

Rate Adaptation and Admission Control for Video Transmission with Subjective Quality Constraints

Chao Chen, *Student Member, IEEE*, Xiaoqing Zhu, *Member, IEEE*,
 Gustavo de Veciana, *Fellow, IEEE*, Alan C. Bovik, *Fellow, IEEE*,
 Robert W. Heath Jr., *Fellow, IEEE*

Abstract

Adapting video data rate during streaming can effectively reduce the risk of playback interruptions caused by channel throughput fluctuations. The variations in rate, however, also introduce video quality fluctuations and thus potentially affects viewers' Quality of Experience (QoE). We show how the QoE of video users can be improved by rate adaptation and admission control. We conducted a subjective study wherein we found that viewers' QoE was strongly correlated with the empirical cumulative distribution function (eCDF) of the predicted video quality. Based on this observation, we propose a rate-adaptation algorithm that can incorporate QoE constraints on the empirical cumulative quality distribution per user. We then propose a threshold-based admission control policy to block users whose empirical cumulative quality distribution is not likely to satisfy their QoE constraint. We further devise an online adaptation algorithm to automatically optimize the threshold. Extensive simulation results show that the proposed scheme can reduce network resource consumption by 40% over conventional average-quality maximized rate-adaptation algorithms.

Index Terms

Quality of Experience, Video transport, Rate adaptation, Admission control, Wireless networks.

I. INTRODUCTION

Video traffic is currently a rapidly growing fraction of the data traffic on wireless networks. As reported in [1], video traffic accounted for more than 50% of the mobile data traffic in 2012. Efficiently

Chao Chen, Gustavo de Veciana, Alan Bovik, and Robert Heath are with Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712-0240, USA (email: tochenchao@gmail.com; gustavo@ece.utexas.edu; bovik@ece.utexas.edu; rheath@ece.utexas.edu). Xiaoqing Zhu is with the Advanced Architecture and Research Group, Cisco Systems, San Jose, CA 95134, USA (email: xiaoqzhu@cisco.com).

This research was supported in part by Intel Inc. and Cisco Corp. under the VAWN program.

utilizing network resources to satisfy video users' expectations regarding their Quality of Experience (QoE) is an important research topic. In this paper, we study approaches to share wireless down-link resources among video users via QoE-based rate adaptation and admission control.

We focus on a setting in which stored video content is streamed over wireless networks. When a video is streamed, the received video data is first buffered at the receiver and then played out to the viewer. Because the throughput of a wireless channel generally varies over time, the amount of buffered video decreases when the channel throughput falls below the current video data rate. Once all the video data buffered at the receiver has been played out, the playback process stalls, significantly affecting the viewers' QoE [2]. To address this problem, various video rate-adaptation techniques based on scalable video coding or adaptive bitrate switching have been proposed to match the video data rate to the varying channel capacity [3]–[7]. Although these rate adaptation techniques can effectively reduce the risk of playback interruptions, the variable bitrate causes quality fluctuations, which, in turn, affect viewers' QoE. In most existing rate-adaptation algorithms such as [8]–[13], the average video quality is employed as the proxy for QoE. The average quality, however, does not reflect the impact of quality fluctuations on the QoE, i.e., two videos with the same average quality can have significantly different levels of quality fluctuation. In this paper, we propose to characterize and predict the users' QoE using the second order empirical cumulative distribution function (2^{nd} -order eCDF) of the delivered video quality; this is defined as

$$F^{(2)}(x; \mathbf{q}) = \frac{1}{T} \sum_{t=1}^T \max\{x - q(t), 0\}, \quad (1)$$

where $q(t)$ represents the predicted quality [14] of the t^{th} second of the video and T is the video length. Note that $\max\{x - q(t), 0\} > 0$ if and only if $q(t) < x$, so the 2^{nd} -order eCDF captures for how long and by how much the predicted video quality falls below x . If we interpret x as the threshold below which users judge the video quality to be unacceptable, then the 2^{nd} -order eCDF reflects the impact of the unacceptable periods on the QoE. Since it has been recognized that the worst parts of a video tend to dominate the overall quality of an entire video [15]–[19], the 2^{nd} -order eCDF can be used to predict the QoE.

The efficacy of the 2^{nd} -order eCDF in capturing QoE can be validated through subjective experiments. In [20] and [21], we reported a subjective study of this type. For each of the 15 quality-varying videos involved in the subjective study, we asked 25 subjects to score its quality. We computed the linear correlation coefficients (LCCs) between various QoE metrics and the subjects' Mean Opinion Scores (MOSs) in Table I. The 2^{nd} -order eCDF achieves the strongest linear correlation (0.84). In

comparison, the average video quality only achieves a correlation of 0.57. This lends strong support for eCDF as a good proxy for video QoE.

TABLE I

THE LINEAR CORRELATION COEFFICIENTS (LCCs) OF SEVERAL METRICS WITH QoE. THE METRICS $\text{mean}\{q\}$, $\text{min}\{q\}$ AND $\text{var}\{q\}$ ARE THE AVERAGE VALUE, THE MAXIMUM VALUE AND THE VARIANCE OF THE TIME SERIES $q(1), \dots, q(T)$, WHERE $q(t)$ IS THE PREDICTED VIDEO QUALITY AT TIME t .

QoE metrics	$\text{mean}\{q\}$	$\text{min}\{q\}$	$\text{mean}\{q\} + \mu\sqrt{\text{var}\{q\}}$	$F^{(2)}(x; q)$
LCC	0.5659	0.4022	0.7377	0.8446

In this paper, we design adaptive video streaming algorithms that incorporate QoE constraints on the 2nd-order eCDFs of the video qualities seen by users. In particular, we consider a wireless network in which a base station transmits videos to multiple users. The user population is dynamic, i.e., users arrive and depart from the network at random times. When a new user joins the network, the base station starts streaming a video to the user. A rate adaptation algorithm is employed to control the video data rate of all active video streams according to varying wireless channel conditions.

When the base station is shared by too many users simultaneously, the QoE served to each user can be poor. Instead of serving every user with poor QoE, it is preferable to satisfy the QoE constraints of existing users by selectively blocking newcomers. Therefore, in addition to rate adaptation algorithms, we introduce a new admission control strategy that is designed to maximize the number of video users satisfying the QoE constraints on their 2nd-order eCDFs. As will be shown in the paper, although the admission control strategy damages the QoE of the blocked users, the overall percentage of users satisfying the QoE constraints among both admitted and blocked users can be significantly improved. The contributions of our work are twofold:

- 1) *An online rate-adaptation algorithm aimed at meeting QoE constraints based on the 2nd-order eCDFs of users' video qualities.*

Since the 2nd-order eCDF is determined by the overall spatio-temporal pattern of video, simply maximizing the video quality all the times is not sufficient to satisfy the QoE constraints. Instead, we propose an online rate-adaptation algorithm that jointly considers the channel conditions, the video rate-quality characteristics, and the 2nd-order eCDFs of all video users. We show significant performance gains over conventional average-quality optimized algorithms.

- 2) *An admission control strategy, which blocks video users who will likely be unable to meet the constraints on their 2nd-order eCDFs.*

Specifically, we propose an algorithm that predicts the video quality experienced by each user who

is new to the network. We then employ a threshold-based admission control policy to block those users whose estimated qualities fall below the threshold. An online algorithm is proposed to adjust the threshold to its optimal value. The proposed admission control strategy further improves the performance of our rate-adaptation algorithm, especially when the network resources are limited. The remainder of this paper is organized as follows: Section II discusses related work. In Section III, we give an overview of the structure of the video streaming system studied in this paper. In Section IV, we explain our system model and the QoE constraints. The proposed rate adaptation algorithm and the admission control strategy are introduced in Section V and VI. We demonstrate their performance via extensive simulation results. Section VII concludes the paper with discussions on future work.

II. RELATED WORK

Let us first review related work in QoE assessment, rate adaptation, and admission control.

A. QoE Metrics for Rate-Adaptive Video Steaming:

Most existing rate-adaptation algorithms such as [8]–[13] employ average video quality as the proxy of QoE due to its simplicity in analysis. As shown in Table I, however, average quality does not efficiently capture the impact of quality fluctuations. To address this problem, temporal quality pooling has been studied [15]–[19]. The pooling algorithms, however, have only been designed and validated for short videos that are a few seconds long. Furthermore, they are too complicated to be incorporated into a tractable analytical framework for the design of rate-adaptation algorithms. The authors of [22] propose a simple QoE metric, which is the weighted sum of the time-average and the standard deviation of the predicted video quality. As shown in Table I, this metric achieves a better correlation of 0.73 than the time-average quality. The QoE metric that we propose here achieves an even higher correlation of 0.84. Moreover, because the 2nd-order eCDF is simply a temporal average of the function $\max\{x - q(t), 0\}$, its analysis is just as simple as for average video quality.

B. Rate Adaptive Video Streaming:

Extensive research efforts have been applied to the problem of rate adaptation for wireless video streaming. Most existing work employs the time-average video quality as the QoE metric due to its simplicity [8]–[13]. In [23], a rate adaptation algorithm is proposed to optimize the QoE metric of [22]. Although the proposed algorithm mitigates video quality variations, it assumes a fixed set of users and thus does not incorporate the impact of user arrival and departure on video quality. In [24], the

quality fluctuations caused by user arrival and departure are analyzed for large wireless links shared by many video streams. The algorithms proposed here do not rely on such assumptions and can be applied to small wireless cells such as Wi-Fi networks.

C. Admission Control for Video Streaming:

Admission control for variable bitrate videos has been studied in [25]–[27]. In [25], an admission control algorithm is proposed for variable bitrate videos stored on disk arrays and transmitted over cable television networks. In contrast to wireless networks, the transmission bottleneck of cable television networks is the buffer at the disk array. The admission control algorithm is designed to ensure that the buffer does not overflow while the video is played back continuously. In [26] and [27], threshold-based admission control algorithms for video streams delivered over packet-switch networks are studied. In [26], thresholds are applied to the number of video users. The heterogeneity in the data rate of the video streams is not considered. In [27], the threshold is applied to the aggregated data rate requested by admitted videos. In both [26] and [27], a statistical model of the video traffic is necessary to optimize the admission threshold. In practice, however, modeling the video traffic is difficult. Our proposed admission control strategy does not rely on the *a priori* knowledge of the traffic statistics. The threshold is learned online and optimized automatically.

III. SYSTEM OVERVIEW

We first discuss the architecture of the wireless networks considered in this paper. Then, we explain how the proposed online rate adaptation algorithm and threshold-based admission control strategy fit into the existing network architecture.

A. Architecture of the Wireless Network

We consider a wireless network where video users share the down-link with high-priority traffic (e.g., voice traffic) and thus video users will need to adapt their rates accordingly. (see Fig. 1). All users arrive to and depart from the network at random times. A high-priority user requires a random amount of wireless resource (e.g., transmission time in TDMA systems) per unit time throughout its sojourn in the system. The transmission resources not used by the high-priority users can be allocated to video users. When a video user arrives, it requests a video that is stored at a content delivery network (CDN) and is streamed to the user via the base station. When a video is being streamed, the video data is first delivered to a receive buffer and then decoded for display. Paralleling prior work such as [13] [23] [28], we assume a proxy is colocated with the base station. We treat the high-priority users as background traffic, and the proxy is only used to control the video streams.

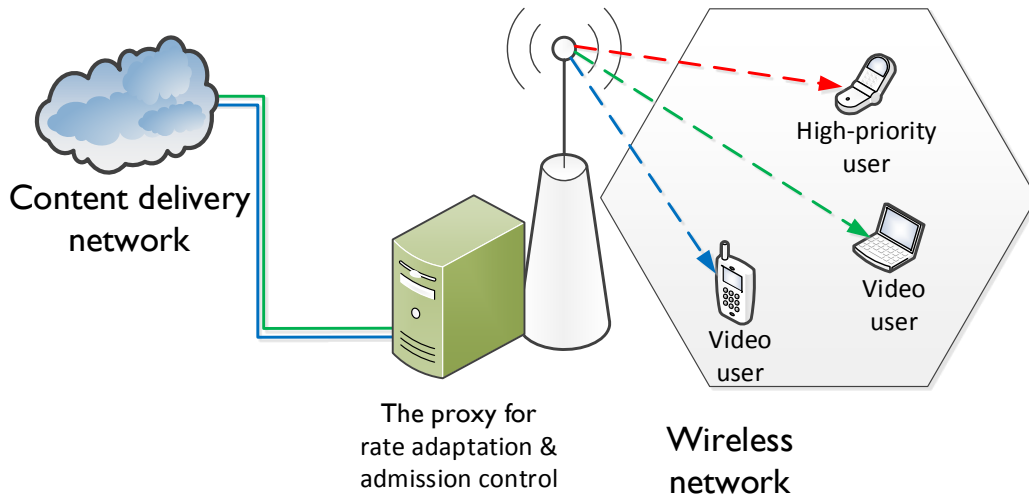


Fig. 1. The wireless network considered here: The video is stored at the content delivery network. The proxy for rate adaptation and admission control is colocated with the base station. The base station serves both video users and high-priority users.

B. Video Streaming Proxy

The function of the proxy is twofold (see Fig. 2). First, when a video user arrives, the proxy decides whether the user should be admitted to share the channel or not. Second, for admitted video users, the proxy adapts the transmission data rates according to the varying channel conditions. The admission control strategy only acts on newly arrived video users, and the rate adaptation algorithm does its best to satisfy QoE constraints for all admitted video users. The rate adaptation algorithm accompanies the admission control algorithm by feeding back necessary information (see Section V-B for more detail) regarding the current status of the admitted users.

We assume the proxy operates in a time-slotted manner where the duration of each slot is ΔT seconds. The admission control and the rate adaptation actions are conducted at the beginning of every time slot. Note that, in a video stream, the video frames are partitioned in Groups of Pictures (GoPs) and the video data rate can only be adapted at the boundary of GoPs [29] [30]. Because the duration of a GoP is usually configured to be larger than one second to achieve high compression efficiency, we assume ΔT is at least 1 second.

IV. SYSTEM MODEL

Before proceeding further, we introduce the notation used in the paper. Then, we describe the channel model and the video rate-quality model. At the end of this section, we explain the QoE constraints considered in our problem formulation.

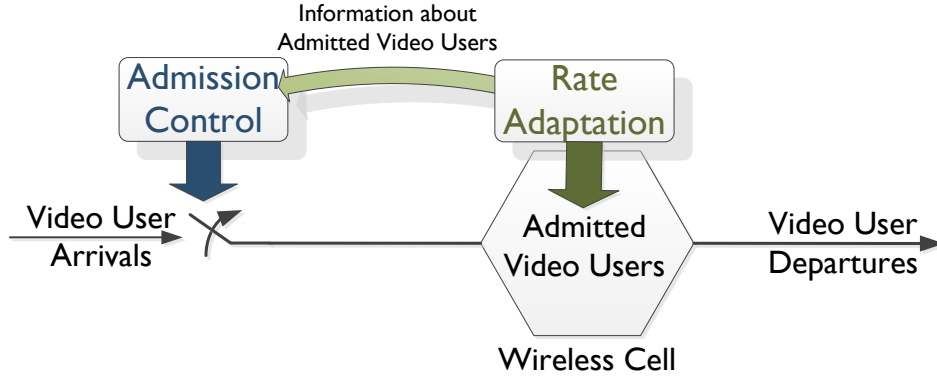


Fig. 2. The proposed QoE-constrained video streaming system: Admission control only affects newly arrived video users and the rate adaptation algorithm does its best to satisfy the QoE constraints of all admitted video users.

A. Notation

In the rest of the paper, the time slots are indexed with $t = 1, 2, \dots$. The notation $(x(t), t = 1, 2, \dots)$ denote a discrete time series. Lower-case symbols such as a denote scalar variables and boldface symbols such as \mathbf{a} denote vectors. Random variables are denoted by uppercase letters such as A . Calligraphic symbols such as \mathcal{A} denote sets, while $|\mathcal{A}|$ is the cardinality of \mathcal{A} . The set of positive integers is denoted \mathbb{N}^+ . The set of real numbers is denoted \mathbb{R} . Finally, the function $[x]^+ = \max\{x, 0\}$.

B. Wireless Channel Model

We label users (including both video users and high-priority users) according to their arrival times, i.e., user u is the u^{th} user to arrive to the network. For each user u , we let A_u and D_u be random variables denoting the arrival and departure times, respectively. The time spent by a user in the network is denoted by $T_u = D_u - A_u + 1$.

We let by $\mathcal{U}^p(t)$ be the set of high-priority users in the network at slot t . For each high-priority user $u \in \mathcal{U}^p(t)$, we let the random variable $W_u(t)$ represent the amount of data received in slot t . The data rate is thus $R_u(t) = W_u(t)/\Delta T$. We denote by $\mathcal{U}^v(t)$ the set of video users that would be in the network at slot t if all were admitted. The set of video users in $\mathcal{U}^v(t)$ who are actually admitted to share the wireless channel is denoted by $\mathcal{U}^{\text{av}}(t)$. We assume that an admission decision is made upon the arrival of each video user. Once admitted, the video user shares the channel until it leaves the network. For an admitted video user $u \in \mathcal{U}^{\text{av}}(t)$, we denote by $w_u(t)$ the amount of received video data in slot t . The video data rate delivered to the user is thus $r_u(t) = w_u(t)/\Delta T$. We call

$\mathbf{r}(t) = (r_u(t) : u \in \mathcal{U}^{\text{av}}(t))$ the *video rate vector* at time t . Because high-priority users are treated as background traffic, the proxy only controls the video rate vector and regards $(R_u(t) : u \in \mathcal{U}^{\text{p}}(t))$ as exogenous variables.

We assume the set of video rate vectors that can be supported by the wireless channel is given by

$$\mathcal{C}(t) = \{\mathbf{r} : C_t(\mathbf{r}) \leq 0\}, \quad (2)$$

where $C_t : \mathbb{R}^{|\mathcal{U}^{\text{av}}(t)|} \rightarrow \mathbb{R}$ is a time-varying multivariate convex function. The specific form of $C_t(\cdot)$ depends on the multiuser multiplexing techniques used in the wireless network. For example, in a time-division multiple access (TDMA) system [13] [23], the channel is occupied by a single user at any moment. Denote by $P_u(t)$ the peak transmission rate of user u , i.e., the transmission rate at which user u can be served during slot t . Then, in slot t , video user $u \in \mathcal{U}^{\text{av}}(t)$ spends $\frac{w_u(t)}{P_u(t)}$ seconds to download video. Similarly, each high-priority user $u' \in \mathcal{U}^{\text{p}}(t)$ spends $\frac{W_{u'}(t)}{P_{u'}(t)}$ seconds downloading data. Since the total transmission across all users is less than ΔT , we have $\sum_{u \in \mathcal{U}^{\text{av}}(t)} \frac{w_u(t)}{P_u(t)} + \sum_{u' \in \mathcal{U}^{\text{p}}(t)} \frac{W_{u'}(t)}{P_{u'}(t)} \leq \Delta T$. Because $r_u(t) = w_u(t)/\Delta T$ and $R_{u'}(t) = W_{u'}(t)/\Delta T$, we have $\sum_{u \in \mathcal{U}^{\text{av}}(t)} \frac{r_u(t)}{P_u(t)} + \sum_{u' \in \mathcal{U}^{\text{p}}(t)} \frac{R_{u'}(t)}{P_{u'}(t)} \leq 1$. Therefore, for TDMA systems, we have $C_t(\mathbf{r}(t)) = \sum_{u \in \mathcal{U}^{\text{av}}(t)} \frac{r_u(t)}{P_u(t)} + \sum_{u' \in \mathcal{U}^{\text{p}}(t)} \frac{R_{u'}(t)}{P_{u'}(t)} - 1$.

In rate-adaptive video streaming systems such as [5]–[7], the video data rate can only take values in a finite and discrete set. In our problem formulation, we relax the constraint and allow $r_u(t)$ to take values in a compact interval $[r_u^{\min}(t), r_u^{\max}(t)]$, where $r_u^{\min}(t)$ and $r_u^{\max}(t)$ denote the minimum and maximum data rate available for user u , respectively. Letting $\mathcal{R}(t) = \prod_{u \in \mathcal{U}^{\text{av}}(t)} [r_u^{\min}(t), r_u^{\max}(t)]$, we have

$$\mathbf{r}(t) \in \mathcal{R}(t). \quad (3)$$

In our algorithm implementation, we round up the optimal video data rate obtained under this relaxed constraint to the nearest available data rate.

C. Video Rate-Quality Model

We assume that the quality of the video downloaded in each slot is represented by a Difference Mean Opinion Score (DMOS) [31], which ranges from 0 to 100 where lower values indicate better quality. To represent video quality more naturally, so that higher numbers indicate better video quality, we deploy a Reversed DMOS (RDMOS). Denote by $q_u^{\text{dmos}}(t)$ the DMOS of the video delivered to user u at slot t , the RDMOS is given by $q_u^{\text{rdmos}}(t) = 100 - q_u^{\text{dmos}}(t)$. We employ the following rate-quality model to predict $q_u^{\text{rdmos}}(t)$ using the video data rate $r_u(t)$:

$$q_u(t) = \alpha_u(t) \log(r_u(t)) + \beta_u(t), \quad (4)$$

where $q_u(t)$ is the predicted RDMOS. The model parameters $\alpha_u(t)$ and $\beta_u(t)$ can be determined by minimizing the prediction error between $q_u^{\text{rdmos}}(t)$ and $q_u(t)$. For stored video streaming, the video file is stored at the CDN. Thus, the model parameters in (4) can be obtained before transmission. Here, we assume the parameters $\alpha_u(t)$ and $\beta_u(t)$ are known *a priori*.

We validated this model on a video database that includes twenty-five different pristine and representative videos [20]. The rest of the database is created by encoding and decoding each video at different rates with the widely used H.264 codec [32]. Then, we predict the RDMOSs for each second of the decoded videos using the high-accuracy video-RRED index [33]. In Fig. 3, we show the rate-RDMOS mapping of one second of a video randomly chosen from the database. It may be observed that the model (4) can accurately predict the RDMOSs. On the whole database, the mean prediction error of (4) is less than 1.5, which is visually negligible. Thus, in the following, we shall refer to $q_u(t)$ as the video quality.

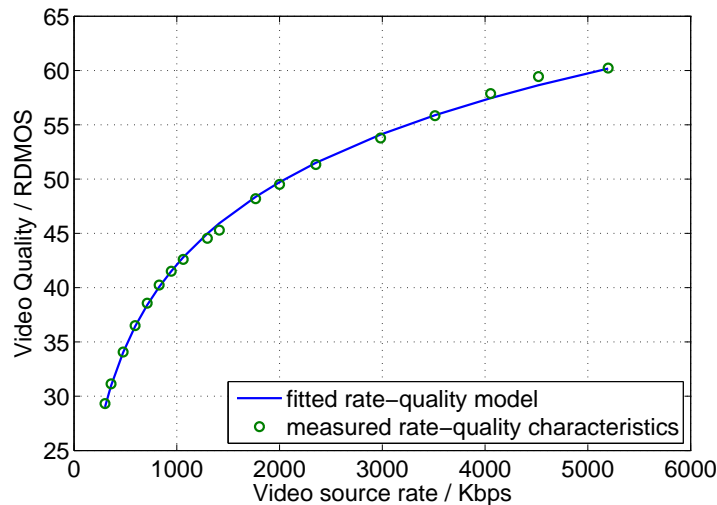


Fig. 3. The performance of the rate-quality model on one second of a video randomly chosen from the database [20]. The rate-quality characteristics are shown in circles while the fitted rate-quality model (4) is the solid line.

D. Constraints on the Quality of Experience

We capture video users' QoE using the 2th-order eCDF $F^{(2)}(x; q)$, which was defined in (1). As illustrated in Fig. 4(a), for a given x , the right-hand side of (1) is proportional to the area where $q(t)$ falls below x . If $q(t)$ falls below x for a long while, the QoE is poor and $F^{(2)}(x; q)$ is large. Otherwise, the QoE is good and $F^{(2)}(x; q)$ is small.

To justify the use of the $F^{(2)}(x; q)$ as the QoE metric, we conducted a subjective study following the guidelines of [31]. The study involved twenty-five subjects and fifteen quality-varying long videos (for more details, see [20] and [21]). Based on the subjects' feedback, we obtained the Mean Opinion Scores (MOSs) of each video's overall quality. Given an x , we computed $F^{(2)}(x; q)$ for all the videos in the database and then calculated the linear correlation coefficient (LCC) between the computed $F^{(2)}(x; q)$ s and the MOSs. In Fig 4(b), we plot the absolute value of the LCCs as a function of x . We found that, at $x^* = 37$, $F^{(2)}(x^*; q)$ achieves a strong correlation of 0.84 with the MOSs. Since $F^{(2)}(x^*; q)$ is determined by the area where $q(t)$ falls below x^* , we interpret x^* as the users' *video quality expectation*, which is used by the users as a threshold in judging whether the video quality is acceptable or not. In our subjective study, all subjects viewed the videos in a controlled environment and every subject viewed the videos on the same device. Broadly speaking, the video quality expectation x^* can be environment-dependent. For example, viewers tend to have higher expectation for videos shown on a laptop than videos shown on a smartphone. Therefore, in a practical wireless network, x^* may vary across users. In the following, we denote by x_u^* the video quality expectation of user u and

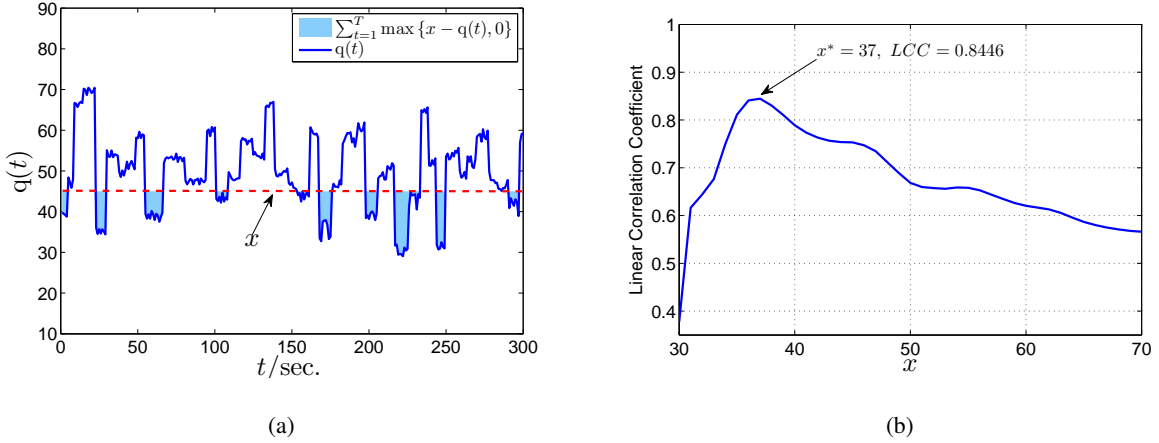


Fig. 4. (a) An example of $F^{(2)}(x; q)$ at $x = 45$; (b) The absolute value of the linear correlation coefficient (LCC) between $F^{(2)}(x; q)$ and the mean opinion scores.

study the following two cases:

Case I: Users' video quality expectation is unavailable. If user u is in the system from A_u to D_u and sees video qualities $(q_u(t) : t \in [A_u, D_u])$, according to (1), its 2nd-order eCDF is given by

$$F^{(2)}(x; q_u) = \frac{1}{T_u} \sum_{t=A_u}^{D_u} [x - q_u(t)]^+ . \quad (5)$$

If the users' video quality expectation x_u^* is not known *a priori*, we may impose constraints on all x

and for all video users. In particular, we consider the following QoE constraints:

$$F^{(2)}(x; q_u) \leq h(x), \quad \forall x \in [0, 100], \quad \forall u \in \cup_{t=1}^{\infty} \mathcal{U}^{\text{av}}(t), \quad (6)$$

where $h(x)$ is a function of x . In practice, we cannot apply constraints on all values of $x \in [0, 100]$. Therefore, we consider a relaxed version of (6) as follows:

$$F^{(2)}(x_i; q_u) \leq h(x_i), \quad \forall x_i \in \mathcal{I}, \quad \forall u \in \cup_{t=1}^{\infty} \mathcal{U}^{\text{av}}(t). \quad (7)$$

Here, \mathcal{I} is a discrete set of points on $[0, 100]$. The following property of 2nd-order eCDFs shows that (7) will approximate (6) if \mathcal{I} is dense. Its proof is given in Appendix A.

Theorem 1. *Let $\bar{h}(x)$ be the piece-wise linear function that connects the points $\{(x_i, h(x_i)) : \forall x_i \in \mathcal{I}\}$. The constraint (7) is equivalent to $F^{(2)}(x; q_u) \leq \bar{h}(x)$, $\forall x \in [0, 100]$.*

Case II: Users' video quality expectation is known. We also consider the case where users' video quality expectation x_u^* is known or specified by the service provider. For example, users with different viewing devices tend to have different quality expectations. Desktop users usually watch high-definition television programs on large screens and smartphone users usually watch low resolution videos. Thus, we may conduct subjective studies on different devices. Based on the results of the studies, the users' typical video quality expectations x_u^* on each type of device can be deduced. Then, we can categorize video users according to their respective devices and provide differentiated QoE guarantees.

We define a finite set $\mathcal{G} = \{g_1, \dots, g_{|\mathcal{G}|}\}$ that represents different video quality expectations and assume $x_u^* \in \mathcal{G}$, $\forall u \in \cup_{t=1}^{\infty} \mathcal{U}^{\text{av}}(t)$. Let $\mathcal{U}_j^{\text{av}}(t) = \{u \in \mathcal{U}^{\text{av}}(t) : x_u^* = g_j\}$ denote the video users whose quality expectation is g_j . We consider the following constraints:

$$F^{(2)}(g_j; q_u) \leq h_j, \quad \forall g_j \in \mathcal{G}, \quad \forall u \in \cup_{t=1}^{\infty} \mathcal{U}_j^{\text{av}}(t), \quad (8)$$

where h_j is the QoE constraint for users with quality expectation g_j .

In sum, the goal of the admission control strategy and the rate adaptation algorithm is to maximize the number of users satisfying constraints (2)-(4), (7), or (8). In Section V, we introduce our rate adaptation algorithm and the admission control strategy when users' quality expectations are unknown. Then, in Section VI, we extend our rate adaptation and admission control algorithms to the case where each user's video quality expectation is known. For ease of reading, a summary of the notation used in this paper is given in Table II.

TABLE II
NOTATION SUMMARY

A_u, D_u	The arrival and departure time of user u
T_u	Time spent by user u in the network
$\mathcal{U}^v(t)$	Video users at time t ;
$\mathcal{U}^{\text{av}}(t)$	Admitted video users at time t ;
$r_u(t)$	Data rate of user u at t
$\mathcal{C}(t)$	Rate region at t
$\mathcal{R}(t)$	The set of available source rates at t
$q_u(t)$	Video quality of user u at t
$F^{(2)}(x; q_u)$	The second order eCDF of user u
x_u^*	Video quality expectation of user u
\mathcal{I}	Constrained points on eCDFs
$h(x_i)$	The QoE constraint at $x_i \in \mathcal{I}$
\mathcal{G}	Quality expectations of video users
$\mathcal{U}_j^{\text{av}}$	Admitted video users with $x_u^* = g_j \in \mathcal{G}$
h_j	The QoE constraint for the users in $\mathcal{U}_j^{\text{av}}$

V. RATE ADAPTATION AND ADMISSION CONTROL WITH UNKNOWN VIDEO QUALITY EXPECTATION

If the quality expectations of video users are not known, we apply the same constraint on the second-order eCDFs of all video users. We first propose a rate adaptation algorithm and a corresponding admission control strategy. Then, we evaluate their performance via numerical simulation.

A. Rate Adaptation Algorithm

To clarify the design of our rate adaptation algorithm, we present an off-line problem formulation in which the future channel conditions and admission decisions are assumed to be known. Then, based on the analysis of this offline problem, we propose a new on-line rate adaptation algorithm.

If we consider a finite horizon T and assume that the realizations of channel conditions $\mathcal{C}(1), \dots, \mathcal{C}(T)$ are known, the rate adaptation algorithm should solve the following feasibility problem:

$$\text{find } \mathbf{r}_{1:T} \tag{9a}$$

$$\text{subject to: } \mathbf{r}(t) \in \mathcal{C}(t) \cap \mathcal{R}(t), \forall 1 \leq t \leq T \tag{9b}$$

$$q_u(t) = \alpha_u(t) \log(r_u(t)) + \beta_u(t), \forall u \in \mathcal{U}^{\text{av}}(t), \forall 1 \leq t \leq T \tag{9c}$$

$$F^{(2)}(x_i; q_u) \leq h(x_i), \forall x_i \in \mathcal{I}, \forall u \in \cup_{t=1}^T \mathcal{U}^{\text{av}}(t). \tag{9d}$$

The constraint (9b) is associated with the achievable rate region (2) and the available video source rates in (3). The constraint (9c) is because of the rate-quality model (4). The constraints (9d) are the QoE constraints (7) that were discussed in Section IV-D. For each admitted user, a series of QoE constraints are applied to the 2nd-order eCDF at discrete points in $\mathcal{I} = \{x_1, \dots, x_{|\mathcal{I}|}\}$. Since the rate-quality function (9c) is concave, according to [34], the problem (9) is equivalent to the following convex optimization problem:

$$\begin{aligned} & \underset{\mathbf{r}_{1:T}, (\hat{q}_u)_{1:T}}{\text{maximize}} && 0 \end{aligned} \tag{10a}$$

$$\text{subject to: } \mathbf{r}(t) \in \mathcal{C}(t) \cap \mathcal{R}(t), \forall 1 \leq t \leq T \tag{10b}$$

$$\hat{q}_u(t) \leq \alpha_u(t) \log(r_u(t)) + \beta_u(t), \forall u \in \mathcal{U}^{\text{av}}(t), \forall 1 \leq t \leq T \tag{10c}$$

$$F^{(2)}(x_i; \hat{q}_u) \leq h(x_i), \forall x_i \in \mathcal{I}, \forall u \in \cup_{t=1}^T \mathcal{U}^{\text{av}}(t), \tag{10d}$$

where $\hat{q}_u(t)$ are virtual variables introduced here to make the constraint (10c) convex. Note that the right-hand side of constraint (10c) equals $q_u(t)$. For any $\hat{q}_u(t)$ satisfying (10c), we have $\hat{q}_u(t) \leq q_u(t)$. Hence, if $\hat{q}_u(t)$ satisfies (10d), the constraint $F^{(2)}(x_i; q_u) \leq h(x_i)$ is satisfied as well.

By the definition in (5), the 2nd-order eCDF in constraint (10d) is determined by the entire process $\hat{q}_u(\mathbf{A}_u), \dots, \hat{q}_u(\mathbf{D}_u)$. Due to the constraints (10b) and (10c), $\hat{q}_u(t)$ depends on the rate $r_u(t)$ and thus also depends on the rate region $\mathcal{C}(t)$. Therefore, the solution of (10) depends on the entire process $\mathcal{C}(1), \dots, \mathcal{C}(T)$. In practice, the future channel conditions are unavailable to the rate-adaptation algorithm. In the following, we transform problem (10) to a simpler form that inspires our online rate adaptation algorithm.

Since (10) is a convex problem, if it is feasible, there exists a set of Lagrange multipliers $\lambda_{u,i}^* \geq 0$ for the constraints in (10d) such that a solution of (10) can be obtained by solving the following problem (see [35]):

$$\begin{aligned} & \underset{\mathbf{r}_{1:T}, (\hat{q}_u)_{1:T}}{\text{minimize}} && \sum_{u \in \cup_{t=1}^T \mathcal{U}^{\text{av}}(t)} \sum_{x_i \in \mathcal{I}} \lambda_{u,i}^* (F^{(2)}(x_i; \hat{q}_u) - h(x_i)) \end{aligned} \tag{11a}$$

$$\text{subject to: } \mathbf{r}(t) \in \mathcal{C}(t) \cap \mathcal{R}(t), \forall 1 \leq t \leq T \tag{11b}$$

$$\hat{q}_u(t) \leq \alpha_u(t) \log(r_u(t)) + \beta_u(t), \forall u \in \mathcal{U}^{\text{av}}(t), \forall 1 \leq t \leq T. \tag{11c}$$

If we define a function $s_{u,i}(t)$ as

$$s_{u,i}(t) = \begin{cases} \frac{1}{T_u} ([x_i - \hat{q}_u(t)]^+ - h(x_i)) & \text{if } \mathbf{A}_u \leq t \leq \mathbf{D}_u, \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

the term $F^{(2)}(x_i; \hat{q}_u) - h(x_i)$ in (11a) can be rewritten as

$$\begin{aligned} & F^{(2)}(x_i; \hat{q}_u) - h(x_i) \\ &= \sum_{t=A_u}^{D_u} \frac{1}{T_u} ([x_i - \hat{q}_u(t)]^+ - h(x_i)) \\ &= \sum_{t=1}^T s_{u,i}(t). \end{aligned} \quad (13)$$

Thus, $s_{u,i}(t)$ indicates to what extent the variable $\hat{q}_u(t)$ violates the constraint $F^{(2)}(x_i; \hat{q}_u) \leq h(x_i)$ in each slot. Substituting (13) in (11a) and changing the order of summation, the optimization in (11) becomes

$$\begin{aligned} & \underset{\mathbf{r}_{1:T}, (\hat{q}_u)_{1:T}}{\text{minimize}} \quad \sum_{t=1}^T \left(\sum_{u \in \mathcal{U}^{\text{av}}(t)} \sum_{x_i \in \mathcal{I}} \lambda_{u,i}^* s_{u,i}(t) \right) \\ & \text{subject to:} \quad \mathbf{r}(t) \in \mathcal{C}(t) \cap \mathcal{R}(t), \quad \forall 1 \leq t \leq T \\ & \quad \hat{q}_u(t) \leq \alpha_u(t) \log(\mathbf{r}_u(t)) + \beta_u(t), \quad \forall u \in \mathcal{U}^{\text{av}}(t), \quad \forall 1 \leq t \leq T. \end{aligned} \quad (14)$$

Note that, except for the Lagrange multipliers, the optimization in (14) does not involve variables that depend on the entire process $\{\hat{q}_u(t) : 1 \leq t \leq T\}$. Thus, (14) can be solved by minimizing the weighted sum $\sum_{u \in \mathcal{U}^{\text{av}}(t)} \sum_{x_i \in \mathcal{I}} \lambda_{u,i}^* s_{u,i}(t)$ in every slot. That is, if it is possible to estimate the Lagrange multiplier $\lambda_{u,i}^*$, then (10) can be solved by greedily choosing the rate vector $\mathbf{r}(t)$ at each time slot as the solution of the following problem:

$$\begin{aligned} & \underset{\mathbf{r}(t), \hat{q}_u(t)}{\text{minimize}} \quad \sum_{u \in \mathcal{U}^{\text{av}}(t)} \sum_{x_i \in \mathcal{I}} \lambda_{u,i}^* s_{u,i}(t) \\ & \text{subject to:} \quad \mathbf{r}(t) \in \mathcal{C}(t) \cap \mathcal{R}(t) \\ & \quad \hat{q}_u(t) \leq \alpha_u(t) \log(\mathbf{r}_u(t)) + \beta_u(t), \quad \forall u \in \mathcal{U}^{\text{av}}(t). \end{aligned} \quad (15)$$

We introduce a method to approximate the Lagrange multiplier $\lambda_{u,i}^*$. We know that the Lagrange multiplier $\lambda_{u,i}^*$ indicates the difficulty in satisfying the constraint $F_u^{(2)}(x_i; \hat{q}_u) \leq h(x_i)$ [34]. Inspired by prior work in [36] and [37], we employ a virtual queue to capture this difficulty. For each admitted user u and each $x_i \in \mathcal{I}_u$, define the virtual queue as

$$v_{u,i}(t) = \begin{cases} [v_{u,i}(t-1) + s_{u,i}(t)]^+ & \text{if } A_u \leq t \leq D_u, \\ 0 & \text{otherwise.} \end{cases} \quad (16)$$

From (13) it follows that, if the summation of $s_{u,i}(t)$ is large, then it is difficult to satisfy the constraint $F_u^{(2)}(x_i; \hat{q}_u) \leq h(x_i)$. The virtual queue captures the cumulative summation of $s_{u,i}(t)$ up to slot t . Hence, the virtual queue reflects the level of difficulty in satisfying $F_u^{(2)}(x_i; \hat{q}_u) \leq h(x_i)$. Actually, for the special case where user set $\mathcal{U}^{\text{av}}(t)$ is fixed for all t , it can be proved that the virtual queue asymptotically approaches $\lambda_{u,i}^*$ as $T \rightarrow \infty$ [36]. Hence, we replace the Lagrange multipliers in (15)

with virtual queue $v_{u,i}(t)$ and our online rate adaptation algorithm is summarized in Algorithm 1. In every slot, we maximize the weighted sum of $s_{u,i}(t)$, where the weight is given by $v_{u,i}(t-1)$. Thus users with larger virtual queues tend to be allocated more network resources. This helps users satisfy their QoE constraints.

Next, we introduce an admission control policy that is combined with our rate-adaptation algorithm to further improve performance.

Algorithm 1 Online algorithm for video data rate adaptation

1: **for** $t = 1 \rightarrow \infty$ **do**

2: Choose rate vector $\mathbf{r}(t)$ that solves the problem

$$\begin{aligned} & \underset{\mathbf{r}(t), \hat{q}_u(t)}{\text{minimize}} && \sum_{u \in \mathcal{U}^{\text{av}}(t)} \sum_{x_i \in \mathcal{I}} v_{u,i}(t-1) s_{u,i}(t) \\ & \text{subject to:} && \mathbf{r}(t) \in \mathcal{C}(t) \cap \mathcal{R}(t) \\ & && \hat{q}_u(t) \leq \alpha_u(t) \log(r_u(t)) + \beta_u(t), \quad \forall u \in \mathcal{U}^{\text{av}}(t), \end{aligned} \tag{17}$$

where $s_{u,i}(t) = \frac{1}{T_u} ([x_i - \hat{q}_u(t)]^+ - h(x_i))$.

3: For $\forall u \in \mathcal{U}^{\text{av}}(t), \forall x_i \in \mathcal{I}$, update virtual queues with

$$v_{u,i}(t) = [v_{u,i}(t-1) + s_{u,i}(t)]^+.$$

4: **end for**

B. Admission Control Strategy

Since a video stream typically has high data rate and thus consumes a large amount of resources, the arrival and departure of a single video user can have a significant impact on other video users' QoE. Our admission control strategy is designed to identify and block those video users who may consume excessive network resources. As has been discussed in Algorithm 1, resource allocation in each slot is determined by the solution of the optimization problem (17). Therefore, it is possible to estimate the QoE of a newly arrived user by solving (17) as if the user had already been admitted. Based on this idea, we propose a threshold-based admission control strategy, which is summarized in Algorithm 2. For each newly arrived video user \bar{u} , we first estimate its video quality \bar{q} by solving the optimization problem (20), which is similar to the optimization problem (17). Then, we compare \bar{q} with a threshold θ . If \bar{q} is larger than θ , it is admitted to the network. Otherwise, it is rejected.

Algorithm 2 Admission control when video quality expectation is not known.

Inputs: Threshold θ , admitted users $\mathcal{U}^{\text{av}}(t-1)$, new user \bar{u}

- 1: Initialize video user set $\mathcal{U}^{\text{av}+} \leftarrow \mathcal{U}^{\text{av}}(t-1) \cup \{\bar{u}\}$
- 2: Estimate mean rate-quality parameters for $\forall u \in \mathcal{U}^{\text{av}+}$:

$$\begin{aligned}\hat{\alpha}_u &\leftarrow 1/\mathbb{T}_u \sum_{t=A_u}^{D_u} \alpha_u(t), \\ \hat{\beta}_u &\leftarrow 1/\mathbb{T}_u \sum_{t=A_u}^{D_u} \beta_u(t).\end{aligned}\tag{18}$$

- 3: Initialize virtual queue for a new user \bar{u} :

$$v_{\bar{u},i}(t-1) \leftarrow \frac{1}{|\mathcal{U}^{\text{av}}(t-1)|} \sum_{u \in \mathcal{U}^{\text{av}}(t-1)} v_{u,i}(t-1), \quad \forall i \in \mathcal{I}\tag{19}$$

- 4: Define variables $\hat{\mathbf{r}} = (\hat{r}_u : u \in \mathcal{U}^{\text{av}+})$, $\hat{\mathbf{q}} = (\hat{q}_u : u \in \mathcal{U}^{\text{av}+})$. Also define variables $\hat{s}_{u,i} = \frac{1}{\mathbb{T}_u} ([x_i - \hat{q}_u]^+ - h(x_i))$. Find the solution $\mathbf{r}^* = (r_u^* : u \in \mathcal{U}^{\text{av}+})$ of the optimization problem

$$\begin{aligned}\underset{\hat{\mathbf{r}}, \hat{\mathbf{q}}}{\text{minimize}} \quad & \sum_{u \in \mathcal{U}^{\text{av}+}} \sum_{x_i \in \mathcal{I}} v_{u,i}(t-1) \hat{s}_{u,i} \\ \text{subject to:} \quad & \hat{\mathbf{r}} \in \mathcal{C} \cap \mathcal{R},\end{aligned}\tag{20}$$

$$\hat{q}_u \leq \hat{\alpha}_u \log(\hat{r}_u) + \hat{\beta}_u, \quad \forall u \in \mathcal{U}^{\text{av}+},$$

where the sets $\mathcal{C} = \{\mathbf{r} : \mathbb{E}[C_t(\mathbf{r})] \leq 0\}$ and $\mathcal{R} = \Pi_{u \in \mathcal{U}^{\text{av}+}} \left[\frac{1}{\mathbb{T}_u} \sum_{t=A_u}^{D_u} r_u^{\min}(t), \frac{1}{\mathbb{T}_u} \sum_{t=A_u}^{D_u} r_u^{\max}(t) \right]$.

- 5: Estimate the video quality delivered to new user via

$$\bar{q} = \hat{\alpha}_{\bar{u}} \log(r_{\bar{u}}^*) + \hat{\beta}_{\bar{u}}.\tag{21}$$

- 6: If $\bar{q} > \theta$, admit the new user; otherwise, reject it.
-

The optimization problem (20) is different from (17) in the following three aspects. First, to predict the long-term QoE of users, we replace the instantaneous rate-quality parameters $\alpha_u(t)$ and $\beta_u(t)$ in (17) with the average rate-quality parameters $\hat{\alpha}_u$ and $\hat{\beta}_u$ (see the second step in Algorithm 2):

$$\begin{aligned}\hat{\alpha}_u &= \frac{1}{\mathbb{T}_u} \sum_{t=A_u}^{D_u} \alpha_u(t), \\ \hat{\beta}_u &= \frac{1}{\mathbb{T}_u} \sum_{t=A_u}^{D_u} \beta_u(t).\end{aligned}\tag{22}$$

Second, we replace the instantaneous rate region $\mathcal{C}(t)$ with a rate region estimated by

$$\mathcal{C} = \{\mathbf{r} : \mathbb{E}[C_t(\mathbf{r})] \leq 0\}.\tag{23}$$

Similarly, we replace the set of available video source rate $\mathcal{R}(t)$ by

$$\mathcal{R} = \Pi_{u \in \mathcal{U}^{\text{av}}(t-1) \cup \{\bar{u}\}} \left[\frac{1}{\mathbb{T}_u} \sum_{t=A_u}^{D_u} r_u^{\min}(t), \frac{1}{\mathbb{T}_u} \sum_{t=A_u}^{D_u} r_u^{\max}(t) \right].\tag{24}$$

Third, for a newly arrived user \bar{u} , we initialize its virtual queue with the average virtual queues of the existing users (see the third step in Algorithm 2), i.e.,

$$v_{\bar{u},i}(t-1) \leftarrow \frac{1}{|\mathcal{U}^{\text{av}}(t-1)|} \sum_{u \in \mathcal{U}^{\text{av}}(t-1)} v_{u,i}(t-1), \quad \forall i \in \mathcal{I}. \quad (25)$$

In (22), the rate-quality parameters $\alpha_u(\cdot)$ and $\beta_u(\cdot)$ are needed. For stored video streaming systems, the videos are pre-encoded. Thus, we assume the rate-quality parameters for the entire video stream are known. Also, the rate region \mathcal{C} in (23) can be estimated using the time-average of the previous channel conditions. For example, in TDMA systems, we have $C_t(\mathbf{r}) = \sum_{u \in \mathcal{U}^{\text{av}}(t)} \frac{r_u}{\bar{P}_u(t)} + \sum_{u' \in \mathcal{U}^{\text{p}}(t)} \frac{R_{u'}(t)}{\bar{P}_{u'}(t)} - 1$ (see Section IV-B). Thus, we can estimate $\mathbb{E} \left[\frac{1}{\bar{P}_u} \right]$ and $\mathbb{E} \left[\sum_{u' \in \mathcal{U}^{\text{p}}(t)} \frac{R_{u'}(t)}{\bar{P}_{u'}(t)} \right]$ using the previous observations of the peak rate \bar{P}_u and the high-priority users' data rate $R_{u'}$. The estimated rate region is therefore $\mathcal{C} = \left\{ \mathbf{r} : \sum_{u \in \mathcal{U}^{\text{av}}(t-1) \cup \{\bar{u}\}} \mathbb{E} \left[\frac{1}{\bar{P}_u} \right] r_u \leq 1 - \mathbb{E} \left[\sum_{u' \in \mathcal{U}^{\text{p}}(t)} \frac{R_{u'}(t)}{\bar{P}_{u'}(t)} \right] \right\}$.

As was discussed in Section V-A, the virtual queue $v_{u,i}(t)$ captures the difficulty for an admitted video user to satisfy the QoE constraints. Thus, users with large virtual queues tend to be allocated more network resources. In other words, the virtual queues drive the priorities in resource allocation. Because it is difficult to estimate the length of a virtual queue before a new user is admitted, we simply initialize the virtual queue of the newly arrived user with the average virtual queues of all existing video users. In this way, we actually estimate the video quality \bar{q} when an average priority is assigned to the new user. Next, we introduce an approach to optimize the admission threshold θ in Algorithm 2.

C. Online Algorithm for Threshold Optimization

Denote by $g(\theta)$ the probability that a video user's QoE constraints are satisfied when the threshold is θ . Also, denote by $e(\theta)$ the probability that a video user is admitted into the network but its QoE constraints are not satisfied. Our goal is to find the threshold θ^* that maximizes $g(\theta)$. We have conducted extensive simulations under different channel conditions and QoE constraints. From all the simulation results, we observed that the optimal threshold θ^* maximizes $g(\theta)$ if

$$\begin{cases} e(\theta) > 0, & \forall \theta < \theta^* \\ e(\theta) = 0, & \forall \theta \geq \theta^* \end{cases}. \quad (26)$$

This means that $g(\theta)$ is maximized if θ is just large enough to make all the admitted users satisfy the QoE constraints. As an example, using the same simulation configurations that are detailed later in Section V-D, we simulated and plotted the functions $e(\theta)$ and $g(\theta)$ in Fig. 5. It is seen that $\theta^* = 60$

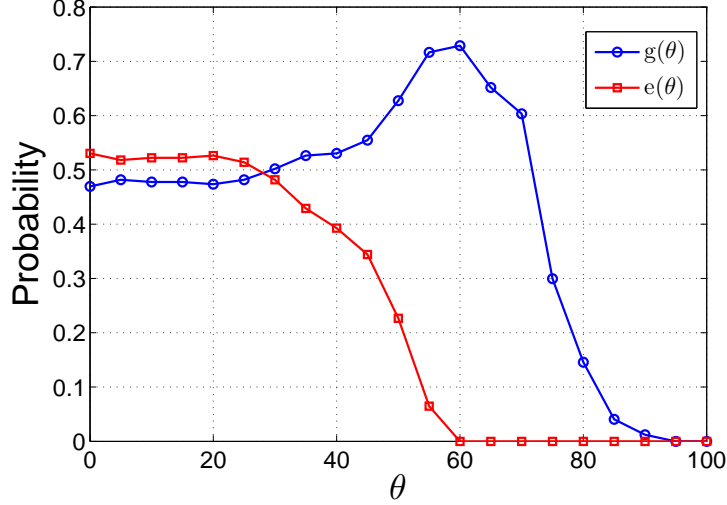


Fig. 5. The two plots show (i) the percentage of video users who satisfy the QoE constraints and (ii) the percentage of video users who are admitted into the network but do not satisfy the QoE constraints under different admission control thresholds.

satisfies (26) and $g(\theta)$ is also maximized at θ^* . Therefore, to find the optimal threshold θ^* , it is sufficient to find a threshold that satisfies (26).

We propose an iterative algorithm that automatically adjusts the threshold to θ^* . This is summarized in Algorithm 3. In each iteration, the algorithm observes the 2nd-order eCDFs of L video users who have been admitted into the network since the end of the last iteration. Then the algorithm updates the threshold via

$$\theta^{n+1} = \theta^n + \epsilon^n y^n, \quad (27)$$

where θ^n denote the admission control threshold in the n^{th} iteration. The value $y^n \in \{-1, 1\}$ determines whether to increase or to decrease the threshold. The quantity $\epsilon^n > 0$ is an updating step size. If the algorithm observes a video user whose 2nd-order eCDF violates the QoE constraints, then it is probable that $e(\theta^n) > 0$ and $\theta^n < \theta^*$. Therefore, the algorithm increases the threshold by setting $y^n = 1$. Otherwise, if all the L video users satisfy the constraints, the threshold is possibly larger than θ^* . Thus the algorithm decreases the threshold by setting $y^n = -1$. The updating step size is

$$\epsilon^n = \epsilon^0 / m, \quad (28)$$

where m counts the number of sign changes in the series $\{y^1, \dots, y^n\}$ (see step 8) and ϵ^0 is the initial step-size. Here, m is introduced to accelerate the convergence of the algorithm. The reason is as follows: If θ^n is far from θ^* , the sign of y^n does not change frequently and m increases slowly. Thus, the step-size ϵ^n stays large and θ^n moves towards θ^* quickly. When θ^n is moved to a small

neighborhood of θ^* , the sign of y^n changes frequently and thus m increases rapidly. Therefore, the step-size ϵ^n decreases to zero quickly, which makes θ^n converge.

Algorithm 3 The threshold optimization algorithm when the video quality expectation is unknown.

Inputs: $L = 100$, initial threshold $\theta^0 = 0$, initial step-size $\epsilon^0 = 10$, and initial counter $m = 1$

- 1: **for** $n = 1 \rightarrow \infty$ **do**
- 2: Observe the 2nd-order eCDFs of L admitted video users.
- 3: **if** there exists a user that does not satisfy the QoE constraints **then**
- 4: $y^n \leftarrow 1$
- 5: **else**
- 6: $y^n \leftarrow -1$
- 7: **end if**
- 8: If $y^n \neq y^{n-1}$, $m \leftarrow m + 1$.
- 9: Update threshold with

$$\theta^{n+1} = \theta^n + \epsilon^n y^n, \quad (29)$$

where $\epsilon^n = \epsilon^0/m$.

- 10: **end for**
-

In the following, we analyze the convergence of Algorithm 3. Based on our observations from the simulations, we make the following assumption:

Assumption 1. The function $e(\theta)$ is a continuous function and is strictly decreasing on $[0, \theta^*]$.

We define $p^L(\theta)$ as the probability that all the L admitted video users in an iteration satisfy the QoE constraints. Since increasing the threshold θ would block more users and thus reserve more network resources to the admitted users, we assume that $p^L(\theta)$ is a continuously increasing function of θ . Also, according to Assumption 1, when $\theta > \theta^*$, all the admitted users satisfy the QoE constraints and thus $p^L(\theta) = 1$. Thus, we have the following assumption on $p^L(\theta)$:

Assumption 2. The function $p^L(\theta)$ is a continuous and increasing function of θ . For $\forall \theta > \theta^*$, we have $p^L(\theta) = 1$. Furthermore, assume that there exists a constant $M > 0$ such that $|p^L(\theta) - p^L(\theta')| \geq M|\theta - \theta'|$ for all $\theta' < \theta$ and $p^L(\theta) < 1$.

The following theorem assures that if L is sufficiently large, θ^n converges to an arbitrarily small neighborhood of θ^* as $n \rightarrow \infty$. Its proof is given in Appendix B.

Theorem 2. *Let $\delta > 0$ be an arbitrarily small number. If Assumptions 1 and 2 are satisfied and $L \geq \frac{-\log 2}{\log(1-e^{-(\theta^*-\delta)})}$, then θ^n converges as $n \rightarrow \infty$ and $\lim_{n \rightarrow \infty} \theta^n \in [\theta^* - \delta, \theta^*]$.*

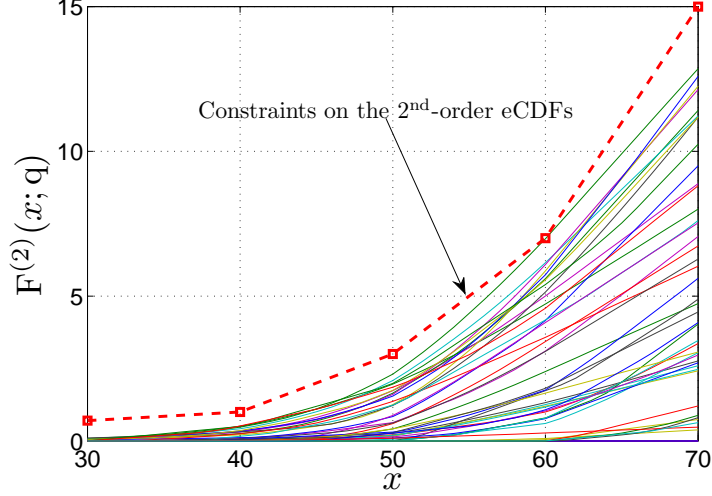
D. Simulation Results

Below, we evaluate our rate-adaptation algorithm and the admission control strategy via numerical simulations. We assume the duration of a time slot is $\Delta T = 1$ second. The high-priority users' arrivals follow a Poisson process with average arrival rate of $\frac{1}{20}$ users/second. The time spent by a high-priority user in the network is exponentially distributed with a mean value of 200 seconds. Video users also arrival as a Poisson process with a average arrival rate of $\frac{1}{20}$ users/second. Since video streams are typically more than tens of seconds long, we assume the time spent by a video user in the network is at least 40 seconds. In particular, for all video users, we set $T_u = \max\{T'_u, 40\}$, where T'_u is exponentially distributed with a mean value of 200 seconds.

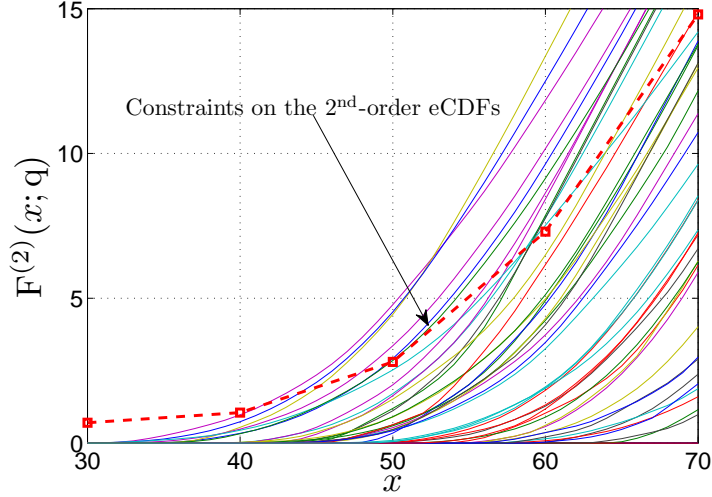
To simulate variations of the rate-quality characteristics in each video stream, we assume the rate-quality parameters $(\alpha_u(t), \beta_u(t))$ of each slot are independently sampled from the rate-quality parameters in the video database [20]. We assume the minimum and maximum available data rate for video users in (3) are $r_u^{\min}(t) = 302$ kbps and $r_u^{\max}(t) = 6412$ kbps, which are the minimum and maximum rate of the videos in the database [20]. For high-priority users, the downloading data rate R_u is assumed to be uniformly distributed in $[100, 300]$ kbps.

We assume the wireless system is a TDMA system. The rate region $\mathcal{C}(t)$ is that introduced in Section IV-B. We model the peak transmission rate $P_u(t)$ as the product of two independent random variables, i.e., $P_u(t) = P_u^{\text{avg}} \times P_u^*(t)$. The random variable P_u^{avg} is employed to simulate the heterogeneity of channel condition across users and remains constant during a user's sojourn. We assume that P_u^{avg} is uniformly distributed on $[1250\gamma, 3750\gamma]$ kbps, where the parameter γ is used to scale the channel capacity in our simulations. The random variable $P_u^*(t)$ is employed to simulate channel variation across time slots. We assume that $\{P_u^*(t) : t \in \mathbb{N}^+\}$ is an i.i.d. process with $P_u^*(t)$ being uniformly distributed on $[0.5, 1.5]$. In our simulations, we set $\mathcal{I} = \{30, 40, 50, 60, 70\}$. Correspondingly, for $x_i = 30, 40, 50, 60,$ and 70 , we let the constraints $h(x_i) = 0.7, 1.0, 3.0, 7.0$ and 15.0 , respectively.

We first evaluate the performance of the rate-adaptation algorithm when admission control is not applied. We set the scaling parameter $\gamma = 12$ and simulate Algorithm 1 until 100 users have arrived and departed the network. We plot the 2nd-order eCDFs of the video users in Fig. 6(a). It may be seen that, using Algorithm 1, the 2nd-order eCDFs of the video users all satisfy the constraints. By



(a) Proposed rate-adaptation algorithm



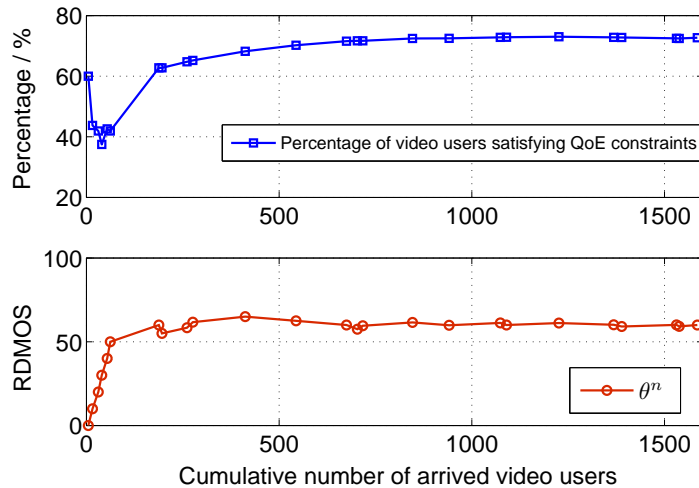
(b) Average-quality maximized rate adaptation

Fig. 6. Simulation results of rate-adaptation algorithms when admission control is not applied. (a) The 2nd-order eCDFs of the video users when the proposed rate-adaptation is used. (b) The 2nd-order eCDFs of the video users when the rate vector is adapted to maximize the sum of users’ video qualities.

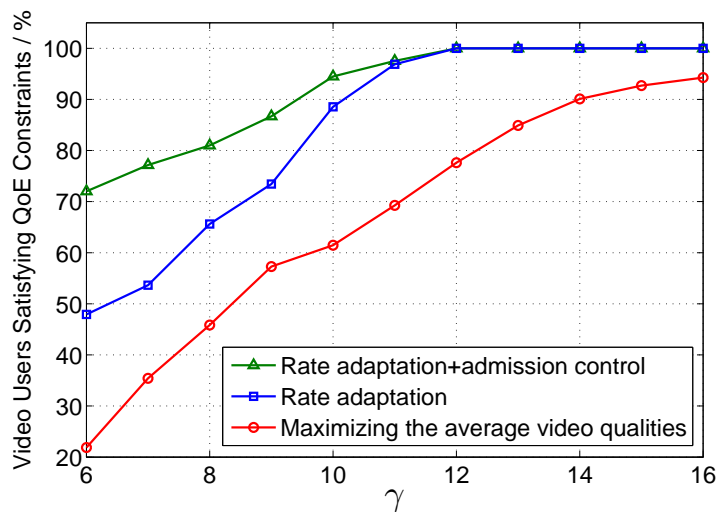
comparison, if we adapt the rate vector to maximize the sum of the average-quality of all users¹, the QoE constraint is violated by many users (see Fig. 6(b)).

Next, we evaluate the proposed admission control strategy. In Fig. 7(a), we fix the channel scaling parameter to be $\gamma = 6$ and plot the threshold θ^n at every iteration of the online threshold optimization algorithm. Recall that the optimal threshold is $\theta^* = 60$ (see Fig. 5). Fig. 7(a) shows that θ^n converges

¹This is achieved by maximizing $\sum_{u \in \mathcal{U}^{av}(t)} q_u(t) / T_u$ in each slot.



(a)



(b)

Fig. 7. (a) The performance of the proposed admission control strategy when the scaling parameter is $\gamma = 6$. (b) Simulation results of the proposed algorithms under different channel scaling parameters. Each data point on the figure is obtained by simulating 2000 video user arrivals.

to θ^* after 200 video user arrivals. We have assumed that the average arrival rate of video users is $1/20$ users/seconds. Thus, 200 video user arrivals require about 200×20 seconds = 1.1 hours. Since our goal is to optimize the performance of the network in the long run, this convergence speed is acceptable.

We scale the channel scaling parameter from $\gamma = 6$ to $\gamma = 16$. The percentage of video users whose video qualities satisfy the QoE constraints is shown in Fig. 7(b). When compared with the

average-quality-maximized rate-adaptation algorithm, the percentage of video users who satisfy the QoE constraints is improved significantly even if no admission control is applied. The admission control policy further improves the performance especially when the channel condition is poor. For example, at $\gamma = 6$, the proposed algorithms satisfy the QoE constraints of 70% of the video users while the average-quality-maximized rate-adaptation algorithm only satisfies the constraints of 20% of the video users. At a moderate channel condition of $\gamma = 12$, about 77% of the video users satisfy the QoE constraints when the average-quality maximizing algorithm is applied. The proposed algorithms achieve the same performance at $\gamma = 7.5$, reducing the consumption of resources by $(12 - 7.5)/12 = 38\%$.

VI. RATE ADAPTATION AND ADMISSION CONTROL WITH KNOWN VIDEO QUALITY EXPECTATION

In this section, we extend the rate adaptation algorithm and the admission control strategy to the case where the video quality expectation of each user is known. We first explain the extended algorithms and then evaluate their performance via simulation.

A. The Extended Rate Adaptation and Admission Control Algorithms

In Section IV-D, we defined the finite set $\mathcal{G} = \{g_1, \dots, g_{|\mathcal{G}|}\}$ to represent different video quality expectations among video users. In the following, we call users with $x_u^* = g_j \in \mathcal{G}$ the Type- j users. Each Type- j video user need only satisfy one QoE constraint, i.e.,

$$F^{(2)}(g_j; q_u) \leq h_j. \quad (30)$$

Thus, we extend the rate adaptation method in Algorithm 1 by maintaining one virtual queue for each user. In particular, the virtual queue of a Type- j user u is defined as

$$v_u(t) = \begin{cases} [v_u(t-1) + s_u(t)]^+ & \text{if } A_u \leq t \leq D_u, \\ 0 & \text{otherwise} \end{cases} \quad (31)$$

where

$$s_u(t) = \begin{cases} \frac{1}{T_u} ([g_j - \hat{q}_u(t)]^+ - h_j) & \text{if } A_u \leq t \leq D_u, \\ 0 & \text{otherwise} \end{cases}. \quad (32)$$

In each slot, the rate vector $\mathbf{r}(t)$ is adapted by solving

$$\begin{aligned} & \underset{\mathbf{r}(t), \hat{q}_u(t)}{\text{minimize}} && \sum_{u \in \mathcal{U}^{\text{av}}(t)} v_u(t-1) s_u(t) \\ & \text{subject to:} && \mathbf{r}(t) \in \mathcal{C}(t) \cap \mathcal{R}(t), \\ & && \hat{q}_u(t) \leq \alpha_u(t) \log(r_u(t)) + \beta_u(t), \quad \forall u \in \mathcal{U}^{\text{av}}(t). \end{aligned} \quad (33)$$

For admission control, we extend Algorithm 2 by applying different thresholds to different types of users. In particular, for a newly arrived Type- j video user \bar{u} , we initialize its virtual queue by averaging the virtual queues of all existing video users i.e.,

$$v_{\bar{u}}(t-1) \leftarrow \frac{1}{|\mathcal{U}^{\text{av}}(t-1)|} \sum_{u \in \mathcal{U}^{\text{av}}(t-1)} v_u(t-1). \quad (34)$$

Letting $\mathcal{U}^{\text{av}+} = \mathcal{U}^{\text{av}}(t) \cup \{\bar{u}\}$, we define variables $\hat{\mathbf{r}} = (\hat{r}_u : u \in \mathcal{U}^{\text{av}+})$, $\hat{\mathbf{q}} = (\hat{q}_u : u \in \mathcal{U}^{\text{av}+})$, and $\hat{s}_u = \frac{1}{\Gamma_u} ([g_j - \hat{q}_u]^+ - h_j)$. We then find the solution $\mathbf{r}^* = (r_u^* : u \in \mathcal{U}^{\text{av}+})$ of the following problem:

$$\begin{aligned} & \underset{\hat{\mathbf{r}}, \hat{\mathbf{q}}}{\text{minimize}} && \sum_{u \in \mathcal{U}^{\text{av}+}} v_u(t-1) \hat{s}_u \\ & \text{subject to:} && \hat{\mathbf{r}} \in \mathcal{C} \cap \mathcal{R}, \\ & && \hat{q}_u \leq \hat{\alpha}_u \log(\hat{r}_u) + \hat{\beta}_u, \quad \forall u \in \mathcal{U}^{\text{av}+} \end{aligned} \quad (35)$$

where $\hat{\alpha}_u$, $\hat{\beta}_u$, \mathcal{C} , and \mathcal{R} are given by (22), (23), and (24), respectively. Finally, we estimate the video quality of the new user by $\bar{q} = \hat{\alpha}_{\bar{u}} \log(r_{\bar{u}}^*) + \hat{\beta}_{\bar{u}}$ and compare \bar{q} with a threshold θ_j to make the admission decision.

Next, we discuss how to optimize the threshold θ_j for Type- j users.

B. The Extended Threshold Optimization Algorithm

Denote by $\boldsymbol{\theta} = (\theta_1, \dots, \theta_{|\mathcal{G}|})$ the vector of thresholds for all types of video users. Define $g(\boldsymbol{\theta})$ to be the probability that a video user satisfies the QoE constraints when the threshold vector is $\boldsymbol{\theta}$. Also, define $e_j(\boldsymbol{\theta})$ as the probability that a Type- j video user's QoE constraint is not satisfied. To determine the optimal threshold vector $\boldsymbol{\theta}$ that maximizes $g(\boldsymbol{\theta})$, we ran simulations under a variety of relevant channel conditions and QoE constraints. We found that a threshold vector $\boldsymbol{\theta}^* = (\theta_1^*, \dots, \theta_{|\mathcal{G}|}^*)$ maximizes $g(\boldsymbol{\theta})$ if

$$\begin{cases} e_j(\boldsymbol{\theta}) > 0, & \forall \boldsymbol{\theta} \prec \boldsymbol{\theta}^* \\ e_j(\boldsymbol{\theta}) = 0, & \forall \boldsymbol{\theta} \succeq \boldsymbol{\theta}^*. \end{cases}, \quad \forall 1 \leq j \leq |\mathcal{G}|. \quad (36)$$

Here, the partial order $\boldsymbol{\theta} \prec \boldsymbol{\theta}^*$ indicates that $\boldsymbol{\theta} \neq \boldsymbol{\theta}^*$ and $\theta_j \leq \theta_j^*$, $\forall j$. The partial order $\boldsymbol{\theta} \succeq \boldsymbol{\theta}^*$ indicates that $\theta_j \geq \theta_j^*$, $\forall j$. The condition in (36) means that if $\boldsymbol{\theta}^*$ is an optimal threshold vector and we increase all entries of $\boldsymbol{\theta}^*$, the QoE constraints of all the admitted users can still be satisfied. Conversely, if we decrease all the entries of $\boldsymbol{\theta}^*$, the QoE constraints of all types of users will be violated with a non-zero probability. To illustrate this, we considered two types of video users who arrive to the network with equal probability and simulated the function $e_1(\boldsymbol{\theta})$, $e_2(\boldsymbol{\theta})$, and $g_1(\boldsymbol{\theta})$ using the same setting as

in Section V-D. From Fig. 8(a) and Fig. 8(b), it can be seen that the θ s in $[0, 42] \times [63, 64]$ satisfy the condition (36) because they lie on the boundary of the region where $\{e_1(\theta) > 0 \text{ and } e_2(\theta) > 0\}$. From Fig. 8(c), it is seen that the function $g(\theta)$ is also maximized on the region $[0, 42] \times [63, 64]$.

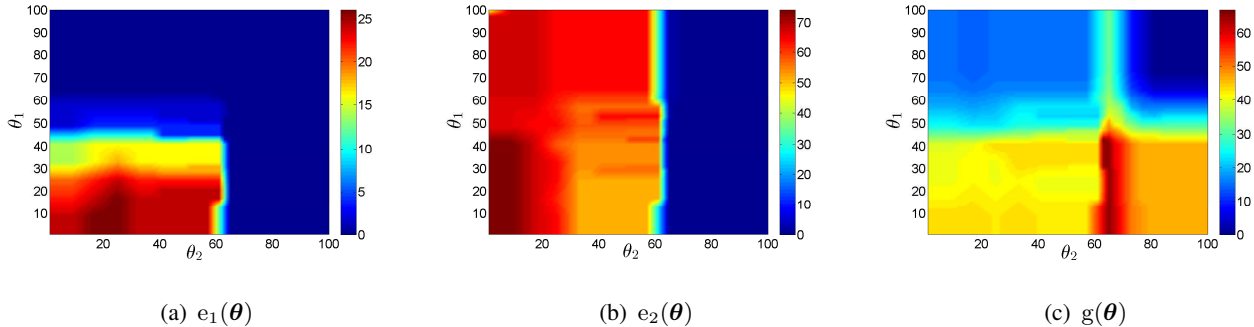


Fig. 8. (a) The probability of admitted Type-1 video users whose QoE constraints are violated. (b) The probability of admitted Type-2 video users whose QoE constraints are violated. (c) The probability of video users whose QoE constraints are satisfied. All results are shown in percentages. We assume these two types of user have $x_u^* = 40$ and 60 , respectively. The QoE constraint in (30) is assumed to be $h_1 = h_2 = 1$. The channel scaling parameter is set as $\gamma = 6$.

We devise an iterative algorithm to find the threshold vector θ^* satisfying (36). Denote by θ^n the threshold vector at the n^{th} iteration, the algorithm observes the 2^{nd} -order eCDFs of L admitted video users and updates the threshold vector using

$$\theta^{n+1} = \theta^n + \text{diag}(\epsilon^n) \mathbf{y}^n, \quad (37)$$

where $\text{diag}(\epsilon^n)$ is the diagonal matrix with diagonal entries being $\epsilon_1, \dots, \epsilon_{|\mathcal{G}|}$. In (37), $\mathbf{y}^n = (y_1^n, \dots, y_{|\mathcal{G}|}^n)$ is a $|\mathcal{G}|$ -dimensional vector where $y_j^n \in \{-1, 1\}$ is the updating direction for θ_j . The vector $\epsilon = (\epsilon_1^n, \dots, \epsilon_{|\mathcal{G}|}^n)$ is the corresponding update step-size. Among the L video users, if a Type- j video user's QoE constraint is not satisfied, the algorithm sets $y_j^n = 1$. Otherwise, if all the Type- j video users' QoE constraints are satisfied, the algorithm sets $y_j^n = -1$. The step-size ϵ_j^n is given by $\epsilon_j^n = \epsilon_j^0 / m_j$, where m_j counts the sign changes in $\{y_j^1, \dots, y_j^n\}$ and ϵ_j^0 is the initial step-size. Next, we show the performance of our rate adaptation algorithm and the admission control strategy via simulation.

C. Simulation Results

In our simulations, we assume that there are two types of video users. Both types of video users arrive as a Poisson process with arrival rate $1/40$ users/second. We assume that Type-1 users have $x_u^* = 40$ while Type-2 users have $x_u^* = 60$. We also set $h_1 = h_2 = 1$. In Fig. 9, we plot the percentage of video users whose video qualities satisfy the QoE constraint (8). It can be seen that, for all tested

channel scaling parameters, our rate adaptation algorithm outperforms the average-quality maximizing algorithm. The percentage of video users who satisfy the QoE constraints improved significantly even when the admission control strategy was not applied. At $\gamma = 12$, about 71% of video users satisfied the QoE constraints when the average-quality maximizing algorithm was applied. The proposed algorithms achieve the same performance at $\gamma = 6.5$. Thus, the proposed algorithms reduce the consumption of resources by $(12 - 6.5)/12 = 46\%$.

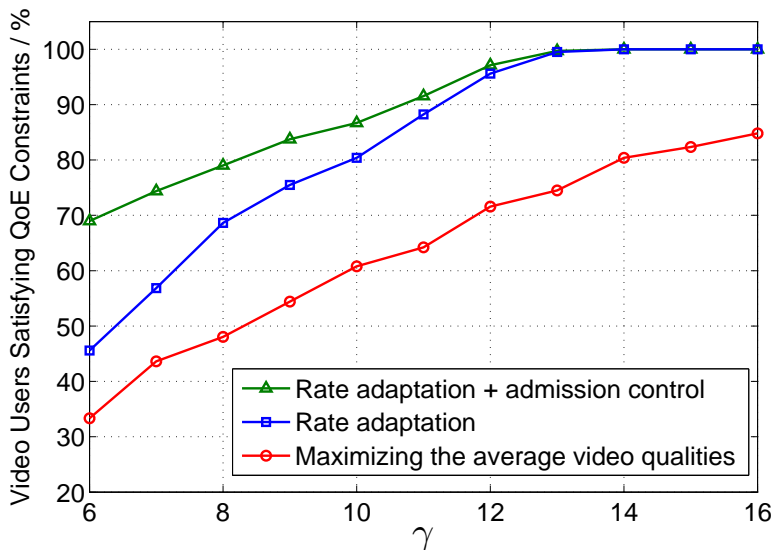


Fig. 9. Simulation results of the proposed admission control policy under different channel scaling parameters.

In Fig. 10, we plot the threshold vectors θ^n in the proposed threshold optimizing algorithm when the channel scaling parameter is fixed at $\gamma = 6$. We set the initial updating step-size $\epsilon_j^0 = 10, \forall j$. It is apparent that the threshold vector converges quickly to the area where $g(\theta)$ is maximized.

VII. CONCLUSIONS AND FUTURE WORK

We created a new QoE metric based on the second-order empirical cumulative distribution function (eCDF) of time-varying video quality. We then proposed an online rate adaptation algorithm to maximize the percentage of video users who satisfy the QoE constraints on the second-order cumulative distribution function. Furthermore, we devised a threshold-based admission control strategy that blocks new video users whose QoE constraints cannot be satisfied. Simulation results showed that combining the proposed approaches leads to a 40% reduction in wireless network resource consumption.

The users' video quality expectation play a critical role in our QoE metric. It is important to observe that our subjective study was conducted in a controlled environment. In reality, however,

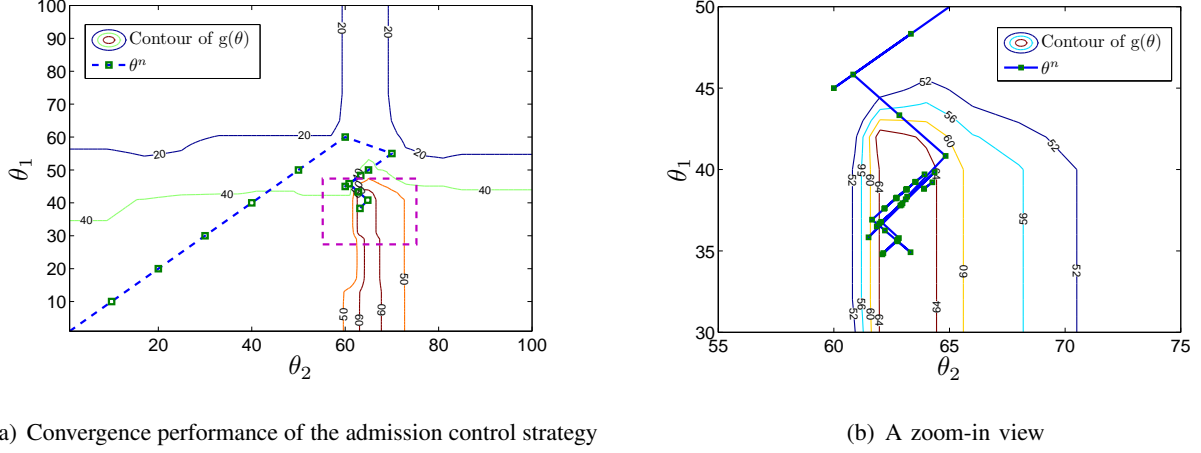


Fig. 10. The updated threshold vector θ^n s of the proposed threshold optimization algorithm are shown in (a). The dashed box is shown blown up in (b) to illustrate more detail. The contours of $g(\theta)$ are also shown on the figure for reference.

users' expectations for video quality may depend on various factors in the environment (e.g., user mobility, device type, lighting conditions). In the future, we plan to conduct subjective study in more diversified, "worldly" environments to obtain a better understanding of and an improved ability to predict users' quality expectations.

APPENDIX A

PROOF OF THEOREM 1

Proof: Since $[x - q(t)]^+$ is a convex function of x , $F^{(2)}(x; q)$ is a linear combination of $[x - q(t)]^+$ and is thus also a convex function of x . Without loss of generality, assume function $h(x)$ is also convex². Let $x_i < x_j$, where $i, j \in \mathcal{I}$. If (7) is satisfied, then $F^{(2)}(x_i; q) \leq h(x_i)$ and $F^{(2)}(x_j; q) \leq h(x_j)$. For any $\lambda \in [0, 1]$ and $x = \lambda x_i + (1 - \lambda)x_j$, we have $F^{(2)}(x; q) = F^{(2)}(\lambda x_i + (1 - \lambda)x_j; q) \leq \lambda F^{(2)}(x_i; q) + (1 - \lambda)F^{(2)}(x_j; q) \leq \lambda h(x_i) + (1 - \lambda)h(x_j) = \bar{h}(x)$. Because $[0, 100]$ is a compact set, the convexity of $h(x)$ implies its continuity. Therefore, $h(x)$ can be approximated by piece-wise linear functions to arbitrary accuracy. ■

²Otherwise, we can simply replace $h(x)$ with another function whose epigraph is the convex hull of $h(x)$'s epigraph.

APPENDIX B
PROOF OF THEOREM 2

Proof: Note that Algorithm 3 can be viewed as a stochastic approximation algorithm [38] with an associated mean ordinary differential equation (ODE)

$$\frac{d\theta(t)}{dt} = \mathbb{E}[Y(\theta(t))], \quad (38)$$

where $Y(\theta)$ is a random variable that denotes the updating direction when the threshold is θ , we have

$$\mathbb{E}[Y(\theta(t))] = 1 - 2p^L(\theta(t)). \quad (39)$$

According to Assumption 2, there exists a unique θ' such that $p^L(\theta') = 1/2$ and $\mathbb{E}[Y(\theta')] = 0$. By the monotonicity of $p^L(\theta)$, we have $\mathbb{E}[Y(\theta)] > 0$, $\forall \theta < \theta'$ and $\mathbb{E}[Y(\theta)] < 0$, $\forall \theta > \theta'$. If we define a function $V(\theta) = 1/2(\theta - \theta')^2$, then

$$\begin{aligned} \frac{dV(\theta(t))}{dt} &= (\theta(t) - \theta') \frac{d\theta(t)}{dt} \\ &= (\theta(t) - \theta') (1 - 2p^L(\theta(t))) \\ &= -(\theta(t) - \theta') (2p^L(\theta(t)) - 2p^L(\theta')) \\ &= -2(\theta(t) - \theta') (p^L(\theta(t)) - p^L(\theta')) \end{aligned}$$

For $\forall \theta < \theta'$, we have $p^L(\theta) - p^L(\theta') \leq M(\theta - \theta')$. For $\forall \theta > \theta'$, $p^L(\theta) - p^L(\theta') \geq M(\theta - \theta')$. In sum, we have

$$\frac{dV(\theta(t))}{dt} \leq -2M(\theta(t) - \theta')^2.$$

By Theorem 5.4.1 in [38, p.145], we have $\lim_{n \rightarrow \infty} \theta^n = \theta'$. Next, we prove that $\theta' \in [\theta^* - \delta, \theta^*]$.

We define a binary random variable S_u such that $S_u = 1$ if video user u satisfies the QoE constraints $\{F^{(2)}(x_i; q_u) \leq h_i, \forall x_i \in \mathcal{I}\}$. Otherwise, we define $S_u = 0$. Denote by $\mathcal{U}^L = \{u_1, \dots, u_L\}$ the indices of the L admitted video users in an iteration of Algorithm 2. Let π_θ be the joint distribution of the variables $\{S_u, \forall u \in \mathcal{U}^L\}$ when the admission threshold is θ . Then, we have

$$\begin{aligned} p^L(\theta) &= \mathbb{P}^{\pi_\theta} (S_u = 1, \forall u \in \mathcal{U}^L) \\ &= \mathbb{P}^{\pi_\theta} (S_{u_1} = 1) \mathbb{P}^{\pi_\theta} (S_{u_\ell} = 1, 2 \leq \ell \leq L | S_{u_1} = 1). \end{aligned} \quad (40)$$

Since the users are competing with each other for network resources, if the QoE constraints of a video user are satisfied, the probability of satisfying other users' QoE constraints is reduced. Thus,

$$\begin{aligned} &\mathbb{P}^{\pi_\theta} (S_{u_\ell} = 1, \forall 2 \leq \ell \leq L | S_{u_1} = 1) \\ &\leq \mathbb{P}^{\pi_\theta} (S_{u_\ell} = 1, \forall 2 \leq \ell \leq L). \end{aligned} \quad (41)$$

Substitute (41) into (40) yields

$$\begin{aligned}
p^L(\theta) &\leq \mathbb{P}^{\pi_\theta}(S_{u_1} = 1) \mathbb{P}^{\pi_\theta}(S_{u_\ell} = 1, \forall 2 \leq \ell \leq L) \\
&\leq \mathbb{P}^{\pi_\theta}(S_{u_1} = 1) \mathbb{P}^{\pi_\theta}(S_{u_2} = 1) \mathbb{P}^{\pi_\theta}(S_{u_\ell} = 1, \forall 3 \leq \ell \leq L) \\
&\leq \prod_{\ell=1}^L \mathbb{P}^{\pi_\theta}(S_{u_\ell} = 1) \\
&= (1 - e(\theta))^L.
\end{aligned} \tag{42}$$

Since $L \geq \frac{-\log 2}{\log(1-e(\theta^*-\delta))}$, it follows that $p^L(\theta^* - \delta) \leq (1 - e(\theta^* - \delta))^L \leq 1/2$. Because of the monotonicity of $p^L(\theta)$, we know that $\theta^* - \delta \leq \theta'$. Moreover, because $p^L(\theta') = 1/2 > 0$, we have $\theta' < \theta^*$. ■

REFERENCES

- [1] Cisco, “Cisco visual networking index: Global mobile data traffic forecast update, 2011-2016,” Feb. 2012.
- [2] F. Dobrian, V. Sekar, A. Awan, I. Stoica, D. Joseph, A. Ganjam, J. Zhan, and H. Zhang, “Understanding the impact of video quality on user engagement,” in *Proceedings of the ACM SIGCOMM 2011 conference*, 2011, pp. 362–373.
- [3] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the scalable video coding extension of the H.264/AVC standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sept. 2007.
- [4] Microsoft Corporation, “IIS smooth streaming technical overview,” <http://www.microsoft.com/en-us/download/default.aspx>, Sept. 2009.
- [5] R. Pantos and E. W. May, “HTTP live streaming,” *IETF Internet Draft, work in progress*, Mar. 2011.
- [6] Adobe Systems, “HTTP dynamics streaming,” <http://www.adobe.com/products/hds-dynamic-streaming.html>, Mar. 2012.
- [7] MPEG Requirements Group, “ISO/IEC FCD 23001-6 Part 6: Dynamics adaptive streaming over HTTP (DASH),” http://mpeg.chiariglione.org/working_documents/mpeg-b/dash/dash-dis.zip, Jan. 2011.
- [8] C. Luna, Y. Eisenberg, R. Berry, T. Pappas, and A. Katsaggelos, “Joint source coding and data rate adaptation for energy efficient wireless video streaming,” *IEEE J. Sel. Areas Commun.*, vol. 21, no. 10, pp. 1710–1720, 2003.
- [9] J. Chakareski and P. Frossard, “Rate-distortion optimized distributed packet scheduling of multiple video streams over shared communication resources,” *IEEE Trans. Multimedia*, vol. 8, no. 2, pp. 207–218, 2006.
- [10] J. Huang, Z. Li, M. Chiang, and A. Katsaggelos, “Joint source adaptation and resource allocation for multi-user wireless video streaming,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 5, pp. 582–595, 2008.
- [11] X. Zhu and B. Girod, “Distributed media-aware rate allocation for wireless video streaming,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1462–1474, 2010.
- [12] K. Lin, W.-L. Shen, C.-C. Hsu, and C.-F. Chou, “Quality-differentiated video multicast in multirate wireless networks,” *IEEE Trans. Mobile Comput.*, vol. 12, no. 1, pp. 21–34, 2013.
- [13] H. Hu, X. Zhu, Y. Wang, R. Pan, J. Zhu, and F. Bonomi, “Proxy-based multi-stream scalable video adaptation over wireless networks using subjective quality and rate models,” *to appear on IEEE Trans. Multimedia*, 2013.
- [14] A. Bovik, “Automatic prediction of perceptual image and video quality,” *Proceedings of the IEEE*, vol. 101, no. 9, pp. 2008–2024, 2013.
- [15] M. Barkowsky, B. Eskofier, R. Bitto, J. Bialkowski, and A. Kaup, “Perceptually motivated spatial and temporal integration of pixel based video quality measures,” in *Welcome to Mobile Content Quality of Experience*, Mar. 2007, pp. 1–7.
- [16] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba, “Considering temporal variations of spatial visual distortions in video quality assessment,” *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 253–265, 2009.

- [17] F. Yang, S. Wan, Q. Xie, and H. R. Wu, "No-reference quality assessment for networked video via primary analysis of bit stream," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1544–1554, 2010.
- [18] K. Seshadrinathan and A. C. Bovik, "Temporal hysteresis model of time-varying subjective video quality," *IEEE Int'l Conf. Acoust. Speech Signal Process*, May. 2011.
- [19] J. Park, K. Seshadrinathan, S. Lee, and A. Bovik, "Video quality pooling adaptive to perceptual distortion severity," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 610–620, 2013.
- [20] C. Chen, L. K. Choi, G. de Veciana, C. Caramanis, R. W. Heath Jr., and A. C. Bovik, "A dynamic system model of time-varying subjective quality of video streams over HTTP," *IEEE Int'l Conf. Acoust. Speech Signal Process*, May 2013.
- [21] ——. (2013, Jun.) A model of the time-varying subjective quality of http video streams. [Online]. Available: <https://webspace.utexas.edu/cc39488/pdf/report2013.pdf>
- [22] C. Yim and A. C. Bovik, "Evaluation of temporal variation of video quality in packet loss networks," *Signal Process: Image Commun.*, vol. 26, no. 1, pp. 24–38, Jan. 2011.
- [23] V. Joseph and G. de Veciana, "Jointly optimizing multi-user rate adaptation for video transport over wireless systems: Mean-fairness-variability tradeoffs," in *2012 Proceedings IEEE INFOCOM*, 2012, pp. 567–575.
- [24] S. Weber and G. de Veciana, "Rate adaptive multimedia streams: optimization and admission control," *IEEE/ACM Trans. Netw.*, vol. 13, no. 6, pp. 1275–1288, 2005.
- [25] F.-S. Lin, "Optimal real-time admission control algorithms for the video-on-demand (VOD) service," *IEEE Trans. Broadcast.*, vol. 44, no. 4, pp. 402–408, 1998.
- [26] P. Mundur, A. Sood, and R. Simon, "Class-based access control for distributed video-on-demand systems," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 7, pp. 844–853, 2005.
- [27] Y.-H. Tseng, E.-K. Wu, and G.-H. Chen, "An admission control scheme based on online measurement for VBR video streams over wireless home networks," *IEEE Trans. Multimedia*, vol. 10, no. 3, pp. 470–479, 2008.
- [28] W. Pu, Z. Zou, and C. W. Chen, "Video adaptation proxy for wireless dynamic adaptive streaming over http," in *19th Int'l Packet Video Wkshp. (PV)*, 2012, pp. 65–70.
- [29] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h.264/avc video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, 2003.
- [30] G. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (hevc) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, 2012.
- [31] ITU-R Recommendation BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," http://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.500-13-201201-I!!PDF-E.pdf, Jan. 2012.
- [32] F. Bellard and M. Niedermayer, "FFmpeg," <http://ffmpeg.org/>, 2012.
- [33] R. Soundararajan and A. C. Bovik, "Video quality assessment by reduced reference spatio-temporal entropic differencing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 4, pp. 684–694, 2013.
- [34] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.
- [35] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, 1996.
- [36] A. L. Stolyar, "Maximizing queueing network utility subject to stability: Greedy primal-dual algorithm," *Queueing Systems*, vol. 50, no. 4, pp. 401–457, 2005.
- [37] M. J. Neely, *Stochastic Network Optimization with Application to Communication and Queueing Systems*. Morgan & Claypool, 2010.
- [38] H. J. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*. Springer Press, 2003.