

Encoding Bitrate Optimization Using Playback Statistics for HTTP-based Adaptive Video Streaming

Chao Chen
Google Inc.
1600 Amphitheatre Parkway
Mountain View, California
chaochen@google.com

Anil Kokaram
Department of Electronic and
Electrical Engineering, Trinity
College Dublin,
College Green, Dublin, Ireland
anil.kokaram@tcd.ie

Yao-Chung Lin
Google Inc.
1600 Amphitheatre Parkway
Mountain View, California
yclin@google.com

Steve Benting
Google Inc.
1600 Amphitheatre Parkway
Mountain View, California
sbenting@google.com

ABSTRACT

HTTP video streaming is in wide use to deliver video over the Internet. With HTTP adaptive steaming, a video playback dynamically selects a video stream from a pre-encoded representation based on available bandwidth and viewport (screen) size. The viewer's video quality is therefore influenced by the encoded bitrates. We minimize the average delivered bitrate subject to a quality lower bound on a per-chunk basis by modeling the probability that a player selects a particular encoding. Through simulation and real-world experiments, the proposed method saves 9.6% of bandwidth while average delivered video quality comparing with state of the art while keeping average delivered video quality.

Keywords

Video Streaming, MPEG-DASH, HLS, QoE, Rate-Quality Optimization

1. INTRODUCTION

To avoid buffering stalls while streaming videos over communication networks with time-varying throughput, adaptive video streaming protocols based on Hypertext Transfer Protocol (HTTP) have been widely deployed [11]. Among them, the most popular ones are Dynamic Adaptive Streaming over HTTP (DASH) [4] issued by Moving Picture Experts Group (MPEG) and the HTTP Live Streaming (HLS) protocol [9] proposed by Apple Inc. As illustrated in Fig. 1, with HTTP-based adaptive streaming, videos are encoded into multiple representations at different bitrates and resolutions. Each representation is then partitioned into short video segments. At any moment, the video player dynami-

cally requests the streaming server to send a segment from an appropriate representation: one where the bitrate of the segment is lower than the available network bandwidth capacity and the resolution of the segment fits into the player's viewport.

From the perspective of streaming service providers, e.g., YouTube and Netflix, it is always desirable to deliver maximum video quality with minimum data bitrates. The encoding bitrates and qualities of the representations, to a large extent, determine the delivered video quality and the volume of data traffic. However, it is challenging to find out the optimal encoding bitrates, because the adaptation behavior of the video player affects users' viewing experience and bandwidth cost as well. For example, increasing the encoding bitrate of a representation may not lead to better delivered quality, because that may cause players to switch to other representations with lower bitrates and result in worse video quality presented to viewers. Similarly, decreasing the encoding bitrate of a representation may not reduce the traffic cost, because the players are more likely to select other representations with higher quality, which incurs increased traffic cost.

Rate-quality optimization has been extensively studied in the realm of video compression [7, 8, 15]. These technologies optimize the bit allocation within a single video stream to minimize the encoding bitrate at a given encoding quality. For encoding a single bitstream, the performance of a video codec can be fully characterized by its rate-quality curve. If the rate-quality curve of one codec dominates the rate-quality curve of another codec, the former codec can compress a video at a given quality with a lower bitrate, and thus has better performance. In this paper, we investigate the rate-quality optimization problem for HTTP-based adaptive streaming where players dynamically switch among multiple encoded representations. In this scenario, the average streaming bitrate and delivered quality depend on the encoding bits allocated for all representations.

A method proposed in [3] configures encoding bitrates solely based on the rate-quality curves of representations. The impact of bandwidth and viewport size distributions on the average bitrate and delivered video quality was not considered and is thus sub-optimal for adaptive video streaming.

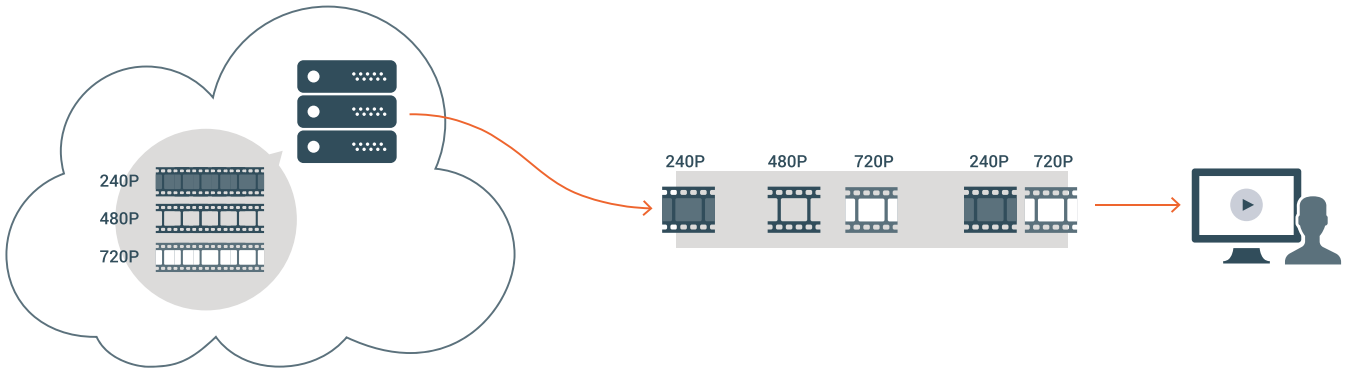


Figure 1: Adaptive video streaming over HTTP.

The optimal selection of DASH representations is also investigated in [5], [6], and [13]. The problem is formulated as an integer programming that attempts to maximize users’ satisfaction given content delivery network (CDN) capacity constraints, content type, and end-user characteristics. Rainer *et al.* proposed an optimization approach for rate adaptation on video streaming when the set of representations are given [10]. The idea is to solve a general optimization problem that maximizes delivery quality given download capacity constraint and yield quality upper bound for each streaming session. Then the quality upper bound is folded into another optimization problem that selects which representation to stream. The selected representation minimizes the bandwidth cost while delivering quality close to the upper bound.

In this paper, we aim to minimize the bandwidth cost by optimizing the encoding bitrates for a given set of representations. We consider the large-scale video streaming systems such as YouTube or Netflix where CDN bandwidth cost dominates the computational cost of encoding. In this scenario, we can spend computational resources to obtain rate-quality curves at all resolutions on a per video chunk basis. It enables us to take into account the unique rate-quality characteristics of each video chunk for encoding bitrate optimization. In addition, we collected real-world playback traces from our video streaming platform and obtained the empirical distributions of estimated client bandwidth and viewport sizes. These empirical distributions enable us to establish a model for the adaptation behavior of players, which is critical to the optimization of encoding bitrates. The optimization step is on a per-chunk basis, i.e. each video chunk can have different representation bitrates instead of a fixed bitrate set per video content type in contrast with the method proposed in [13].

We first describe a simple model for the adaptation behavior in players in Section 2. In addition to the rate-quality characteristics of the encoded videos at per chunk basis, the model also incorporates the real-world statistics of network bandwidth and the viewport size of players. Using this model, we are able to establish the mapping from the encoding bitrates of encoded representations to the average streaming bitrate and the video quality delivered to viewers. Then, we propose a simple optimization framework to identify the optimal encoding bitrates that minimize the average streaming bitrate, subject to a given lower bound on delivered quality.

We simulated the performance of the proposed method. The results are presented in Section 3. The average streaming bitrate can be reduced by 9.45% to 12.07%. We then implemented the proposed method in our transcoding system. The experimental results show that the average video traffic is reduced by 9.6% to 14.37% without degrading users’ Quality-of-Experience. Please see Section 4 for more detailed results.

2. ENCODING BITRATE OPTIMIZATION METHOD

In this section, we first establish a theorem that characterizes the achievable performance for HTTP-adaptive streaming. Then, we introduce a mathematical model for characterizing the adaptation behavior of video players and formulate the encoding bitrate configurations as an optimization problem. Finally, we describe the overall system implementation.

2.1 Rate Quality Region

Let $\mathcal{I} \subset \mathbb{N}$ be the index set of the encoded representations of an input video. Let $\mathcal{V} = \{144, 240, \dots, 2160\}$ denote the set of supported viewport heights on players. The i ’th representation is obtained by encoding the input video at bitrate $r_i \in \mathbb{R}$ and resolution $v_i \in \mathcal{V}$. We assume the representations can be ordered by their encoding bitrates and viewport heights in ascending order, i.e., $\forall i < j$ in \mathcal{I} , we have $r_i \leq r_j$ and $v_i \leq v_j$. Note that the equality in $v_i \leq v_j$ allows multiple representations per resolution.

Let q_i be the encoding quality of the i ’th representation. The encoding quality can be measured with any video quality metric such as PSNR or SSIM [14]. Since different representations may have different resolutions, to make their encoding quality comparable, q_i is obtained by first upscaling the representation to the resolution of the input video and then calculating the PSNR or SSIM against the input video.

For a given input video and an output resolution $v \in \mathcal{V}$, the encoding quality q at arbitrary encoding bitrate r can be modeled by its rate-quality characteristic function $q_v(r)$, i.e.,

$$q_i = q_{v_i}(r_i), \forall i \in \mathcal{I}. \quad (1)$$

The function $q_v(\cdot)$ depends on the nature of the video content and the compression algorithm adopted by the codec.

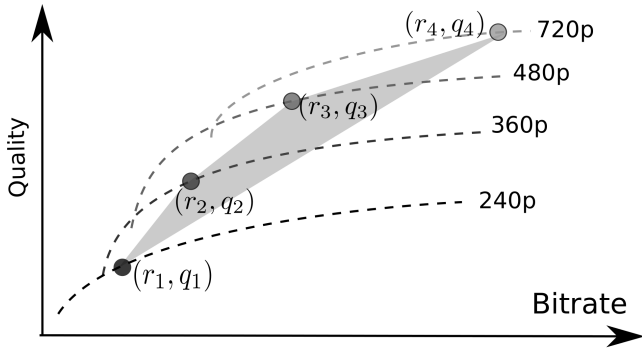


Figure 2: Typical rate-quality curves of a video encoded at different resolutions. The achievable rate-quality region is shown by the shaded area.

In Fig. 2, we illustrate the typical rate-quality functions of a video corresponding to different encoding resolutions. For each representation, the rate-quality operating point (r_i, q_i) will always fall on the rate-quality curve $q_{v_i}(\cdot)$.

Assuming the viewers spent t_i seconds watching representation i , the fraction of time viewers spent on representation i is thus

$$\lambda_i = t_i / \sum_{j \in \mathcal{I}} t_j. \quad (2)$$

The average streaming bitrate over time is given by

$$\mathbf{R}(\mathbf{r}) = \frac{1}{\sum_{j \in \mathcal{I}} t_j} \sum_{i \in \mathcal{I}} t_i r_i = \sum_{i \in \mathcal{I}} \lambda_i r_i = \boldsymbol{\lambda}^\top \mathbf{r}, \quad (3)$$

where $\mathbf{r} = (r_i : i \in \mathcal{I})$ and $\boldsymbol{\lambda} = (\lambda_i : i \in \mathcal{I})$. Similarly, the average quality delivered to users is given by

$$\mathbf{Q}(\mathbf{r}) = \boldsymbol{\lambda}^\top \mathbf{q}, \quad (4)$$

where $\mathbf{q} = (q_i : i \in \mathcal{I}) = (q_{v_i}(r_i) : i \in \mathcal{I})$ is the vector of encoding qualities. Because of $\sum_i \lambda_i = 1$ and $\lambda_i \geq 0 \forall i \in \mathcal{I}$, the average bitrate-quality point $(\mathbf{R}(\mathbf{r}), \mathbf{Q}(\mathbf{r}))$ falls in the convex hull spanned by the encoding rate-quality points $\{(r_i, q_i) : i \in \mathcal{I}\}$, which is illustrated in Fig. 2. In other words, any rate-quality point outside the convex hull cannot be achieved by an adaptive streaming system. This is summarized in Theorem 1.

THEOREM 1. *For an adaptive video streaming service, the achievable region of average streaming bitrate R and average quality Q is given by the convex hull*

$$\mathcal{S}(\mathbf{r}) = \{(R, Q) : R = \boldsymbol{\lambda}^\top \mathbf{r}, Q = \boldsymbol{\lambda}^\top \mathbf{q}, \mathbf{1}^\top \boldsymbol{\lambda} = 1, \lambda_i \geq 0, \forall i \in \mathcal{I}\}. \quad (5)$$

This theorem reveals some of the rationale of the encoding configurations proposed in [3]. This configuration selects the encoding points $\{(r_i, q_i) : i \in \mathcal{I}\}$ from the upper boundary of the convex hull spanned by the rate quality curves. It pushes the achievable region $\mathcal{S}(\mathbf{r})$ to the low bitrate and high quality region in the rate-quality space. The corresponding average bitrate tends to be reduced and the average delivered quality tends to be improved. However, the exact position of point $(\mathbf{R}(\mathbf{r}), \mathbf{Q}(\mathbf{r}))$ depends on $\boldsymbol{\lambda}$, i.e., the adaptation behavior of players, which is not considered in [3]. In the next section, we propose a model to characterize the adaptation behavior of players.

2.2 Player Model

We model the player-estimated bandwidth and viewport size at players as two stationary random processes $\{\mathbf{R}_t : t \in \mathcal{R}^+\}$ and $\{\mathbf{V}_t : t \in \mathcal{R}^+\}$, respectively. We assume a player selects the streamed representations according to two rules:

1. The player always requests a representation whose resolution is lower than or equals to player viewport size \mathbf{V}_t . This is to save bandwidth by not streaming unnecessary pixels to viewers.
2. Among the representations satisfying the first rule, the player always selects the highest representation whose bitrate is lower than its estimated bandwidth \mathbf{R}_t . This is to ensure that the bandwidth is fully utilized while the streamed representation can be smoothly played without stalls.

These two rules are widely followed by video players in practice. At any moment t , a player requests representation i in the following two cases. If the viewport size \mathbf{V}_t equals the resolution of representation i , the player requests representation i when the bandwidth \mathbf{R}_t is greater than the encoding bitrate r_i . If the viewport size \mathbf{V}_t is larger than the resolution of representation i , the player requests representation i if \mathbf{R}_t is greater than the bitrate of representation i but is less than that of higher representations. The probability for a player to select representation i is thus

$$p_i(\mathbf{r}) = \mathbb{P}[\mathbf{V}_t = v_i, \mathbf{R}_t > r_i] + \mathbb{P}[\mathbf{V}_t > v_i, r_{i+1} \geq \mathbf{R}_t > r_i]. \quad (6)$$

The two terms on the right hand side of (6) correspond to the probability of the two cases above. We estimated the statistical distributions of bandwidth and viewport sizes from playback statistics and found that the bandwidth distribution does not vary significantly with the viewport size v . This is because streaming bandwidth is mainly determined by network conditions when a video is played, which are not related to the viewport size of devices. Therefore, we may assume that viewport size \mathbf{V}_t and bandwidth \mathbf{R}_t are two independent processes. The viewing probability $p_i(\mathbf{r})$ in (6) can thus be rewritten as

$$p_i(\mathbf{r}) = \mathbb{P}[\mathbf{V}_t = v_i] \mathbb{P}[\mathbf{R}_t > r_i] + \mathbb{P}[\mathbf{V}_t > v_i] \mathbb{P}[r_{i+1} \geq \mathbf{R}_t > r_i]. \quad (7)$$

Assuming \mathbf{R}_t and \mathbf{V}_t are ergodic random processes, we have

$$p_i(\mathbf{r}) = \lim_{\sum_i t_i \rightarrow \infty} \frac{t_i}{\sum_i t_i} = \lambda_i \quad (8)$$

where λ_i is defined in equation (2) as the fraction of time that players stream representation i . Substituting (8) into (3) and (4), we have

$$\mathbf{R}(\mathbf{r}) = \sum_{i \in \mathcal{I}} r_i p_i(\mathbf{r}), \quad (9)$$

$$\mathbf{Q}(\mathbf{r}) = \sum_{i \in \mathcal{I}} q_i p_i(\mathbf{r}). \quad (10)$$

Substituting (7) into (9) and (10), we obtain the explicit expression for average bitrate and delivered quality in (11) and (12).

$$\mathbf{R}(\mathbf{r}) = \sum_{i \in \mathcal{I}} \left[r_i \left(\mathbb{P}[\mathbf{V}_t = v_i] \mathbb{P}[\mathbf{R}_t > r_i] \right. \right. \\ \left. \left. + \mathbb{P}[\mathbf{V}_t > v_i] \mathbb{P}[r_{i+1} \geq \mathbf{R}_t > r_i] \right) \right], \quad (11)$$

and

$$\mathbf{Q}(\mathbf{r}) = \sum_{i \in \mathcal{I}} \left[q_{v_i}(r_i) \left(\mathbb{P}[\mathbf{V}_t = v_i] \mathbb{P}[\mathbf{R}_t > r_i] \right. \right. \\ \left. \left. + \mathbb{P}[\mathbf{V}_t > v_i] \mathbb{P}[r_{i+1} \geq \mathbf{R}_t > r_i] \right) \right]. \quad (12)$$

These expressions can then be used to estimate the corresponding average bitrate and average delivered quality for a given encoding bitrate configuration \mathbf{r} .

2.3 Encoding Bitrate Optimization

We propose minimizing the average bitrate, subject to a given lower bound on average delivered quality, by solving the following optimization problem

$$\begin{aligned} & \underset{\mathbf{r}}{\text{minimize}} && \mathbf{R}(\mathbf{r}) \\ & \text{subject to} && \mathbf{Q}(\mathbf{r}) \geq \mathbf{Q}_0, \end{aligned} \quad (13)$$

where $\mathbf{R}(\mathbf{r})$ and $\mathbf{Q}(\mathbf{r})$ are given by equation (11) and (12). The value \mathbf{Q}_0 is a desired lower bound for delivered quality.

Here, we target minimizing average bitrate given a constraint on quality. In 2015, video traffic accounted for 70% of IP network traffic and 55% of mobile network traffic[2]. Reducing bandwidth costs is critical for the success of streaming services.

To solve the optimization problem in (13), gradient-based optimization algorithms such as method of moving asymptotes (MMA) [12] can be employed. It is worth noting that the optimization problem (13) is non-convex because the bitrate cumulative distribution function $\mathbb{P}[\mathbf{R}_t \leq x]$ is not a convex function of x in general. Thus, there is no guarantee that the local optimal solution generated by gradient-based optimization algorithms could achieve a global optimum. However, in practice, we found that the local optimal solutions still provided a significant reduction in average bitrate.

2.4 System Implementation

The proposed method is integrated in our video processing pipeline that processes and re-encodes ingested videos. An ingested video is first divided into 5-second nonoverlapped chunks. Then we obtain the rate-quality models of each video chunk at all resolutions. For example, we encode a 1080p video chunk into 6 different resolutions including 144p, 240p, 360p, 480p, 720p, and 1080p. For each resolution, we constructed the rate-quality model by sampling rate-quality points from $q_{v_i}(r)$. Specifically, we encoded a video multiple times using the libx264 codec. Each time we applied a different Constant Rate Factor (CRF) sampled from 5 to 55 with a step size of 5 so as to cover a wide range of encoding qualities. Then we scaled up each encoded version to 1080p with bicubic scaling filter and calculated the corresponding PSNR against the original video. The rate-quality model $q_{v_i}(r)$ was approximated using the piece-wise linear function connecting the sampled rate-quality points.

On our video streaming platform, the video players recorded the estimated bandwidth in every few seconds during each playback session. The estimated bandwidth, along with the viewport size is then stored to our backend databases. We collected 1,000,000 such real-world playback traces to obtain the empirical distribution of viewport and estimated bandwidth, i.e., $\mathbb{P}[\mathbf{V}_t]$ and $\mathbb{P}[\mathbf{R}_t]$.

The optimizer described in Sec. 2.3 takes into account the rate-quality curves, the empirical distributions of viewport and estimated downlink bandwidth, and the default delivered quality \mathbf{Q}_0 based on the default settings. It then runs the MMA solver in the NLOpt library [1] to obtain the optimized bitrates. Finally, the video chunks will be encoded into the representations using the optimized bitrates, and muxed into DASH or HLS formats for streaming.

In the next section, we evaluate the gain of the optimized encoding bitrate configurations via numerical simulations.

3. NUMERICAL SIMULATIONS

A set of 1000 1080p videos, the content of which covers a wide-range of spatio-temporal complexities, is randomly selected to run the simulation. The lengths of the selected videos ranged from 1 minute to 20 minutes. Every videos are processed as described in Sec. 2.4.

We first evaluated the performance of the proposed method against a baseline encoding parameter configuration where a fixed CRF of 23 is applied to all representations. Here, we chose a CRF of 23 because it is the default CRF in ffmpeg. We denote by r_i^0 the baseline encoding bitrate for representation i and let $\mathbf{r}_0 = (r_i^0 : i \in \mathcal{I})$. We set $\mathbf{Q}_0 = \mathbf{Q}(\mathbf{r}_0)$ as the lower bound for delivered quality. Then we employed the MMA solver in the NLOpt library [1] to find the optimal encoding bitrates \mathbf{r}^* and the corresponding average streaming bitrate $\mathbf{R}(\mathbf{r}^*)$ in (13).

On the 1000 test videos, we found that the proposed algorithm can reduce the average streaming bitrate by 12.07%. We plot the encoding configurations of an example video in Fig. 3. It can be seen that the optimized encoding bitrates are smaller than the encoding bitrates with fixed CRFs. This leads to reduction in encoding bitrate and degradation in encoding quality for all representations. Interestingly, because the delivered quality depends on the viewing probability distribution, the delivered quality of the optimized encoding configuration is kept the same as that of the baseline configuration. In Fig. 4, we plot the viewing probability of each representation as predicted by our player model in (6). It is seen that the optimized encoding configuration tend to cause the player to spend more time streaming higher representations, thereby compensating for the loss in delivered quality due to reduced encoding bitrates.

We also compared the proposed solution with another baseline configuration. In this baseline configuration, we fixed the encoding CRFs of the 144p and 1080p representations to be 23. Then we selected the encoding bitrates for the other representations such that the achievable rate-quality region given in (5) was maximized. This is similar to the method proposed in [3]. On the 1000 test videos, we found the proposed solution could reduce the average streaming bitrate by 9.45%. Fig. 5 illustrated the baseline and the optimized encoding configurations of an example video. It can be seen that the optimized encoding bitrate is higher than the baseline configurations in low resolution presentations, including 240p, 360p, and 480p. For 720p and 1080p rep-

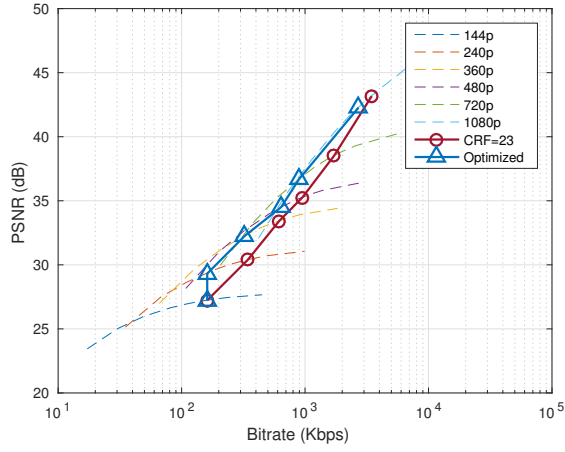


Figure 3: The encoding bitrates when CRF is fixed to 23 (○) and the corresponding encoding bitrates optimized by the proposed method (△) for an example video.

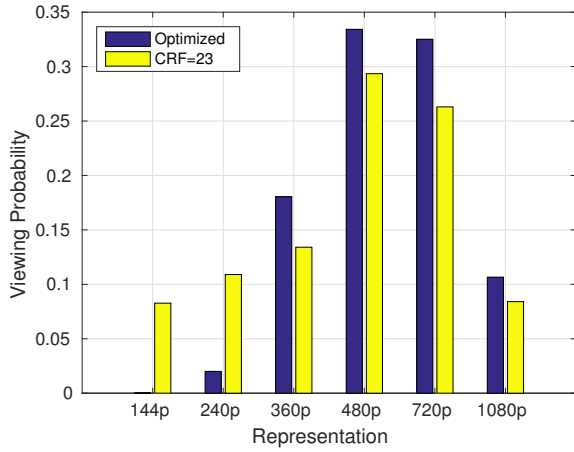


Figure 4: The viewing probability distribution of an example video estimated by the proposed player model.

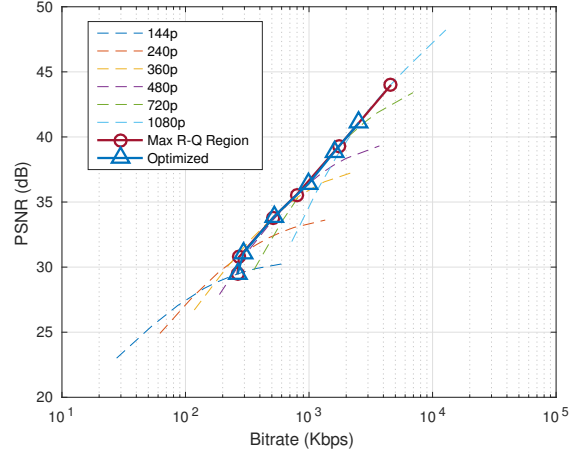


Figure 5: The encoding parameter configurations that maximized the achievable rate quality region (○) and the corresponding encoding bitrates optimized by the proposed method (△) for an example video.

resentation, the optimized encoding bitrates are lower than that of the baseline. As shown in Fig. 6, with the optimized configuration, players would spent more time on the 1080p representations, preserving delivered quality.

In the next section, we validate the effectiveness of our encoding bitrate configuration via experiments on our video streaming platform.

4. EXPERIMENTAL RESULTS

We randomly selected 5,000 1080p videos for a real-world experiment. The selected videos were 1 minute to 20 minutes long. Two pairs of treatments were applied, each containing a baseline encoding bitrate configuration and the corresponding optimized configuration. As in our simulations, the first pair of treatments used the default ffmpeg CRF of 23 as the baseline. In the second pair, the baseline configuration fixed the encoding CRF for 144p and 1080p at 23. The bitrates of other representations were configured such that the achievable rate quality region was maximized.

For each treatment, the bitrate configurations were applied on every 5-second chunks of the videos in order to incorporate the variations in the spatial-temporal characteristics of videos. The playback statistics, which included total watch time and average video streaming bitrates, were collected to evaluate performance. In the following, we will first report the bitrate changes in each treatment pair, and then summarize the overall playback statistics.

We define the relative change in encoding bitrate as $\frac{r^*}{r_0} - 1$, where r^* and r_0 are the encoded bitrates of the optimized configuration and the corresponding baseline configuration, respectively. Fig. 7(a) shows the boxplot of relative encoded bitrate changes at different resolutions for the first pair of treatments. It can be seen that the proposed method selects lower encoding bitrates for almost all resolutions. This is especially true for 240P, where the median of the relative change was -36.5%. Fig. 7(b) shows the relative encoding bitrate changes against the baseline that maximizes the achievable rate-quality region. In this case, the optimizer

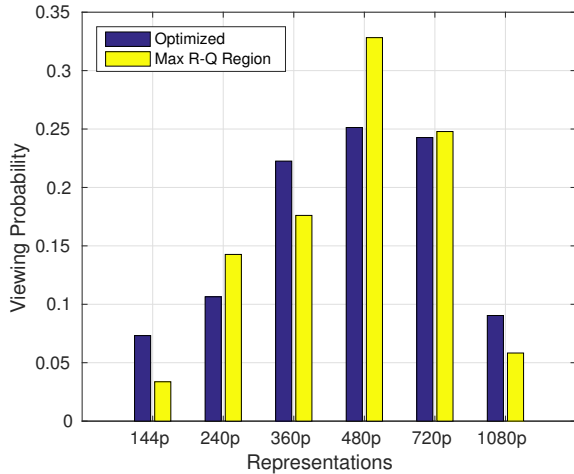


Figure 6: The viewing probability distribution of an example video estimated by the proposed player model.

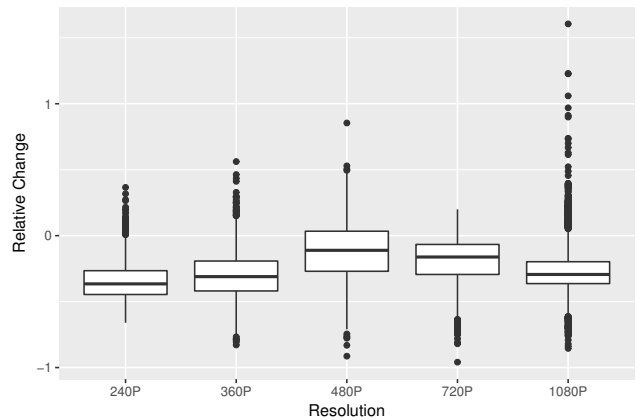
increases the median bitrates of 240p, 360p, and 480p by 13.0%, 23.2% and 16.5%, respectively. The median bitrates of 720p and 1080p are reduced by 10.8%, and 48.5%, respectively.

From the collected statistics, the changes in bitrate affected the distribution of watch time across resolutions. Fig. 8 compares watch time distributions at different resolutions. Fig. 8(a) shows that, for the first pair of treatments, the watch time of the optimized configuration generally shifts towards higher resolutions. For the second pair of treatments, watch time shifts are observed as well. As can be seen in Fig. 8(b), the watch time of the 240p representation is slightly reduced while that of the 1080p representations is increased.

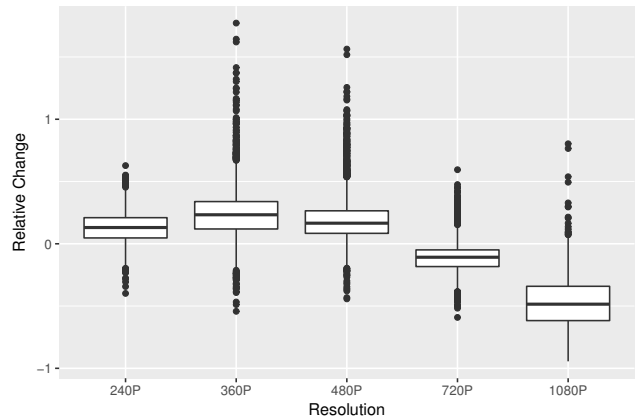
We calculated the average streaming bitrate and observed that the proposed method saves 14.37% in average bitrate against the configuration using a fixed CRF of 23, and 9.65% against the configurations that maximize the rate-quality region.

The optimized encoding bitrate configurations is designed to reduce average streaming bitrate without affecting the quality of experience of viewers. Table. 1 and Table. 2 summarize the total watch time and average delivered PSNR of each configuration. Measured quality loses 0.20 dB in PSNR against the configuration using a fixed CRF of 23, and gains 0.05 dB against the configurations that maximize the rate-quality region. For both treatment pairs, we conducted two-sided log-transformed t-tests at 95% confidence on watch time and average delivered PSNR with. There is no significant changes in both metrics.

We also measured the representation switching rate, the join latency that is the latency from the moment when a playback request is sent to the moment when the video starts to play and the mean time between rebufferings. As shown in Table. 1, comparing with the configuration using a fixed CRF of 23, our method reduces the rate switching rate and initial delay significantly. This can improve the overall QoE of users. The proposed method also increased the mean time between rebuffering events but the improvement is statisti-

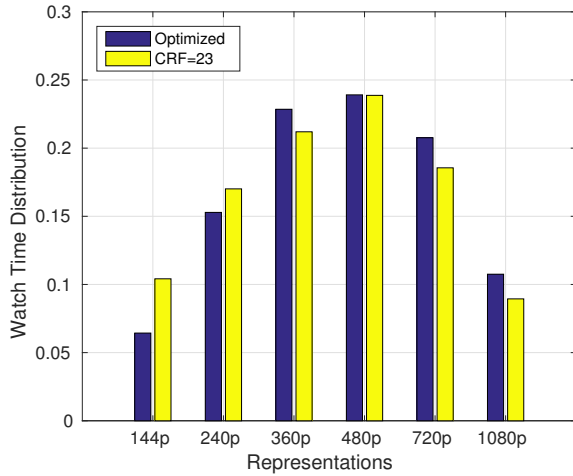


(a)

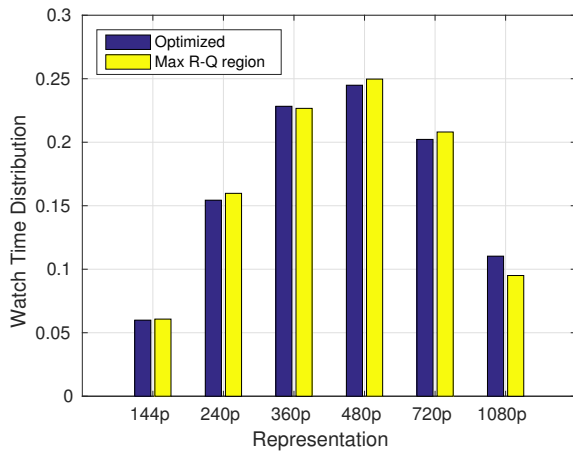


(b)

Figure 7: Relative encoding bitrate changes per resolution under the two experimental settings: (a) Fixed CRF = 23 as the baseline. (b) Maximizing rate-quality region as the baseline.



(a)



(b)

Figure 8: Watch time distribution at different resolutions under the two experiment settings: (a) Fixed CRF as the baseline. (b) Maximizing rate-quality region as the baseline

cally insignificant. Comparing with the configurations that maximize the rate-quality region, our method reduced the rate switching rate slightly. Its impact on initial delay and mean time between rebufferings is statistically insignificant.

5. CONCLUSIONS AND FUTURE WORK

We propose a mathematical model for the adaptation behavior of players in HTTP-based video streaming. In addition to the rate-quality characteristics of videos, the model also incorporates the statistical distribution of available bandwidth and viewport sizes of players. Based on the model, we implemented a method to optimize the encoding bitrates of the video representations. Both numerical simulations and experimental results demonstrated that the proposed method can save 9.6% on the average video streaming bandwidth without degrading users' quality of experience or average video delivered quality.

The optimization method presented in this paper is based on global bandwidth and viewport size distributions. However, a video might be popular in a certain geographic region where the bandwidth/viewport distributions differ from the global ones. As part of future work, we will investigate potential gains of incorporating local bandwidth and viewport distributions to our approach.

6. REFERENCES

- [1] Nlopt introduction. http://ab-initio.mit.edu/wiki/index.php/Nlopt_Introduction.
- [2] Cisco visual networking index. <http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.html>, 2016.
- [3] J. D. Cock, Z. Li, M. Manohara, and A. Aaron. Complexity-based consistent-quality encoding in the cloud. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 1484–1488, Sept 2016.
- [4] M. R. Group. ISO/IEC FCD 23001-6 part 6: Dynamics adaptive streaming over HTTP (DASH). <http://mpeg.chiariglione.org/working&underscore;documents/mpeg-b/dash/dash-dis.zip>, 2011.
- [5] C. Kreuzberger, B. Rainer, H. Hellwagner, L. Toni, and P. Frossard. A comparative study of dash representation sets using real user characteristics. In *Proceedings of the 26th International Workshop on Network and Operating Systems Support for Digital Audio and Video*, page 4. ACM, 2016.
- [6] C. Li, L. Toni, P. Frossard, H. Xiong, and J. Zou. Complexity constrained representation selection for dynamic adaptive streaming. In *2016 Visual Communications and Image Processing (VCIP)*, pages 1–4, Nov 2016.
- [7] H. Li, B. Li, and J. Xu. Rate-distortion optimized reference picture management for high efficiency video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12):1844–1857, Dec 2012.
- [8] J. R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand. Comparison of the coding efficiency of video coding standards - including high efficiency video coding (HEVC). *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12):1669–1684, Dec 2012.

Metrics	CRF=23	Optimized	Gain	Statistical Significance
Watch Time	178,201	177,499	-0.39%	insignificant
Average Quality	37.93	37.73	-0.53%	insignificant
Normalized Average Bitrate	1	0.855	-14.37%	significant
Normalized Adaptive Switch Rate	1	0.8699	-13.01%	significant
Normalized Join Latency	1	0.9791	-2.09%	significant
Normalized Mean Time Between Rebuffers	1	0.9904	0.96%	insignificant

Table 1: Performance comparing with the baseline where CRF of 23 is applied to all representations. Watch Time: total watch time in hours, Average quality: measured quality in scaled PSNR (dB). Normalized Average Bitrate: measured average bitrate divided by the average bitrate of the baseline. Adaptive switch rate: the number of representation switches per hour of video playbacks. Join latency: the latency from the moment when a playback request is sent to the moment when the video starts to play. Mean time between rebuffering: the average time between two playback rebuffering events.

Metrics	CRF=23	Optimized	Gain	Statistical Significance
Watch Time	177,265	177,797	0.30%	insignificant
Average Quality	37.49	37.54	0.13%	insignificant
Normalized Average Bitrate	1	0.903	-9.65%	significant
Normalized Adaptive Switch Rate	1	0.8699	-0.76%	significant
Normalized Join Latency	1	0.9791	0.07%	insignificant
Normalized Mean Time Between Rebuffers	1	0.9904	-0.74%	insignificant

Table 2: Performance comparing with the baseline where bitrates are configured to maximize the rate quality region. Watch Time: total watch time in hours. Average quality: measured quality in scaled PSNR (dB). Normalized Average Bitrate: measured average bitrate divided by the average bitrate of the baseline.

- [9] R. Pantos and E. W. May. HTTP live streaming. IETF internet draft. <https://tools.ietf.org/html/draft-pantos-http-live-streaming-20>, 2016.
- [10] B. Rainer, S. Petschornig, C. Timmerer, and H. Hellwagner. Statistically indifferent quality variation: An approach for reducing multimedia distribution cost for adaptive video streaming services. *IEEE Transactions on Multimedia*, 19(4):849–860, 2017.
- [11] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hobfeld, and P. Tran-Gia. A survey on quality of experience of HTTP adaptive streaming. *IEEE Communications Surveys & Tutorials*, 17(1):469–492, 2015.
- [12] K. Svanberg. A class of globally convergent optimization methods based on conservative convex separable approximations. *SIAM Journal on Optimization*, 12(2):555–573, Feb. 2002.
- [13] L. Toni, R. Aparicio-Pardo, K. Pires, G. Simon, A. Blanc, and P. Frossard. Optimal selection of adaptive streaming representations. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 11(2s):43, 2015.
- [14] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004.
- [15] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan. Rate-constrained coder control and comparison of video coding standards. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(7):688–703, July 2003.