# QoE-Driven Cache Management for HTTP Adaptive Bit Rate (ABR) Streaming over Wireless Networks

Weiwen Zhang[1], Yonggang Wen[1], Zhenzhong Chen[2], and Ashish Khisti[3]

[1]School of Computer Engineering, Nanyang Technological University, Singapore
[2]MediaTek USA Inc., USA
[3]School of Electrical and Computer Engineering, University of Toronto, Canada
{wzhang9, ygwen}@ntu.edu.sg, zzchen@ieee.org, akhisti@comm.utoronto.ca

***Abstract*—In this paper, we investigate the problem of how to cache a set of media files with optimal streaming rates, under HTTP adaptive bit rate streaming over wireless networks. The design objective is to achieve the optimal expected QoE under a limited storage budget, which is measured by the logarithmic relation between the required bit rate and the actual streaming bit rate. We formulate the content cache management of streaming files as a constrained optimization problem. Lagrange multiplier method is employed, and we obtain the numerical solution of the optimal streaming files. Particularly, we characterize the properties of the solution, and find there is a fundamental phase change in the optimal solution as the number of cached files grows. Moreover, the simulation results indicate that with the increase of cache size, more copies of different bit rate should be cached for a better QoE. Our comprehensive investigation reveals insightful guidelines to provide HTTP ABR streaming services over wireless networks.**

***Keywords*-QoE; Adaptive Bit Rate streaming; content cache management; optimization.**

## I. Introduction

Mobile video consumption, owing to the rapid adoption of smartphones, is fueling a dramatic growth of mobile data traffic lately. Cisco VNI report [1] predicted that the mobile data traffic will increase 26 times between 2010 and 2015, among which the leading contributor is the video traffic generated by the mobile users worldwide. This growth of mobile video experience, however, is in tandem with a huge concern of the user experience, resulted from the inherent nature of stochastic warless channels (e.g., multi-path and shadowing fading effects). As a result, it has become a technical challenge to provide a high Quality of Experience (QoE) for the rising demand of video streaming over wireless network.

Recently HTTP adaptive bit rate (ABR) streaming [2], [3], [4] is emerging as a prominent solution to improve the user experience and the network resource utilization in mobile video. In practical systems (e.g., Cisco's CDS-IS [5]), the streaming engine switches, in a real-time manner, among a set of video files (for the same content) with different playback rates, in response to the channel condition and the device model [6]. However, such a practical solution is stressed by the huge growth of user-generated contents. It has been observed in Cisco's deployment that transcoding a large set of video contents into files with different playback rates on the streaming engine can have the storage to be filled up rapidly.

This research aims to address the technical challenge of content cache management for HTTP ABR streaming. Previous studies on HTTP ABR streaming mainly focus on mechanisms to adjust the streaming bit rate for varying network conditions. In [3], [4], [6], [7], [8], various solutions were proposed to improve the performance of the HTTP ABR streaming services, with an objective to optimize the Quality of Service (QoS) (e.g., the reduction of end-to-end delay, better buffer management, bandwidth savings, or higher resource utilization). On a different track, research efforts [9], [10], [11], [12] have been devoted to investigating QoE-aware adaptation schemes. These solutions aim to maximize content provisioning and network resources under the QoE requirement, or maximize the QoE under the bandwidth constraints. Neither approach considers the storage aspect of HTTP ABR streaming. In this research, we extend the research scope of HTTP ABR streaming with QoE-driven content cache management.

In this paper, we aim to develop an optimal scheme for QoE-driven content cache management in HTTP ABR streaming. Our design objective is to provide the best possible QoE for the mobile users, while avoiding the content storage to be filled up rapidly. Specifically, the content provided by the original server is transformed into HTTP streaming formats on the streaming engine in advance. When a client consumes some content with a required streaming rate, the streaming engine will reply a content file with the closest bit rate to requested, hoping that the distortion is acceptable by the client. In this case, the research problem is how to choose a set of video files with different playback rates, for a given storage budget, so as to maximize the expected QoE for a pool of mobile users.

Our contributions are multi-fold. First, we formulate the content cache management for HTTP ABR streaming into a convex optimization problem, thus giving this engineering problem an analytical framework. Second, we apply the Lagrange multiplier method to solve the optimization problem and provide engineering guidelines for how to cache a set of media files with different playback rates. Third, we characterize the optimal solution analytically and show that the optimal solution would go through a phase change as the number of cached files for one content grows. Finally, combining experimental results, we verify our analytical framework in a great

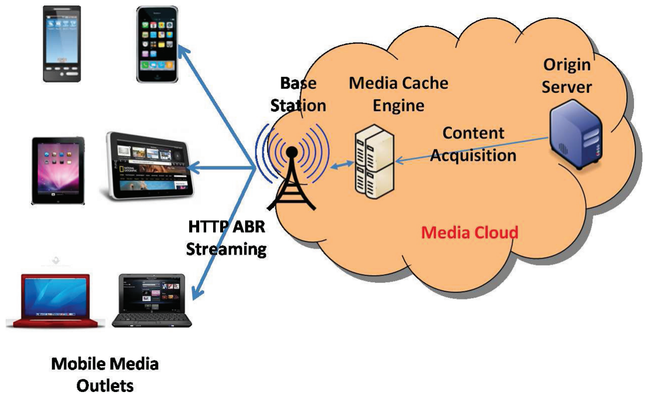



Fig. 1. A schematic diagram for HTTP ABR streaming system over the wireless network: contents acquired from original servers are transcoded into a set of HTTP ABR files with different playback rates, and cached in the streaming engine.

accuracy. Our comprehensive investigation reveals insightful guidelines to provide HTTP ABR streaming services over commercially-available platforms, e.g., CDS-IS from Cisco.

The rest of the paper is organized as follows. Section II presents the system model and problem formulation. In section III, a mathematical solution is given for the optimal content cache management of the adaptive bit rate streaming. Numerical simulations are provided in section IV. Finally, section V concludes the paper and suggests future work.

## II. System Model and Problem Formulation

In this section, we first describe a generic architecture for HTTP ABR streaming system, modeled after a real system deployment. Following that, we present the models for the HTTP ABR streaming system, including a user-request model, a QoE model and a content caching model. Using these models, we formulate the content cache management as a constrained optimization model.

### A. System Architecture

A generic HTTP ABR streaming system, adapted from the real deployment, is illustrated in Figure 1. It consists of three parts, including a content origin server, a media cache engine and a pool of mobile media outlets. The content origin server stores media files in their original format and transfers them to the intermediate media cache engine. In any media cache engine, the original content file is transformed into HTTP streaming format (e.g., SmoothHD, Adobe Zeri, MoveNet, etc) for adaptive-bit-rate streaming. Specifically, a few files of the same content are created locally and stored at the content cache. Each file corresponds to a different streaming rate. Moreover, the format of each file includes two parts: i) a manifest file for meta data, and ii) a set of media content files each of which contains video content of a fixed playback duration (e.g., 2 seconds). When the user requests some content, the streaming engine replies with a required streaming rate, which is determined by the physical capability of the media outlet and the network status. Based on the required streaming rate, the content engine will stream the content from a chosen file among all available copies cached locally. If a particular playback rate is not immediately available, the most logic approach is to stream the content with the closest rate from below.

### B. System Models

*1) User Request Model:* User request from a mobile media outlet is characterized by a required playback rate, denoted as $r$. The required playback rate depends on both the physical capability of the media outlet (e.g., screen size) and the network channel condition (e.g., available bandwidth). Normally, $r$ is modeled as a random variable with a specific probability density function of $f_R(r)$ for $r \in [r_0, r_n]$, where $r_0$ and $r_n$ are the lower and upper bounds of $r$, respectively. In this paper, we assume the user request follows a uniform distribution, with the following probability density function,

$$f_R(r) = \frac{1}{r_n - r_0}, r \in [r_0, r_n]\,. \tag{1}$$

The assumption simplifies our analytical derivation and the obtained closed-form solutions can be adapted to provide operational guidelines for practical HTTP ABR systems.

*2) QoE Model:* QoE is a subjective measurement of a media consumer's experience with a video. In this paper, we assume a QoE model, in which the user's experience depends on two system parameters, including the required playback rate of $r$ and the actual playback rate of $r_i$. Practically, $r_i \leq r$.

In [13], user experience follows the logarithmic laws, and QoE function can be modeled in the logarithmic form for applications of file downloading and web browsing. As such, in this paper, we adopt the QoE model as the logarithmic function between $r_i$ and $r$ in the HTTP ABR scheme, which is specified as

$$Q(r_i, r) = a_1 ln\frac{a_2 r_i}{r}, \tag{2}$$

where the constant parameters $a_1$ and $a_2$ are both positive, and they can be different for videos of different features.

*3) Content Caching Model:* In the HTTP ABR streaming system, the cache engine transcodes the received media file into the HTTP ABR format. Specifically, the received content file is transformed into a set of content files, each of which represents one playback rate.

We assume that $n$ files with different playback rates are created in the cache, and each file has a streaming rate of $r_i > 0, i = 0, 1, \cdots, n-1$. Without loss of generality, we assume that $r_0 < r_1 < \cdots < r_{n-1}$. Moreover, the file size is assumed to be an affine function, i.e.,

$$g(r_i) = ar_i + b, \tag{3}$$

where $ar_i$ represents the size of the media content stored on the server and $b$ represents the meta data. Then, the total storage capacity required at the content cache engine is

$$C_{tot} = \sum_{i=0}^{n-1} g(r_i). \tag{4}$$

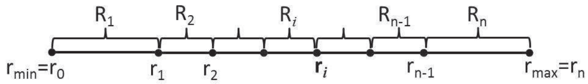



Fig. 2. Illustration of playback rates for cached media files and request playback rates.

### C. Problem Formulation

The research problem is how to maximize the expected QoE metrics by optimally caching HTTP ABR content files. Obviously, it is beneficial to cache streaming files in different bit rates as many as possible for a higher QoE. However, in a real system deployment, the storage space in the content cache can be filled up rapidly, in order to meet all the QoE requirement. It does not allow numerous files to be cached for streaming. As such, one also aims to control the storage budget, by optimally choosing a subset of playback rates, for which a copy of the media content is cached.

Figure 2 shows the scenario of the cached playback rates and the requested playback rates. There are $n$ copies to be cached, i.e., $r_0, r_1, ..., r_{n-1}$. If the requested bit rate $r$ falls within the region $R_{i+1}$ (i.e., $r \in [r_i, r_{i+1})$), we will assign $r_i$ as the replied bit rate. As such, the problem is to find $r_1, r_2, ..., r_{n-1}$ between the given minimal rate $r_0$ and the maximal rate $r_n$ to maximize the average QoE metrics, while respecting a given storage budget constraint. Mathematically, the problem can be formulated as the following constrained optimization problem,

$$\max_{n,\vec{r_i}} \quad \mathbb{E}[Q], \tag{5}$$

$$\text{s.t.} \quad C_{tot} \leq C, \tag{6}$$

where $\vec{r_i} = (r_1, r_2, \cdots, r_{n-1})$. The expectation is taken over the distribution of user request (i.e., $f_R(r)$). Therefore, the optimization problem can re-written as

$$\max_{\vec{r_i}} \quad F = \sum_{i=0}^{n-1} \int_{r_i}^{r_{i+1}} Q(r_i, r) f_R(r) dr, \tag{7}$$

$$\text{s.t.} \quad \sum_{i=0}^{n-1} (ar_i + b) \leq C, \tag{8}$$

$$r_i - r_{i+1} < 0 (i = 0, 1, ..., n-1). \tag{9}$$

Replacing the QoE function with its reverse, we transfer the maximization problem into the following minimization problem,

$$\min_{\vec{r_i}} \quad D = -\sum_{i=0}^{n-1} \int_{r_i}^{r_{i+1}} Q(r_i, r) f_R(r) dr, \tag{10}$$

$$\text{s.t.} \quad \sum_{i=0}^{n-1} (ar_i + b) \leq C, \tag{11}$$

$$r_i - r_{i+1} < 0 (i = 0, 1, ..., n-1). \tag{12}$$

## III. Optimal QoE-Driven Cache Management for HTTP ABR Streaming

In this section, we first use the Lagrange multiplier method to solve the constrained optimization problem, and then characterize the optimal solution for content cache management. Our investigation reveals a fundamental phase in exploring the available storage budget to maximize the offered QoE metrics.

### A. Derivation of Optimal Solution

The Lagrangian function of the optimization problem (10) is given by

$$\begin{aligned} L(r_i, \lambda, \mu_i) = D + \lambda[\sum_{i=0}^{n-1} (ar_i + b) - C] \\ + \sum_{i=0}^{n-1} \mu_{i+1}(r_i - r_{i+1}), \end{aligned} \tag{13}$$

where

$$D = -\sum_{i=0}^{n-1} \int_{r_i}^{r_{i+1}} a_1 ln \frac{a_2 r_i}{r} \frac{1}{r_n - r_0} dr.$$

First, it can be shown that it is a convex optimization problem, because the constrained set is convex and the hessian matrix ($H$) of the objective function is positive definite. The latter is verified by the fact that the determinant of all the principal minors of $H$ are greater than zero, i.e.,

$$|H_k| = \frac{a_1}{(r_n - r_0) \prod_{i=1}^{k} r_i} \left[ 1 + r_{k+1} \sum_{i=1}^{k} \frac{1}{r_i} \right] > 0. \tag{14}$$

Second, the KKT conditions are necessary and sufficient for a global minimum of $D$, subject to the inequality constraints, because it is a convex optimization problem. To solve the optimization problem, we introduce a set of slack variables $z$ and $d_i$ as follows,

$$\begin{aligned} L(r_i, \lambda, z, \mu_i, d_i) = D + \lambda[\sum_{i=0}^{n-1} (ar_i + b) + z^2 - C] \\ + \sum_{i=0}^{n-1} \mu_{i+1}(r_i - r_{i+1} + d_{i+1}^2). \end{aligned} \tag{15}$$

Using the KKT conditions, we have the following equations,

$$\begin{aligned} \frac{\partial L}{\partial r_i} &= -\frac{a_1}{r_n - r_0}[ln \frac{r_{i-1}}{r_i} + \frac{r_{i+1}}{r_i} - 1] + \lambda a + (\mu_{i+1} - \mu_i) \\ &= 0, \end{aligned} \tag{16}$$

$$\frac{\partial L}{\partial \lambda} = \sum_{i=0}^{n-1} (ar_i + b) + z^2 - C = 0, \tag{17}$$

$$\frac{\partial L}{\partial z} = 2\lambda z = 0, \tag{18}$$

$$\frac{\partial L}{\partial \mu_i} = r_{i-1} - r_i + d_i^2 = 0, \tag{19}$$

$$\frac{\partial L}{\partial d_i} = 2\mu_i d_i = 0, \tag{20}$$

$$\lambda \geq 0, \tag{21}$$

$$\mu_i \geq 0. \tag{22}$$

Since $d_i \neq 0$, the variables $\mu_i$ must be zero. As a result, organizing these equations together, we obtain

$$\begin{cases} \frac{r_2}{r_1} - ln\frac{r_1}{r_0} - 1 - \frac{\lambda a(r_n - r_0)}{a_1} = 0 \\ \frac{r_3}{r_2} - ln\frac{r_2}{r_1} - 1 - \frac{\lambda a(r_n - r_0)}{a_1} = 0 \\ \ldots\ldots \\ \frac{r_n}{r_{n-1}} - ln\frac{r_{n-1}}{r_{n-2}} - 1 - \frac{\lambda a(r_n - r_0)}{a_1} = 0 \\ \sum_{i=0}^{n-1}(ar_i + b) + z^2 - C = 0 \\ 2\lambda z = 0 \\ r_0 - r_1 + d_1^2 = 0 \\ r_1 - r_2 + d_2^2 = 0 \\ \ldots\ldots \\ r_{n-1} - r_n + d_n^2 = 0. \end{cases} \tag{23}$$

Therefore, to solve the optimization problem (10), it is equivalent to find the solution of the system of nonlinear equations (23), in which there are $2n$+1 equations and $2n$+1 unknowns.

Finally, to solve this system of nonlinear equations, we can minimize $s$, where $s$ is defined as the sum of the square of functions on the left hand sides of Eq. (23), as follows:

$$\min s = \sum_{i=1}^{n-1}(\frac{\partial L}{\partial r_i})^2 + (\frac{\partial L}{\partial \lambda})^2 + (\frac{\partial L}{\partial z})^2 + \sum_{i=1}^{n}(\frac{\partial L}{\partial \mu_i})^2. \tag{24}$$

Notice that this is an unconstrained optimization. It can be solved by the method of trust-region-dogleg [14]. In this paper, we adopt the solver (i.e., *fsolve*) provided by Matlab, which implements the trust-region-dogleg algorithm. More details about the numerical solutions will be given in Section IV-B.

### B. Characterization of Optimal Solution

In this section, we characterize the optimal solution for QoE-driven content cache management. It can be shown that a fundamental phase change exists in the optimal solution.

Considering $2\lambda z = 0$ in (23), we can have two possibilities, each of which corresponds to a phase in the optimal solution.

*1) Phase I ($\lambda = 0, z \neq 0$):* This is the case when the available storage budget is not fully utilized, with the remainder of $z^2$. For the first $n-1$ equations in (23), there can be at most one solution for $r_i$ ($i = 1, 2, ..., n-1$); this can be proved by contradiction. It suggests that once we find the solution of $r_i$, it must be optimal and unique cache management for the streaming files.

Plugging the optimal condition of (23) into the objective function Eq. (7), we obtain

$$F^* = a_1(lna_2 + 1) + \frac{a_1}{r_n - r_0}[r_1^* + r_n(ln\frac{r_{n-1}^*}{r_n} - 1)], \tag{25}$$

where $r_1^*$ and $r_{n-1}^*$ are their optimal values.

*Proposition 3.1:* For *Phase I*, the optimal solution exists the following properties, including

(a) for a specific $n$, the increase of cache size will still produce the same values of $r_i$ and QoE, since $C$ is irrelevant to the first $n-1$ equations in (23);

(b) $F^*$ is an increasing function of $n$, i.e., the optimal QoE increases with $n$.

First, Proposition 3.1(a) suggests that when there are more cache storage available, we should increase the value of $n$, i.e., to cache more copies of bit rate streaming files in order to improve the QoE.

Second, Proposition 3.1(b) can be explained as follows. Consider two cache managements: $k$ copies of files, $\vec{r_i} = (r_1, r_2, \cdots, r_k)$, and $k+1$ copies of files, $\vec{r_i'} = (r_1', r_2', \cdots, r_{k+1}')$. The streaming rates follow that $r_i' < r_i$ ($i = 1, 2, ..., k$), and $r_{k+1}' > r_k$ (this can be proved by contradiction). Then plugging the optimal condition of (23) into $F^*(k)$ and $F^*(k+1)$ respectively, we can evaluate the difference between these two terms, and finally obtain $F^*(k) < F^*(k+1)$.

Moreover, since $\sum_i r_i$ is increasing with $n$, the remainder $z^2$ will decrease with the increase of $n$, approaching to be zero eventually, which is *Phase II* as below.

*2) Phase II ($\lambda > 0, z = 0$):* This is the case when the available storage budget is fully utilized. For the first $n-1$ equations in (23), there can be at most one solution for $r_i$ ($i = 1, 2, ..., n-1$) and $\lambda$; this can be proved by contradiction. It suggests that once we find the solution of $r_i$, it must be optimal and unique cache management for the streaming files.

Plugging the optimal condition of (23) into the objective function Eq. (7), we obtain

$$F^* = a_1(lna_2+1) + \frac{a_1}{r_n - r_0}[r_1^* + r_n(ln\frac{r_{n-1}^*}{r_n} - 1)] + \lambda^* a \sum_{i=1}^{n-1} r_i^*,$$

where $r_i^* (i = 1, 2, ..., n-1)$ and $\lambda^*$ are their optimal values.

Based on the analysis of *Phase I* and *Phase II*, we can find that there is a fundamental property in the optimal solution. As $n$ increases, the optimal solution will experience a phase change, going from *Phase I* to *Phase II*. That is, when the number of cached files is small, optimal solution is obtained by partially using the storage budget; when the number of cached files increases, the optimal solution will be obtained by fully using the storage budget.

## IV. Numerical Analysis and Results

In this section, we provide the numerical results of the optimal content cache management for HTTP ABR streaming. First, we identify the QoE functions by approximation for three types of videos. Then, the optimal content streaming cache managements are given accordingly.

### A. Verification of QoE Model

QoE can be measured by the Mean opinion score (MOS) scaling from 1 to 5, where 5 represents the service is excellent and 1 represents bad. We have conducted an experimental QoE study for Scalable Video Coding by evaluating the scores for the adapted SVC bitstreams. The SVC bitstreams

TABLE I
RATE (KBPS) AND MOS

| City | | Crew | | Soccer | |
|---|---|---|---|---|---|
| rate | MOS | rate | MOS | rate | MOS |
| 5332.2 | 5.0 | 6520.8 | 5.0 | 5870 | 5.0 |
| 4212.2 | 5.0 | 4584.5 | 5.0 | 4528.9 | 5.0 |
| 2069.7 | 5.0 | 2428.8 | 4.0 | 2467.4 | 5.0 |
| 896.7 | 4.0 | 1275 | 3.5 | 1188.4 | 4.0 |
| 437.9 | 3.5 | 622.3 | 3.0 | 600.4 | 3.0 |
| 388.1 | 3.0 | 466.6 | 3.0 | 463 | 3.0 |
| 335.5 | 3.0 | 315.7 | 2.4 | 318.8 | 2.6 |
| 137.7 | 2.0 | 178.3 | 2.0 | 154 | 2.0 |
| 44.6 | 1.5 | 72.2 | 1.4 | 63.4 | 2.0 |
| 38.4 | 1.0 | 48.2 | 1.0 | 39.1 | 1.4 |

TABLE II
PARAMETERS FOR THE QOE FUNCTIONS

| | $(a_1,a_2)$ | $(r_0,r_n)$ | $\|MOS-Q\|_2^2$ |
|---|---|---|---|
| City | (0.9511, 157.9) | (38.4, 2069.7) | 0.2421 |
| Crew | (0.8205, 311.9) | (48.2, 4584.5) | 0.2103 |
| Soccer | (0.7955, 286.9) | (39.1, 2467.4) | 0.8187 |

were adapted for different bit rate and subjective tests were conducted accordingly. The videos were compressed by JSVM [15] and the rate adaptation was performed by the Bitstream Extractor in JSVM. In the subjective tests, 22 non-expert viewers with normal or corrected-to-normal vision acuity participated in the single-stimulus test for evaluation based on the Adjectival Categorical Judgment Methods in [16]. The viewing conditions, facility setup and data screening followed the ITU recommendations [16], [17]. Table I provides the rate-MOS results for the selected videos (i.e., city, crew, soccer). We adopt these results directly to verify the QoE function (2).

In order to find the values of $a_1$ and $a_2$ in the QoE function, we minimize $\|MOS-Q\|_2^2 = \sum_i (MOS_i - Q(r_i, r))^2$ by the least-squares method, using curve fitting over $ln\frac{r_i}{r}$ and $Q$. Table II lists the values of $a_1$, $a_2$, $r_0$, $r_n$ and $\|MOS-Q\|_2^2$ for three types of videos. Figure 3 indicates that the fitting functions of QoE approximate well with $r_i$ for all the videos.

### B. Optimal Content Cache Management

As an example, we consider the case of the city video with the cache size $C = 3000$ KB. We set the parameters in the content caching model (3) as $a = 1$ and $b = 0.5$. Given an $n$, we can find the optimal $r_i$ for a video. The optimal results of the content cache management are given in Table III. It can be shown that, when $n < 5$, the constraint of the cache size is inactive, which is the case of *Phase I*; and when $n \geq 5$, the storage budget is fully utilized, which is the case of *Phase II*. The optimal QoE is achieved at $n = 8$. Therefore, it follows that a phase change occurs and the optimal QoE is in *Phase II*, which agrees with the analysis in Section III-B.

We also plot QoE as a function of cache size ($C$) and number of copies ($n$) for the three videos in Figures 4 and 5. Several observations can be drawn from these two figures.

First, for a specific $n$, the curves for different cache size in Figure 4 overlap if $n$ is small (i.e., in *Phase I*), and the shapes of the curves in Figure 5 appear to be a line when $C$ increases. As a result, the increase of cache size may not result in higher QoE, which agrees with Proposition 3.1.

Second, in order to enhance the QoE, $n$ should be increased. This can be explained by Figure 5 in which the curve of QoE with larger $n$ is basically above the one with smaller $n$.

In addition, the optimal QoE is not a monotonically increasing function of $n$. In Figure 4, for all of the three videos, the optimal QoE increases initially, but there is a slight drop when $n$ is relatively large (i.e., in *Phase II*). This is because, as $n$ increases, $r_{n-1}$ should decrease to satisfy the constraint of the limited cache, which largely brings down the QoE for the request bit rate in the region $R_n$. More details are omitted here due to the limited page length. Finally, this observation suggests that when we find a $k$ such that $F^*(k) > F^*(k+1)$, $k$ will be the optimal number of copies for the maximum QoE.

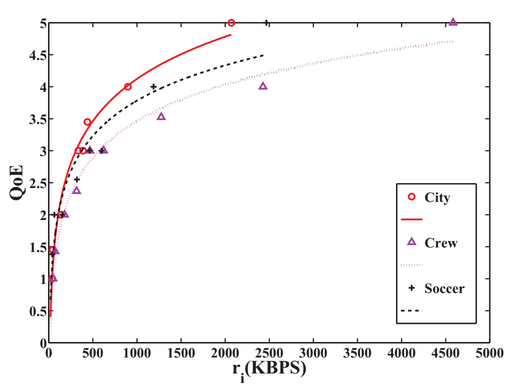


Fig. 3. Approximation of QoE function in logarithmic relation

## V. CONCLUSION

In this paper, we investigated the problem of how to cache a set of media files with optimal streaming rates, under HTTP adaptive bit rate streaming over wireless networks. We provided the mathematical solutions to this problem. We also characterized the properties of the solution, and found that there is a phase change in the optimal solution. Moreover, with more cache size provided, more streaming files with different bit rates should be cached for a better QoE. As future work, we will consider on how to find the optimal number of streaming files ($n$) that can maximize the QoE.

## REFERENCES


[1] Cisco, *Cisco Visual Networking Index: Forecast and Methodology, 2010 - 2015*, 2011.

[2] A. C. Begen, T. Akgul, and M. Baugher, "Watching video over the web," *Internet Computing*, vol. 15, pp. 54–63, 2011.

[3] I. Hofmann, N. Farber, and H. Fuchs, "A study of network performance with application to adaptive HTTP streaming," in *Broadband Multimedia Systems and Broadcasting (BMSB), IEEE International Symposium*, pp. 1–6, 2011.

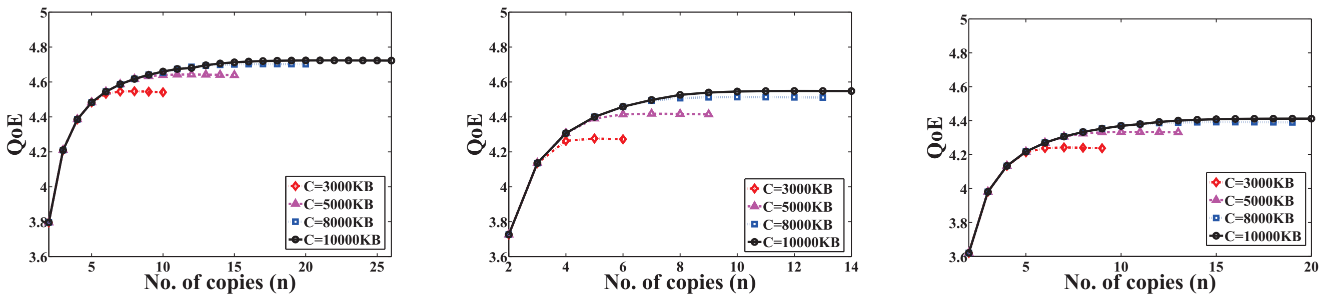
[4] C. Liu, I. Bouazizi, and M. Gabbouj, "Rate adaptation for adaptive HTTP streaming," in *Proceedings of the second annual ACM conference on Multimedia systems*, pp. 169–174, 2011.

[5] Cisco, *Cisco CDS Internet Streaming: Enabling New Video 2.0 Experiences*, 2007.

[6] V. Adzic, H. Kalva, and B. Furht, "Optimized adaptive HTTP streaming for mobile devices," in *Proceedings of the SPIE*, vol. 8135, pp. 81350T–81350T–10, 2011.

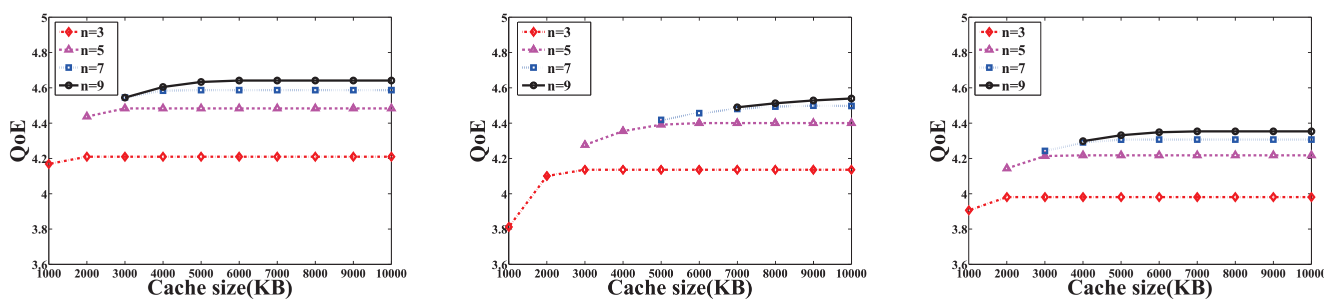
TABLE III
OPTIMAL CONTENT CACHE MANAGEMENT OF $r_i$ FOR CITY VIDEO ($C = 3000$ KB)

| | *Phase I* | | | *Phase II* | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| $z$ | 48.98 | 40.93 | 30.24 | -3.3632e-44 | 0 | 0 | 0 | 0 | 0 |
| $\lambda$ | -1.9e-35 | -2.0e-35 | 5.5e-28 | 1.3e-5 | 4.1e-5 | 6.6e-5 | 7.8e-5 | 8.3e-5 | 8.5e-5 |
| $r_0$ | 38.4 | 38.4 | 38.4 | 38.4 | 38.4 | 38.4 | 38.4 | 38.4 | 38.4 |
| $r_1$ | 561.9155 | 313.3511 | 220.5182 | 183.3648 | 115.1226 | 79.7274 | 59.4591 | 47.0031 | 38.9422 |
| $r_2$ | - | 971.1587 | 605.9671 | 464.8214 | 251.6908 | 149.2230 | 95.3222 | 64.8498 | 46.5990 |
| ... | - | - | 1218.5062 | 883.9555 | 470.7999 | 263.8175 | 156.1283 | 97.2356 | 63.4723 |
| ... | - | - | - | 1426.9583 | 807.2227 | 451.3724 | 259.0700 | 153.8850 | 94.6768 |
| ... | - | - | - | - | 1313.7640 | 757.4690 | 433.2547 | 251.8490 | 149.8223 |
| ... | - | - | - | - | - | 1256.4907 | 727.9343 | 420.6284 | 245.9452 |
| ... | - | - | - | - | - | - | 1226.4315 | 711.0542 | 412.7589 |
| ... | - | - | - | - | - | - | - | 1210.5949 | 701.8355 |
| $r_{n-1}$ | - | - | - | - | - | - | - | - | 1202.5478 |
| $s$ | 6.5e-26 | 1.3e-26 | 3.0e-26 | 2.3e-25 | 2.2e-25 | 6.9e-27 | 2.0e-26 | 2.2e-25 | 2.4e-25 |
| $QoE$ | 3.7961 | 4.2099 | 4.3863 | 4.4825 | 4.5323 | 4.5456 | 4.5470 | 4.5446 | 4.5413 |




(a) City Video (b) Crew Video (c) Soccer Video

Fig. 4. QoE is plot as a function of cache size ($C$) and number of copies ($n$) for three types of videos.




(a) City Video (b) Crew Video (c) Soccer Video

Fig. 5. QoE is plot as a function of cache size ($C$) and number of copies ($n$) for three types of videos.


[7] S. Akhshabi, A. C. Begen, and C. Dovrolis, "An experimental evaluation of rate-adaptation algorithms in adaptive streaming over HTTP," in *Proceedings of the second annual ACM conference on Multimedia systems*, pp. 157–168, 2011.

[8] T. Lohmar., T. Einarsson., P. Frojdh., F. Gabin, and M. Kampmann, "Dynamic adaptive HTTP streaming of live content," in *World of Wireless, Mobile and Multimedia Networks (WoWMoM), IEEE International Symposium*, pp. 1–8, 2011.

[9] A. B.Reis, J. Chakareski, A. Kassler, and S. Sargento, "Distortion optimized multi-service scheduling for next-generation wireless mesh networks," in *INFOCOM IEEE Conference on Computer Communications Workshops*, pp. 1–6, 2010.

[10] S. Thakolsri, W. Kellerer, and E. Steinbach, "QoE-based rate adaptation scheme selection for resource-constrained wireless video transmission," in *ACM Multimedia*, pp. 783–786, 2010.

[11] A. Khan, L. Sun, E. Jammeh, and E. Ifeachor, "Quality of experience-driven adaptation scheme for video applications over wireless networks," *Special Issue on Video Communications over Wireless Networks*, pp. 1337–1347, 2009.

[12] A. Khan, I.-H. Mkwawa, L. Sun, and E. Ifeachor, "QoE-driven sender bitrate adaptation scheme for video applications over IP multimedia subsystem," in *IEEE ICC*, pp. 1–6, 2011.

[13] P. Reichl, B. Tuffin, and R. Schatz, "Logarithmic laws in service quality perception: where microeconomics meets psychophysics and quality of experience," *Telecommunication Systems*, pp. 1–14, 2011.

[14] J. Nocedal and S. J. Wright, *Numerical Optimization*. Operations Research, Springer, 1999.

[15] JSVM, *JSVM Software Manual*, ver. jsvm 9.19 ed., March 2011.

[16] ITU-R Rec. BT.500-11, *Methodology for the subjective assessment of the quality of television pictures*, 2002.

[17] ITU-R Rec. P.910, *Subjective video quality assessment methods for multimedia applications*, 1999.