because it uses at most a minimum-demand rate for each video session based on all available bandwidth resources. In other words, MDRO offers the lowest playable quality for as many video sessions as possible. If the video sessions have the same weight, the optimal solution of MDRO is max-min fair, yet such a scheme may not be the optimal solution of MDRO because MDRO maximizes the rate utility. As a result, we suggest that MDRO can further improve the video quality. MDRO is also different from the common utility maximization problem because the streaming rate in MDRO is, in terms of Eq. (7), a choice between two designated values.

DMRO sets video streaming rates globally based on all delivery paths and the available bandwidth resources except for those used by the rate scheme computed by MDRO. It aims to obtain a desired tradeoff among high rate utility, low bandwidth consumption and high video delivery stability by varying the main video flow weight α ($\alpha \geq 1$). Let $x_{i,k}$ denote the streaming rate of the k th delivery path for video session i and define $\mathbf{x} = [x_{1,1}, x_{1,2}, \dots, x_{1,\delta_1}, \dots, x_{n_r,1}, x_{n_r,2}, \dots, x_{n_r,\delta_{n_r}}]^T$. Note that $x_{i,1}$ is the MDP of video session i . Formally, the DMRO problem can be expressed as

$$\begin{aligned} & \text{maximize } \sum_x \sum_{i=1}^{n_s} U(r_i + \alpha x_{i,1} + \sum_{k=2}^{\delta_i} x_{i,k}) \\ & \text{s.t. } \sum_{i=1}^{n_s} \left(\sum_{k=1}^{\delta_i} \lambda(i, j, k) x_{i,k} + K(i, j) r_i \right) \leq \hat{B}_j, \\ & \quad \forall j \ (1 \leq j \leq n_l), \end{aligned} \quad (8)$$

$$\sum_{k=1}^{\delta_i} x_{i,k} + r_j \leq UB_i, \quad \forall i \ (1 \leq i \leq n_s), \quad (9)$$

$$x_{i,k} \geq 0, \quad \forall i, k \ (1 \leq i \leq n_s, 1 \leq k \leq \delta_i), \quad (10)$$

where

$$\lambda(i, j, k) = \begin{cases} 1 & \text{if } j \text{ is in the } k\text{th delivery path of } i, \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

Note that UB_i is slightly greater than the rate corresponding to the highest streaming quality of video session i and r_i is the rate of video session i in the rate allocation scheme computed by the DMRO.

B. MINIMUM-DEMAND MAIN DELIVERY RATE ORCHESTRATION

In this section, we introduce how to solve the MDRO problem, which maximizes the utility corresponding to the allocated minimum-demand rates for MDPs based on all available bandwidth resources. We prove the following lemma.

Lemma 1: The MDRO problem is NP-hard.

Proof: Consider the following case of the MDRO problem: all delivery paths include the same link l , and l cannot accommodate all video sessions; the links with the exception of l can each accommodate all video sessions. We define an extended weight ω_i^+ by $\omega_i^+ = \omega_i \log r_i$ for video session i . Thus, the MDRO problem in the above case becomes the knapsack problem, which is NP-hard. This lemma has thus been proven. ■

We present an approximate algorithm for solving the MDRO problem, as shown in Algorithm 1. A rate allocation scheme, denoted by \mathbf{r} , is feasible if it satisfies the constraints (3)-(5). Algorithm 1 uses a rate increase priority and a rate decrease priority. The rate increase priority of video session i associated with link set \mathcal{C} is denoted by $\psi(i, \mathcal{C})$ and defined as

$$\psi(i, \mathcal{C}) = \frac{\omega_i \cdot \log(LR_i)}{\sum_{l \in \mathcal{C}} K(i, j)}. \quad (12)$$

The rate decrease priority of video session i associated with rate allocation vector \mathbf{r} is denoted by $\varphi(i, \mathbf{r})$ and defined as

$$\varphi(i, \mathbf{r}) = \frac{\sum_{j \in \Xi_i} \max\{0, \hat{B}_j + r_i - \sum_{k=1}^{n_s} K(k, j) r_k\}}{\omega_i \cdot \log(LR_i)}. \quad (13)$$

Algorithm 1 first sets the minimum-demand rates for video sessions according to the rate increase priority, as shown in Lines 1-6. According to Eq. (12), the above process helps improve the rate utility. When the remaining bandwidth resources are insufficient for providing minimum-demand rates for more video sessions, Algorithm 1 attempts to replace some minimum-demand rates with a minimum-demand rate with a larger absolute rate utility, as shown in Lines 7-14. In the above procedure, the minimum-demand rate replacement is performed in terms of the rate decrease priority. Considering Eq. (13), we observe that the rate decrease priority can effectively reduce the replacement cost. The steps in Lines 7-14 can improve the rate utility by fully using the marginal bandwidth resources, which is a beneficial addition to the utility maximum-increasing rate arrangement (Lines 1-6). The time complexity of Algorithm 1 depends on the computation of the rate increase and rate decrease priorities. According to Eqs. (12) and (13), the time complexity of Algorithm 1 can be expressed as $O(n_s \cdot n_l)$. This time complexity is relatively low, and it facilitates quick adjustments of video streaming rates.

Assume that $\chi(\mathcal{C})$ denotes the set of video sessions passing through a link belonging set \mathcal{C} computed in Line 1. Set $\chi(\mathcal{C})$ can be divided into several disjoint subsets such that (1)

Algorithm 1 Minimum-Demand Rate Arrangement

Input: MDP for each session i ($1 \leq i \leq n_s$).
Output: \mathbf{r} .

- 1 $\mathcal{C} \leftarrow \{j | 1 \leq j \leq n_l, \sum_{i=1}^{n_s} K(i, j)LR_i > \hat{B}_j\}$.
- 2 For each video session i , compute the rate increase priority $\psi(i, \mathcal{C})$; $\mathbf{r} \leftarrow \mathbf{0}$.
- 3 Sort video sessions in the descending order of rate increase priority. /* Assume that the sorted sessions are denoted by q_1, q_2, \dots, q_{n_s} */
- 4 **for** $i=1$ **to** n_s **do**
- 5 $r_{q_i} \leftarrow LR_{q_i}$.
- 6 **if** \mathbf{r} is infeasible **then** $r_{q_i} \leftarrow 0$.
- 7 $\mathcal{W} \leftarrow \{i | 1 \leq i \leq n_s, r_i = 0\}$.
- 8 Sort the sessions in \mathcal{W} in the descending order of rate increase priority. /* Assume that the sorted sessions are denoted by $q'_1, q'_2, \dots, q'_{|\mathcal{W}|}$ */
- 9 **for** $i = 1$ **to** $|\mathcal{W}|$ **do**
- 10 $\mathcal{V} \leftarrow \{j | r_j \neq 0, 1 \leq j \leq n_s\}$.
- 11 $\mathbf{r}' \leftarrow \mathbf{r}, r_{q'_i} \leftarrow LR_{q'_i}, g \leftarrow \omega_{q'_i} \cdot \log(LR_{q'_i})$.
- 12 **while** \mathbf{r} is infeasible and $\mathcal{V} \neq \emptyset$ **do**
- 13 Find the video session k among \mathcal{V} with the maximum rate decrease priority $\varphi(k, \mathbf{r})$;
 $\mathcal{V} \leftarrow \mathcal{V} - \{k\}$.
- 14 $p \leftarrow \omega_k \cdot \log(LR_k)$.
- 15 **if** $p < g$ **then** $r_k \leftarrow 0, g \leftarrow g - p$.
- 16 **if** \mathbf{r} is infeasible **then** $\mathbf{r} \leftarrow \mathbf{r}'$.

each video session of $\chi(\mathcal{C})$ is in a unique subset; (2) for any two video sessions in different subsets, the MDPs of these two sessions contain no identical links; and (3) the above two conditions do not hold if any two subsets are combined. The above subset partition is called the maximum rate irrelevance partition. We use $\Psi(\mathcal{C})$ to denote the combination of the subsets in the maximum rate irrelevance partition of $\chi(\mathcal{C})$. Let $\mathcal{Q}(i)$ denote the set of links in the MDP of session i . We reasonably assume that the available bandwidth of each link is sufficient for video streaming with the highest quality. Then, we obtain the following lemma.

Lemma 2: The approximation ratio of Algorithm 1 is $\frac{1}{\text{MAXD}+1}$, where $\text{MAXD} = \max\{|\mathcal{C} \cap \bigcup_{i \in P} \mathcal{Q}(i)| | P \subseteq \Psi(\mathcal{C})\}$.

Proof: For link l in \mathcal{C} , we call video session i the overflow session of l if i is the first session that cannot be successfully allocated, in Lines 5 and 6, a minimum-demand rate because of the bandwidth limitation of l . For any subset $P \subseteq \Psi(\mathcal{C})$, we use \mathcal{H}_P to denote the set of video sessions that are successfully allocated minimum-demand rates in Lines 5 and 6, and use \mathcal{D}_P to denote the set of overflow sessions of the links in $\mathcal{C} \cap \bigcup_{i \in P} \mathcal{Q}(i)$. Let opt_P and alg_P denote the optimal (i.e., the largest) rate utility and the rate utility computed by

Algorithm 1, respectively. We have

$$opt_P \leq \sum_{i \in \mathcal{H}_P} \omega_i \cdot \log(LR_i) + \sum_{j \in \mathcal{D}_P} \omega_j \cdot \log(LR_j) \quad (14)$$

From Algorithm 1, we can deduce that

$$alg_P \geq \max\left\{ \sum_{i \in \mathcal{H}_P} \omega_i \cdot \log(LR_i), \max\{\omega_j \cdot \log(LR_j) | j \in \mathcal{D}_P\} \right\} \quad (15)$$

Using Eqs. (14) and (15), we obtain $\frac{opt_P}{alg_P} \leq \frac{1}{|\mathcal{D}_P|+1} \leq \frac{1}{\text{MAXD}+1}$. Because the rate allocation schemes for the subsets of $\Psi(\mathcal{C})$ are independent of each other, we have $\frac{opt}{alg} \leq \frac{1}{\text{MAXD}+1}$. Thus, this lemma has been proven. ■

C. UTILITY-ENHANCED MULTIPATH RATE ORCHESTRATION

Algorithm 1 allocates minimum-demand rates to video sessions. Based on the result of the minimum-demand rate allocation, video streaming rates are enhanced using the multipaths to solve the DMRO problem. According to the problem description, DMRO is equivalent to minimizing the following function:

$$f(\mathbf{x}) = - \sum_{i=1}^{n_s} U_i(r_i + \alpha x_{i,1} + \sum_{k=2}^{\delta_i} x_{i,k}). \quad (16)$$

Note that in this section \mathbf{r} is a constant vector that represents the minimum-demand rate allocation scheme computed by Algorithm 1. We can derive the following lemma.

Lemma 3: $f(\mathbf{x})$ is a convex function.

Proof: We first prove that $f_i(\mathbf{x}) = -U_i(r_i + \alpha x_{i,1} + \sum_{k=2}^{\delta_i} x_{i,k})$ is a convex function. Here, r_i is a constant. We can deduce that the Hessian Matrix H_i ($H_i = \nabla^2 f_i(\mathbf{x})$) is

$$H_i = \frac{1}{X^2} \cdot \begin{bmatrix} \alpha^2 & \alpha & \dots & \alpha \\ \alpha & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha & 1 & \dots & 1 \end{bmatrix}_{\delta_i \times \delta_i}, \quad (17)$$

where $X = r_i + \alpha x_{i,1} + \sum_{k=2}^{\delta_i} x_{i,k}$. We observe that $H_i = B_i \cdot (B_i)^T$, where B_i is defined as

$$B_i = \frac{1}{X} \cdot \begin{bmatrix} \alpha & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 0 \end{bmatrix}_{\delta_i \times \delta_i}. \quad (18)$$

Thus, we obtain that $\nabla^2 f_i(\mathbf{x})$ is a positive semidefinite matrix. The sum operation preserves convexity. This lemma has thus been proven. ■

Because $f(\mathbf{x})$ is a convex function and the definition domain of the DMRO problem is a convex set, the optimal

solution of this problem can be computed by convex optimization, which has been widely studied. We can use an existing method to solve this problem. Specifically, in this article, we apply the barrier method to solve the DMRO problem. We use Newton's method to minimize a function of the form

$$\begin{aligned} \phi(\mathbf{x}) = & -t \sum_{i=1}^{n_s} U(r_i + \alpha x_{i,1} + \sum_{k=2}^{\delta_i} x_{i,k}) - \\ & \sum_{j=1}^{n_l} \log \left(\hat{B}_j - \sum_{i=1}^{n_s} \left(\sum_{k=1}^{\delta_i} \lambda(i, j, k) x_{i,k} + K(i, j) r_i \right) \right) \\ & - \sum_{i=1}^{n_s} \log(UB_i - r_i - \sum_{k=1}^{\delta_i} x_{i,k}) - \sum_{i=1}^{n_s} \log \left(\sum_{k=1}^{\delta_i} x_{i,k} \right), \end{aligned} \quad (19)$$

where t is an accuracy parameter. Note that convex optimization using the barrier method is $\frac{1}{t} \sum_{i=1}^{n_s} \delta_i$ -suboptimal [36].

In the optimization procedure, the Newton step Δx_{nt} is determined by solving the linear equations

$$(D_0 + A^T D_1 A + D_2 + D_3) \Delta x_{nt} = -g, \quad (20)$$

where g is the gradient of $\phi(\mathbf{x})$.

$$D_0 = t \text{diag} \left(\frac{\partial^2 f}{\partial x_1^2}, \frac{\partial^2 f}{\partial x_2^2}, \dots, \frac{\partial^2 f}{\partial x_{N_s}^2} \right). \quad (21)$$

$$D_1 = \text{diag} (Y_1^2, Y_2^2, \dots, Y_{n_l}^2), \quad (22)$$

where $Y_j = \hat{B}_j - \sum_{i=1}^{n_s} \left(\sum_{k=1}^{\delta_i} \lambda(i, j, k) x_{i,k} + K(i, 1) r_i \right)$.

$$D_2 = \text{diag} \left(\frac{1}{Z_1^2}, \frac{1}{Z_2^2}, \dots, \frac{1}{Z_{n_s}^2} \right), \quad (23)$$

where $Z_i = UB_i - r_i - \sum_{k=1}^{\delta_i} x_{i,k}$.

$$D_3 = \text{diag} \left(\frac{1}{\left(\sum_{k=1}^{\delta_1} x_{1,k} \right)^2}, \dots, \frac{1}{\left(\sum_{k=1}^{\delta_{n_s}} x_{n_s,k} \right)^2} \right). \quad (24)$$

More details about the Newton's method and the related information on convex optimization can be obtained from [36].

Let $f'(\mathbf{x}) = - \sum_{i=1}^{n_s} U_i(r_i + x_{i,1} + \sum_{k=2}^{\delta_i} x_{i,k})$. For a rate allocation scheme \mathbf{x}^* computed by the barrier-based optimization method, we refer to the value of $-f'(\mathbf{x}^*)$ as real utility. The weight α helps obtain a tradeoff among high rate utility, low bandwidth consumption and high video delivery stability. However, it reduces the real utility. Let alg_m denote the real utility of the rate allocation scheme computed by the barrier-based optimization method, and opt_m denote the optimal real utility of the rate allocation scheme computed without using the weight α . We derive the following lemma.

Lemma 4: Under the assumption that the total rate allocated to each video session i is not less than LR_i , $\frac{opt_m}{alg_m}$ is not larger than $\frac{t(\log \alpha + 1)}{\sum_{i=1}^{n_s} \delta_i}$.

Proof: Let \mathbf{x}^* denote the rate allocation scheme computed by the barrier-based optimization method. Assume that such a method obtains the optimal solution of the DMRO problem. We use alg_m to denote the real utility corresponding to the above optimal solution. We have

$$\begin{aligned} opt_m \leq -f(\mathbf{x}^*) & \leq \sum_{i=1}^{n_s} U_i(\alpha r_i + \alpha x_{i,1}^* + \alpha \sum_{k=2}^{\delta_i} x_{i,k}^*) \\ & \leq \sum_{i=1}^{n_s} \omega_i \log \alpha + alg_m \end{aligned} \quad (26)$$

As noted in Section IV-A, $\log(LR_i) \geq 1$ for any video session i . Under the assumption that the total rate allocated to each video session i is not less than LR_i , we can deduce that

$$\frac{opt_m}{alg_m} \leq \frac{\sum_{i=1}^{N_s} \omega_i \log \alpha}{alg_m} + 1 \leq \log \alpha + 1. \quad (27)$$

Because convex optimization using the barrier method is $\frac{1}{t} \sum_{i=1}^{n_s} \delta_i$ -suboptimal, we have $\frac{opt_m}{alg_m} \leq \frac{t(\log \alpha + 1)}{\sum_{i=1}^{n_s} \delta_i}$. Thus, this lemma has been proven. ■

Based on Lemma 4, we note that α can increase the use of MDPs with a relatively low influence on the real utility. Note that t is a parameter that can be configured in terms of the expected accuracy requirement.

V. EXPERIMENTAL RESULTS

In this section, we evaluate our proposed ESRO solution based on simulations and real-world Internet experiments.

A. SIMULATION RESULTS

Our simulations used a 197-node topology described by the Cogentco dataset in the Internet Topology Zoo [37]. The nodes in the 197-node topology are called routing nodes in this article. We generated 50 host nodes and 5 REP nodes that were connected to random routing nodes. The minimum-demand rates for video sessions were set to 200 Kbps; the maximum rate for a video session was randomly selected among 8 Mbps, 2 Mbps and 1 Mbps, according to the maximum rates corresponding to video resolutions of 1080p, 720p and 360p. The available bandwidth of each link, connecting two routing nodes, was set to a random value between 20 Mbps and 500 Mbps; the available bandwidth of each link connecting a routing node and a host node was set to 50 Mbps; the available bandwidth of each link connecting a routing node and an REP node was set to 100 Mbps. Note that the maximum available bandwidth of a link is obviously less than practical link capacity because our solution does

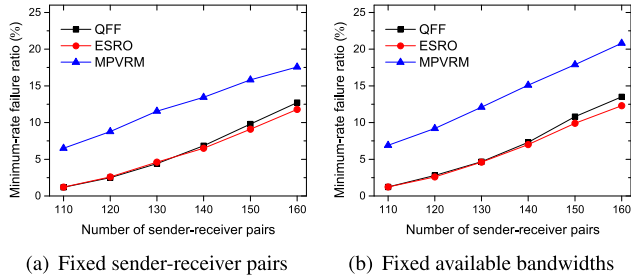


FIGURE 3. Comparison of the minimum-demand failure ratio.

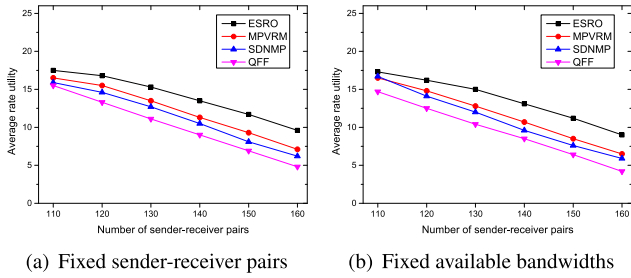


FIGURE 4. Comparison of the rate utility.

not schedule the traffic caused by other applications including real-time video streaming applications and non-video applications. The weights of video sessions were assigned to be integer values between 1 and 5. By default, the main delivery weight α was set to 1.5. The data points shown in the Figs. 3, 4, 5 and 6 represent averages over the 100 runs with a 95% confidence level.

We compare the performance of minimum-demand rate allocation of ESRO with QFF [9] and the rate allocation solution used in [33]. QFF optimizes the rates by the max-min fairness rule, while the rate allocation solution used in [33], called MPVRM for convenience, maximizes the rate utility. Note that MPVRM can be used in both unicast-based and multicast-based video streaming applications. We use the minimum-rate failure ratio to denote the ratio of the number of video sessions with rates that are lower than the minimum-demand rate to the total number of video sessions. Fig. 3 depicts our simulation results. In each run of the simulation shown in Fig. 3a, the available link bandwidths were independently configured, but the sender-receiver pairs were fixed; in each run of the simulation shown in Fig. 3b, the sender-receiver pairs were independently selected, but the available link bandwidths were fixed. The simulation results show that ESRO has a clearly lower minimum-rate failure ratio than that of MPVRM. The main reason is that our proposed solution firstly attempts to arrange the

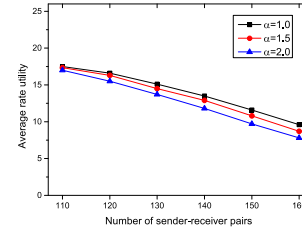


FIGURE 5. The relation between rate utility and α .

minimum-demand rates, based on all the available bandwidth resources, for all the sender-receiver pairs. Because the max-min fairness does not involve effectively avoiding the congestion, the minimum-demand failure ratio with QFF is higher than that with ESRO. However, the max-min fairness also helps reduce the minimum-rate failure ratio. As a result, the minimum-demand failure ratio with QFF is not evidently higher than that with ESRO, as Fig. 3 shows. Note that ESRO mainly optimizes the rate utility.

We compare the performance of the utility optimization of ESRO with that of QFF, MPVRM and a multipath-based solution proposed in [22], called SDNMP for convenience. In the simulation of [22], the rates for sender-receiver pairs were allocated in a random order, and the rates corresponding to the two paths of a sender-receiver pair are allocated evenly until one path cannot be allocated a higher rate because of the limited available bandwidth resources. In our simulations, the available bandwidth resources and sender-receiver pairs are set as in the simulations on minimum-demand rate allocation. Fig. 4 presents our simulation results. From this figure, we note that the average rate utility of our proposed solution appears to be higher than those of other solutions. The above advantage of our solution can be explained as follows: Because the auxiliary paths in our solution are built at the application layer, more delivery paths can be built for each sender-receiver pair in our solution, which indicates that our solution has higher resource utilization capability; additionally, the global streaming rate optimization is superior in improving the whole rate utility.

We investigated the influence of the main delivery weight α . In these experiments, the available bandwidth resources and the sender-receiver pairs were independently reconfigured in each run. Figs. 5 and 7 depict the simulation results. In Fig. 7, the main delivery ratio represents the ratio of the sum of the rates corresponding to MDPs to the total rate. From Fig. 5, the average rate utility slightly decreases as α increases. However, the main delivery ratio has an obvious increase with increasing α , as shown in Fig. 7. In our solution,

$$A = \begin{bmatrix} \lambda(1, 1, 1) & \lambda(1, 1, 2) & \cdots & \lambda(1, 1, \delta_1), \dots, \lambda(n_s, 1, 1) & \lambda(n_s, 1, 2) & \cdots & \lambda(n_s, 1, \delta_{n_s}) \\ \lambda(1, 2, 1) & \lambda(1, 2, 2) & \cdots & \lambda(1, 2, \delta_1), \dots, \lambda(n_s, 2, 1) & \lambda(n_s, 2, 2) & \cdots & \lambda(n_s, 2, \delta_{n_s}) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \lambda(1, n_l, 1) & \lambda(1, n_l, 2) & \cdots & \lambda(1, n_l, \delta_1), \dots, \lambda(n_s, n_l, 1) & \lambda(n_s, n_l, 2) & \cdots & \lambda(n_s, n_l, \delta_{n_s}) \end{bmatrix} \quad (25)$$

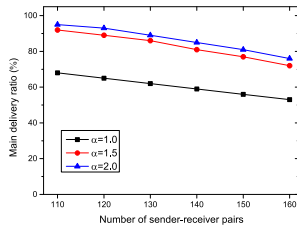


FIGURE 6. The relation between main delivery ratio and α .

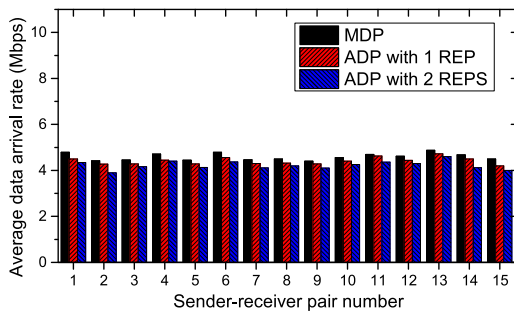


FIGURE 7. Average data arrival rates associated with sender-receiver pairs and delivery paths.

the MDP-based video delivery forwards data packets at the network layer, while the ADP-based delivery forwards data packets at the application layer. The data forwarding at the application layer needs more operations than the data forwarding at the network layer, which indicates that the data flow along a MDP is more stable than that along an ADP. As a result, the above results show that our solution can obtain a desired tradeoff between rate utility and delivery stability by using the main delivery weight.

B. INTERNET EXPERIMENT

We deployed 5 cloud virtual machines at different locations, 2 from EC2 and 3 from the Alibaba Cloud. These virtual machines were used as REPs. We also deployed 30 nodes (physical machines and virtual machines) at different locations, and generated 15 sender-receiver pairs based on those 30 nodes. We arranged an ADP including one REP and an ADP including 2 REPs. Each sender periodically sent UDP data packets to the corresponding receiver at a rate of 5 Mbps along the MDP and the two ADPs. We observed the data arrival rates at receivers, i.e., the data receiving rates of the receivers. Our experiments lasted for a week. Fig. 7 presents our experimental results. Note that each average data arrival rate in Fig. 7 indicates the average value of the arrival rates associated with a designated sender-receiver pair and a designated delivery path. From Fig. 7, the arrival rate of the data delivered along a MDP is higher than the arrival rates of the data delivered along corresponding ADPs, which indicates that the ADP-based video delivery can effectively avoid network congestion. In addition, we notice that the ADP-based video delivery has high arrival rates. The above result indicates that the ADP-based video delivery can well

use the network resources and effectively enhance the streaming rates in real-world networks.

VI. CONCLUSION

In this article, we proposed a multipath-based streaming rate optimization solution. In contrast to the existing multipath-based streaming solutions, our solution uses a differentiated multipath delivery model, which includes a main delivery path and multiple auxiliary paths. The main delivery path usually is the shortest path, which is helpful to improve the network accommodation capability. The auxiliary paths are built at the application layer, thereby overcoming the problem of multipath deployment restriction. Based on the proposed multipath delivery model, we studied a global multipath-based streaming rate optimization problem that fully considers the tradeoff among video streaming quality, network accommodation capability and video delivery stability. The studied optimization problem includes two sub-problems, i.e., MDRO and DMRO. The MDRO subproblem is NP-hard, and we solved it approximately using an approximate algorithm. We solved the DMRO subproblem with a convex optimization technique.

In this article, we only considered unicast-based streaming rate optimization. In our future work, we will study the global multipath-based streaming rate optimization problem for multicast-based video applications.

REFERENCES

- [1] *Cisco Visual Networking Index: Forecast and Methodology*, Cisco, San Jose, CA, USA, 2015.
- [2] *Information Technology: Dynamic Adaptive Streaming over HTTP (DASH) Part 1: Media Presentation Description and Segment Formats*, document ISO/IEC/ISO/IEC 2309-1:2012, 2012.
- [3] J. Kua, G. Armitage, and P. Branch, "A survey of rate adaptation techniques for dynamic adaptive streaming over HTTP," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1842–1866, 3rd Quart., 2017.
- [4] S. Akhshabi, L. Anantkrishnan, A. C. Begen, and C. Dovrolis, "What happens when HTTP adaptive streaming players compete for bandwidth?" in *Proc. NOSSDAV*, 2012, pp. 9–14.
- [5] Z. Li, X. Zhu, J. Gahn, R. Pan, H. Hu, A. C. Begen, and D. Oran, "Probe and adapt: Rate adaptation for HTTP video streaming at scale," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 4, pp. 719–733, Apr. 2014.
- [6] *Openflow Switch Specification Version 1.5.0*. Accessed: Jan. 2020. [Online]. Available: <https://www.opennetworking.org>
- [7] A. Tootoonchian, M. Ghobadi, and Y. Ganjali, "OpenTM: Traffic matrix estimator for OpenFlow networks," in *Proc. Int. Conf. Passive Act. Netw. Meas.*, vol. 2010, pp. 201–210.
- [8] M. Yu, L. Jose, and R. Miao, "Software defined traffic measurement with open sketch," in *Proc. NSDI*, 2013, pp. 29–42.
- [9] P. Georgopoulos, Y. Elkhatib, M. Broadbent, M. Mu, and N. Race, "Towards network-wide QoE fairness using openflow-assisted adaptive video streaming," in *Proc. ACM SIGCOMM Workshop Future Hum.-Centric Multimedia Netw.*, Aug. 2013, pp. 15–20.
- [10] G. Cofano, L. De Cicco, T. Zinner, A. Nguyen-Ngoc, P. Tran-Gia, and S. Mascolo, "Design and experimental evaluation of network-assisted strategies for HTTP adaptive streaming," in *Proc. 7th Int. Conf. Multimedia Syst.*, May 2016, pp. 1–12.
- [11] W. Zhang, J. Tang, C. Wang, and S. de Soysa, "Reliable adaptive multipath provisioning with bandwidth and differential delay constraints," in *Proc. IEEE INFOCOM*, Mar. 2010, pp. 1–9.
- [12] B. Han, F. Qian, L. Ji, and V. Gopalakrishnan, "MP-DASH: Adaptive video streaming over preference-aware multipath," in *Proc. 12th Int. Conf. Emerg. Netw. Exp. Technol.*, Dec. 2016, pp. 129–143.

- [13] Y. Wang, J. Bi, and J. Wu, "AIDR: Aggregation of BGP routing table with AS path stretch," in *Proc. 19th IEEE Int. Conf. Netw. Protocols*, Oct. 2011, pp. 137–138.
- [14] J.-P. Sheu, W.-T. Lin, and G.-Y. Chang, "Efficient TCAM rules distribution algorithms in software-defined networking," *IEEE Trans. Netw. Service Manage.*, vol. 15, no. 2, pp. 854–865, Jun. 2018.
- [15] Y.-C. Wang, Y.-D. Lin, and G.-Y. Chang, "SDN-based dynamic multipath forwarding for inter-data center networking," in *Proc. IEEE Int. Symp. Local Metrop. Area Netw. (LANMAN)*, Jun. 2017, pp. 1–3.
- [16] S. Deb and R. Srikant, "Congestion control for fair resource allocation in networks with multicast flows," *IEEE/ACM Trans. Netw.*, vol. 12, no. 2, pp. 274–285, Apr. 2004.
- [17] S. Akhshabi, L. Anantkrishnan, C. Dovrolis, and A. C. Begen, "Server-based traffic shaping for stabilizing oscillating adaptive streaming players," in *Proc. 23rd ACM Workshop Netw. Oper. Syst. Support Digit. Audio Video*, 2013, pp. 19–24.
- [18] R. K. P. Mok, X. Luo, E. W. W. Chan, and R. K. C. Chang, "QDASH: A QoE-aware DASH system," in *Proc. 3rd Multimedia Syst. Conf.*, 2012, pp. 11–22.
- [19] J. W. Kleinrouweler, S. Cabrero, and P. Cesar, "Delivering stable high-quality video: An SDN architecture with DASH assisting network elements," in *Proc. 7th Int. Conf. Multimedia Syst.*, May 2016, pp. 1–10.
- [20] J. Wu, C. Yuen, B. Cheng, M. Wang, and J. Chen, "Streaming high-quality mobile video with multipath TCP in heterogeneous wireless networks," *IEEE Trans. Mobile Comput.*, vol. 15, no. 9, pp. 2345–2361, Sep. 2016.
- [21] J.-P. Sheu, C.-W. Chang, and Y.-C. Chang, "Efficient multicast algorithms for scalable video coding in software-defined networking," in *Proc. IEEE 26th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Aug. 2015, pp. 455–468.
- [22] M. Doshi and A. Kamdar, "Multi-constraint QoS disjoint multipath routing in SDN," in *Proc. Moscow Workshop Electron. Netw. Technol. (MWENT)*, Mar. 2018, pp. 1–5.
- [23] J. Dai, Z. Hu, B. Li, J. Liu, and B. Li, "Collaborative hierarchical caching with dynamic request routing for massive content distribution," in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 2444–2452.
- [24] X. Zhang, Y. Wang, J. Zhang, L. Wang, and Y. Zhao, "RINGLM: A link-level packet loss monitoring solution for software-defined networks," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 8, pp. 1703–1720, Aug. 2019.
- [25] *OpenDaylight*. Accessed: Feb. 2020. [Online]. Available: <https://www.opendaylight.org>
- [26] *OpenFlow Switch Specification Version 1.3.4*. Accessed: Feb. 2020. [Online]. Available: <https://www.opennetworking.org>
- [27] J. Moy, *OSPF Version 2*, document RFC 2328, IETF, 1998.
- [28] M. Ghobadi, Y. Cheng, A. Jain, and M. Mathis, "Trickle: Rate limiting YouTube video streaming," in *Proc. USENIX Conf. Annu. Tech. Conf.*, 2012, p. 17.
- [29] H.-H. Chu and K. Nahrstedt, "Dynamic multi-path communication for video traffic," in *Proc. 13th Hawaii Int. Conf. Syst. Sci.*, Jan. 1997, pp. 695–704.
- [30] A. Elgabli, V. Aggarwal, S. Hao, F. Qian, and S. Sen, "LBP: Robust rate adaptation algorithm for SVC video streaming," *IEEE/ACM Trans. Netw.*, vol. 26, no. 4, pp. 1633–1645, Aug. 2018.
- [31] P. Reichl, S. Egger, R. Schatz, and A. D'Alconzo, "The logarithmic nature of QoE and the role of the Weber-Fechner law in QoE assessment," in *Proc. IEEE Int. Conf. Commun.*, May 2010, pp. 1–5.
- [32] R. Mahindra, H. Viswanathan, K. Sundaresan, M. Y. Arslan, and S. Rangarajan, "A practical traffic management system for integrated LTE-WiFi networks," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2014, pp. 189–200.
- [33] M. Zhao, B. Jia, M. Wu, H. Yu, and Y. Xu, "Software defined network-enabled multicast for multi-party video conferencing systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2014, pp. 1729–1735.
- [34] Q.-V. Pham and W.-J. Hwang, "Network utility maximization-based congestion control over wireless networks: A survey and potential directives," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 1173–1200, 2nd Quart., 2017.
- [35] J.-W. Lee, M. Chiang, and A. R. Calderbank, "Price-based distributed algorithms for rate-reliability tradeoff in network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 5, pp. 962–976, May 2006.
- [36] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge Univ. Press, 2004.
- [37] *Internet Topology Zoo*. Accessed: May 2020. [Online]. Available: <http://www.topology-zoo.org/dataset.html>



JIACHEN WANG received the M.S. degree from Shandong Normal University, China, in 2008. He is currently pursuing the Ph.D. degree with Panpacific University, Philippines. He is currently a Professor with the School of Information Engineering, Shandong Yingcai University, where he is also the Vice Dean of the School of Information Engineering. He has more than ten articles in research journals and conferences. His research interests include networking, IoT systems, and data processing.



HONGLIN XIE received the M.S. degree from the School of Software Engineering, Shandong University, China, in 2009. He is currently a Professor with Shandong Yingcai University, where he is also the Director of the Information Engineering Experimental Teaching Center. He has more than ten articles in research journals and conferences, and has developed several big data analysis systems. His research interests include big data analysis and networking.



YONGHUI HE was born in January 1981. He received the master's degree in computer science and technology from Shandong University, China. He is currently an Associate Professor with the School of Information Engineering, Shandong Yingcai University, China. His current research interests include information security and image processing, networking, edge computing, and big data.



FENG ZONG received the M.S. degree from Guangxi University, in 2006. He is currently pursuing the Ph.D. degree with Panpacific University, Philippines. He is also an Associate Professor with Shandong Yingcai University, where he is also focusing on the application of the Internet of Things and cyberspace security. He has published more than ten articles in research journals and conferences. He has presided more than a number of research projects, mainly involving the research of security technology of some big data platforms, and the development and application of some IoT systems.



ZHIHONG WANG received the M.S. degree from Shandong University, China, in 2007. She is currently an Associate Professor with Shandong Agriculture and Engineering University. She has more than ten articles in research journals and conferences. Her research interests include information hiding, big data, and reinforcement learning.

...