QoE Analysis of Media Streaming in Wireless Data Networks

Yuedong Xu[1], Eitan Altman[2], Rachid El-Azouzi[1], Salah Eddine Elayoubi[3], and Majed Haddad[1]

[1] University of Avignon, 339 Chemin des Meinajaries, Avignon, France
[2] INRIA Sophia Antipolis, 2004 Route des Lucioles, France
[3] Orange Labs, 38-40 rue du General Leclec, Issy-Les-Moulineaux, France
yuedong.xu@gmail.com, eitan.altman@inria.fr,
Rachid.Elazouzi@univ-avignon.fr, salaheddine.elayoubi@orange.com

Abstract. The purpose of this paper is to model quality of experience (QoE) of media streaming service in a shared fast-fading channel. In this context, the arrival and the service processes of the playout buffer do not have the same job size. We present an analytical framework based on Takács Ballot theorem to compute the probability of buffer starvation and the distribution of playback intervals. We model the arrival processes of Proportional Fair and Round Robin schedulers, and feed them into this framework to study the impact of prefetching on the starvation behavior. Our simulations match the developed model very well if the base station knows the playback rate and the channel gain. Furthermore, we make an important observation that QoE metrics predicted by users are very sensitive to the measurement error of arrival process.

Keywords: Media streaming, Quality of Experience, Starvation Probability, Prefetching Delay, Ballot Theorem.

1 Introduction

With the widespread use of smart phones, Http streaming service is become more and more popular on cellular networks. In contrast to the rapid growth of users and traffic load, the bandwidth provision usually lags behind. Under this context, media servers and network operators face a crucial challenge on how to avoid the degradation of user perceived media quality, namely quality of experience (QoE). The most undesirable case is the interruption of playback when users watch videos or listen to music. The primary goal is to avoid the starvation of playout buffer or to reduce the frequency of starvation. One feasible way is to introduce a start-up (also called prefetching) delay before playing the stream, and a rebuffering delay after each starvation event. After a number of media frames accumulate in the buffer, the media player starts to work. However, a large delay may impair the user perceived streaming quality, which demands an appropriate setting of content prefetching.

In this work, we evaluate the impact of prefetching thresholds on the starvation probability and the playback interval in a cellular downlink system. We aim

R. Bestak et al. (Eds.): NETWORKING 2012, Part II, LNCS 7290, pp. 343–354, 2012.

to answer the following questions: 1) how can we model the starvation behavior in a playout buffer whose arrival and service processes work at different levels (e.g. arrival at the bit level and service at the video frame level); 2) how efficient can a user predict its starvation behavior based on local measurement of arrival process. In our setting, mobile users experience Rayleigh fast-fading. The base station adopts proportional fair (PF) or round-robin (RR) algorithms to schedule a user for transmission in each slot. To carry out the analysis, we first model the playout buffer as a GI/D/1 queue whose arrival process is determined by the wireless scheduler. Then, a probabilistic framework is proposed to compute the starvation probability and the playback interval based on Takács Ballot theorem [1]. In comparison with recent works [2–5], we consider a network with shared resource, instead of a single streaming session. The arrival process and the departure process do not work at the same granularity of jobs. The Ballot theorem approach provides a simple way to compute the starvation behaviors.

Our study leads to the following interesting observation. The QoE metrics predicted by a user are very sensitive to the estimation error of arrival process. Especially, the prediction is not meaningful in the heavy traffic regime (i.e. arrival rate approaches service rate). This phenomenon implies that autonomous QoE optimization of mobile users may not perform well in a wireless downlink.

The rest of this work is organized as follows. In section 2, we characterize the QoE metrics in a Wireless fading environment. Section 3 presents a probabilistic framework for QoE analysis. In section 4 we present numerical studies. Section 5 concludes this paper.

2 Media Streaming Service and Quality of Experience

In this section, we present a GI/D/1 queue model for persistent media streaming service, and define a set of metrics for the user perceived streaming quality.

The wireless downlink is shared by multiple elastic flows that can be http media streams, ftp downloading, or other real-time services with flow control. The system architecture with N streaming users is shown in Fig.1. Mobile users request streaming service from media servers, and the media streams are transmitted by the base station over a fast fading channel. The streaming system in Fig.1 is composed of two types of buffers, the *wireless buffer* at the base station as well as the *playout buffer* at the user side. We assume that the media streams always have backlogged packets in the wireless buffer. This assumption holds due to two reasons. First, the online movies are usually very large, ranging from 100MB to 1GB. Second, most of Internet streaming servers use HTTP protocol (over TCP) to deliver streaming packets. TCP congestion control mechanism exploits the available bandwidth by pumping as more packets as possible to the wireless buffer. Therefore, we can simply regard the wireless buffers to be always saturated. The wireless buffer and the playout buffer work in a tandem way, which means that the departure process of a user in the former is exactly the arrival process in the latter. However, they work at different time scales and at different layers. The wireless buffer works at the bit level due to the bit loading in the lower layers, and the playout buffer functions at the video frame level.

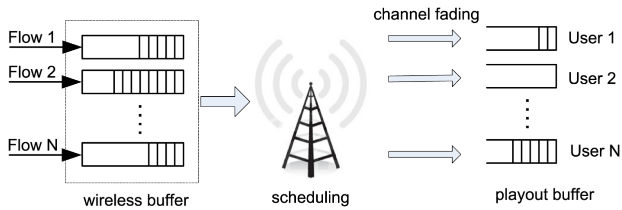

Fig. 1. Media streaming system in wireless cellular networks

The playout buffer at a mobile user stores packets transmitted over the wireless channel. The streaming packets are reassembled into groups of pictures namely "frames" according to the codec of source coding. The size of a frame depends on streaming quality, e.g. a high-definition (HD) video requiring more bits to render each frame. When the media player starts, those frames are served with a deterministic rate. Here, the term "frame rate" refers to the number of pictures that are displayed per second (e.g. 24fps, 25fps, and 30fps). We define **P-slot** D_p as the constant service interval in the playout buffer. For a playback rate 25fps, the duration of a **P-slot** is 40ms. Let $Q(t)$ be the number of frames at time t in the playout buffer and $A(t)$ be the arrival rate of frames. Then, the playout buffer is modeled as a GI/D/1 queue,

$$Q(t+1) = [Q(t) - 1]^+ + A(t+1), \tag{1}$$

where $[x]^+$ denotes the maximum between 0 and x. The general arrival process $A(t)$ is determined by the throughput distribution of a particular flow at the wireless buffer. The maximum buffer size is assumed to be large enough so that the whole media file can be stored without packet drops.

When $Q(t)$ is below a certain threshold, the media player halts. We call this event a *starvation*. Without loss of generality, we set the starvation threshold to be 0 frame at the beginning of each **P-slot**. The starvation property (e.g. probability or frequency) is an important measure of the quality of experience for media streaming services. To avoid starvation, a prefetching stage is introduced by almost all the commercial streaming applications. When a streaming session is initiated, the media player starts to play until the number of accumulated frames reaches a certain value called ***start-up threshold***. One extreme strategy is to download the whole video file before watching. However, the prefetching may cause a long initial waiting time called ***start-up delay***, which is usually unwanted. Therefore, an appropriate choice of the start-up threshold is vital to the users' quality of experience. For a long video stream, starvation events can take place more than once. If the streaming service interrupts during its playing, the media server restarts after a number of packet arrivals. We call this threshold ***rebuffering threshold***, and the induced time ***rebuffering delay***. The streaming provider can either set the same value for the rebuffering and the start-up thresholds, or tune the rebuffering threshold adaptively. Our study in this work is concentrated on long media streams, in which N flows are either persistent or coexisting for a long period. The QoE metrics include the start-up delay, the rebuffering delay, the starvation probability, and the continuous playback interval. Similar metrics has been used in [9] to reveal the interplay between QoE and user satisfaction.

3 Probabilistic Framework for Analyzing QoE

3.1 Preliminaries: Transmissions over a Rayleigh Fading Channel

We consider the media streaming service over a single-cell wireless downlink using HSDPA or EV-DO[1]. The base station is transmitting to a set of mobile users denoted by $\mathbb{N} = \{1, 2, \cdots, N\}$. The wireless channel is shared by all the mobiles where time slot is the allocable resource. We denote by D_s the duration of a scheduling slot (**S-slot**). In general, D_s is smaller than D_p, (e.g. $D_s = 2ms$ [8]). Define an integer $K := D_p/D_s$ (if D_p/D_s is not an integer, our method still applies with slight modification). The wireless channel is assumed to undergo a Rayleigh fast fading due to the mobile environment. This assumption enables us to treat the SNRs of a user to be independent and identically distributed (i.i.d.) in different time slots (see e.g. [6]).

Denote by Γ the symbol of SNR and by γ_i the value of user i's SNR. Then, γ_i is a random variable with the probability density function given below

$$g_\Gamma(\gamma_i) = \frac{1}{\bar{\gamma_i}} \exp(-\frac{\gamma_i}{\bar{\gamma_i}}), \tag{2}$$

where $\bar{\gamma_i}$ is the average SNR of user i. The BS decides not only the user for transmission, but also the modulation and coding scheme (MCS). To help the BS determine the appropriate MCS, each user measures its channel quality and feedbacks channel quality information (CQI) to the BS. We adopt the square M-ary quadrature amplitude modulation (M-QAM) with $M = 2^{2r}, r = 0, 1, 2, \cdots, \hat{r}$, where $\hat{r}$ is the highest modulation order (or equivalently the maximum information bits in each signal). The case $r = 0$ means that the transmission fails. Given the fact that the loaded bits take all values between 0 and $\hat{r}$, the SNR is partitioned into $\hat{r} + 1$ disjoint regions. Denote by $\mathbb{R}_r$ the SNR region of the r^{th} rate. Denote by b_r the lower boundary of $\mathbb{R}_r$ that has $0 = b_0 < b_1 < \cdots < b_{\hat{r}+1} := \infty$. The boundary b_r is obtained by $b_r = -\frac{2}{3}(\ln(5\text{BER}))(2^{2r} - 1)$ [10].Thus, $\mathbb{R}_r$ is written as $[b_r, b_{r+1})$. In this paper, the BER requirement is taken as 10^{-5}. Let B be the spectrum frequency in Hz. If user i is scheduled in a **S-slot** with modulation $M = 2^{2r_i(t)}$ at time t, the number of transmitted bits is $2Br_i(t)D_s$.

In the past decade, a large body of scheduling algorithms have been proposed for various design purposes. Here, we look into two representative schedulers, the SNR-based proportional fairness (PF in abbreviation)[6, 7] and the round robin (RR) algorithms.

[1] Our analysis can be easily extended to 3G long term evolution (LTE) networks if time slot is substituted by OFDMA resource block.

PF Scheduling: PF scheduler selects the user i^* with the highest relative SNR for transmission, that is,

$$i^* = \arg\max_{i \in \mathcal{N}} \frac{\gamma_i}{\bar{\gamma}_i}, \quad \forall i \in \mathcal{N}. \tag{3}$$

We consider a heterogeneous scenario in which different users have different average SNR values due to different path losses. When PF scheduling is adopted, the user with the highest relative SNR is selected for transmission. Then, we obtain the following probability

$$\mathbb{P}\{\text{user } i \text{ scheduled } |\gamma_i = \gamma\} = \mathbb{P}\{\frac{\gamma}{\bar{\gamma}_i} > \frac{\gamma_j}{\bar{\gamma}_j}, \ \forall j \in \mathcal{N} \setminus i\} = \big(1 - \exp(-\frac{\gamma}{\bar{\gamma}_i})\big)^{N-1}.$$

We denote by γ_i^* the SNR of user i if it is scheduled. Its distribution is calculated by:

$$\mathbb{P}\{\gamma_i^* < \Gamma\} = \int_0^{\Gamma} \frac{1}{\bar{\gamma}_i} \exp(-\frac{\gamma}{\bar{\gamma}_i})\big(1 - \exp(-\frac{\gamma}{\bar{\gamma}_i})\big)^{N-1} d\gamma = \frac{1}{N}\big(1 - \exp(-\frac{\Gamma}{\bar{\gamma}_i})\big)^{N}.$$

Denote by $\beta_{i,r}^*$ the probability that user i is scheduled with rate r, $(r \geq 1)$. Then, $\beta_{i,r}^*$ is obtained by

$$\beta_{i,r}^* = P\{b_r \leq \gamma_i^* < b_{r+1}\} = \frac{1}{N}(1 - \exp(-\frac{b_{r+1}}{\bar{\gamma}}))^N - \frac{1}{N}(1 - \exp(-\frac{b_r}{\bar{\gamma}}))^N.$$

Let $\beta_{i,0}^*$ be the probability that user i is not scheduled, or the SNR of user i is below b_1. Then, the equality $\beta_{i,0}^* = 1 - \sum_{r=1}^{\hat{r}} \beta_{i,r}^*$ holds. Given the probability of using r-QAM modulation, we can derive the p.g.f. of U by

$$\Phi_U^i(z) = \beta_{i,0}^* + \sum_{r=1}^{\hat{r}} \beta_{i,r}^* \cdot z^r. \tag{4}$$

The p.g.f. in (4) yields the expectation and the variance

$$\mu = \sum_{r=1}^{\hat{r}} r\beta_{i,r}^* \ \text{ and } \ \sigma^2 = \sum_{r=0}^{\hat{r}} \beta_{i,r}^*(r - \mu)^2. \tag{5}$$

The p.g.f. $\Phi_W(z)$ is obtained by convoluting $\Phi_U(z)$ for K times. Applying our derivations in section 3, we can obtain the QoE metrics of a streaming user with PF scheduling.

Round Robin Scheduling: We consider a simplified scenario where a P-slot is evenly divided into N S-slots. Denote by β_r the probability of SNR within $\mathbb{R}_r$. Then,

$$\beta_r = \exp\big(-\frac{b_r}{\bar{\gamma}}\big) - \exp\big(-\frac{b_{r+1}}{\bar{\gamma}}\big), \ \ r = 0, 1, \cdots, \bar{r}. \tag{6}$$

Therefore, the probability generating function is given by

$$\Phi_U(z) = (1-\frac{1}{N}+\frac{\beta_0}{N}) + \sum_{r=1}^{\hat{r}} \frac{\beta_r}{N} \cdot z^r. \quad (7)$$

The expectation and the variance of U are obtained by

$$\mu = \sum_{r=1}^{\hat{r}} \frac{r\beta_r}{N} \text{ and } \sigma^2 = (1-\frac{1}{N}+\frac{\beta_0}{N})\mu^2 + \sum_{r=1}^{\hat{r}} \frac{\beta_r}{N}(r-\mu)^2. \quad (8)$$

3.2 Interaction between Scheduler and Playout Buffer

We define a set of random variables to measure the user throughput in bits/Hz per S-slot. Let the i.i.d. random variable $U(s)$ denote the modulation scheme at the S-slot s. The number of bits transmitted in S-slot s is $U(s) \times 2BD_s$. We denote by W the sum of $U(s)$ in K S-slots, i.e. $W := \sum_{s=1}^{K} U(s)$. The volume of the transmitted bits in a P-slot is $W(s) \times 2BKD_s$. Let V_k record the transmitted bits in k P-slots, i.e., $V_k := \sum_{t=1}^{k} W(t)$. Recall that $\Phi_U(z)$ is the p.g.f of U that depends on the used scheduler, as detailed previously . Due to the i.i.d. property, the probability generating function of W, $\Phi_W(z)$, is

$$\Phi_W(z) = [\Phi_U(z)]^K, \quad (9)$$

and that of V_k, $\Phi_{V_k}(z)$, is

$$\Phi_{V_k}(z) = [\Phi_W(z)]^k. \quad (10)$$

Denote by μ and σ the mean and the standard deviation of U, respectively. Then, W has a mean $K\mu$ and a standard variation $\sqrt{K}\sigma$. Similarly, the mean and the standard deviation of V_k are $kK\mu$ and $\sqrt{kK}\sigma$. We further define new variables as follows: $\phi_U(j) = \mathbb{P}\{U = j\}$, $\phi_V(k,j) = \mathbb{P}\{V_k = j\}$, $\phi_W(j) = \mathbb{P}\{W(t) = j\}$, $F_U(x) = \mathbb{P}\{U \leq x\}$, $F_W(x) = \mathbb{P}\{W \leq x\}$, and $F_V(k,x) = \mathbb{P}\{V_k \leq x\}$. In what follows, we define a constant $c := 2BD_s/d$ where d is the video frame size in bits. Define a new variable $\rho := cK\mu$ as the traffic load.

The p.g.f. of V_k is obtained from k convolutions of W. The r.v. W takes integer values from 0 to $K\hat{r}$. Then, V_k takes values from 0 to $kK\hat{r}$. As k is large, it needs a heavy computation to obtain $\phi_V(k,j)$. When k is large enough, we can approximate V_k by a diffusion process according to Donsker's theorem [11],

$$V_k \approx kK\mu + \sqrt{K}\sigma \cdot \mathcal{B}(k) + o(\sqrt{k}), \quad (11)$$

where $\mathcal{B}(k)$ is a standard Brownian motion and the symbol $\approx$ denotes "has approximately the same distribution as".

We define X for the frame arrival to the playout buffer in each P-slot. Denote $Y_k := \sum_{t=1}^{k} X(t)$ to be the total frame arrivals in k P-slots. In a P-slot, W takes integer values in the set $[0, K\hat{r}]$. Define a constant $R := \lceil cK\hat{r} \rceil$. The r.v. $X(t)$

takes values from 0 to R. The exact distributions of X and Y_k are intractable because the number of arrived media frames can be a fractional value in reality. Even if the distribution of X is known, it is still complicated to obtain the distribution of Y_k for large k. To carry out our analysis, we make two important assumptions:

- The r.v., $X(t)$, is i.i.d. at all time t;
- The stochastic process $\{Y_k\}$ has

$$Y_k \approx \tilde{Y}_k = ckK\mu + c\sigma\sqrt{K} \cdot \mathcal{B}(k) + o(\sqrt{k}), \tag{12}$$

where $\{\tilde{Y}_k : k > 0\}$ is the approximated diffusion process.

By simplifying the arrival process to the playout buffer, we can approximate the probability mass function by

$$\mathbb{P}\{\tilde{Y}_k = y\} \approx \Phi_{\tilde{Y}}(k, y) = \frac{1}{\sqrt{2\pi kK}c\sigma} \exp\Big(-\frac{(y - ckK\mu)^2}{2kK(c\sigma)^2}\Big). \tag{13}$$

3.3 Computing Playback Interval: A Ballot Theorem Approach

In this paper, we consider the scenario of infinite size of media files. Because a "tagged" user is examined, the user index is dropped in all r.v.s. Given the rebuffering threshold, we are interested in the probability distribution of continuous playback interval T_I (in P-slots). Let P-slot 1 be the starting point of time axis when media player resumes. We denote by $P_{ns}(l, x_r)$ the probability of no starvation in the duration $[1, l]$,

$$P_{ns}(l, x_r) = \mathbb{P}\{\sum_{s=1}^{t} X(s) - (t - x_r) > 0, \;\; \forall \;\; x_r \le t < l\}. \tag{14}$$

An intuitive explanation of eq.(14) is the following. The prefetched x_r frames will be served in x_r P-slots. If there is no new arrival, i.e. $\sum_{s=1}^{x_r} X(s) = 0$, the media player meets with an empty buffer at the beginning of slot x_r+1, thus causing a starvation event. Therefore, $\sum_{s=1}^{x_r} X(s)$ must be greater than 0 if there is no starvation. Similarly, we can replace x_r by an arbitrary $t, (t > x_r)$.

The direct computation of P_{ns} involves a number of complicated iterative equations. We hereby adopt the famous Ballot theorem for a combinatorial analysis. The version of most referenced Ballot theorem is developed by Takács [1].

Theorem 1. *(Takács Ballot Theorem) If $X(1), X(2), \cdots, X(L)$ are cyclically interchangeable random variables taking on nonnegative integer values summing to k, then*

$$\mathbb{P}\Big\{\sum_{s=1}^{t} X(s) < t, \forall\, t \in [1, l]\Big\} = \frac{[l - k]^+}{l}, \;\; (t > x_r). \tag{15}$$

The Takács Ballot Theorem presents a probability that the number of departures is larger than that of arrivals in all l slots. Because $X(s)$ is i.i.d. at different time s, it is cyclically interchangeable. Suppose that the starvation event happens after the media player has worked for t slots, $t > x_r$. The total number of consecutive departure slots is $t-1$, and the total number of arrivals is $t-1-x_r$. We create a backward time axis where the starvation event happens at slot 1. The number of departures is always greater than that of arrivals. Otherwise, the starvation event has already taken place. Hence, according to Takács Ballot theorem, the probability of the departure process leading the arrival process is

$$\mathbb{P}\{\sum_{s=1}^{t} X(s) < t,\ \forall t \in [1,l]\} = \frac{x_r}{t-1}. \tag{16}$$

Therefore, the probability that the first starvation takes place at slot t is

$$\begin{aligned} P_s(t) &= \frac{x_r}{t-1} \cdot \mathbb{P}\{\sum_{s=1}^{t-1} X(s) = t-1-x_r\}, \qquad (t \geq x_r + 1) \\ &\approx \frac{x_r}{t-1} \cdot \mathbb{P}\{\tilde{Y}_{t-1} = t-1-x_r\} = \frac{x_r}{t-1} \cdot \Phi_{\tilde{Y}}(t-1, t-1-x_r) \qquad (17) \\ &= \frac{x_r}{t-1} \cdot \frac{1}{\sqrt{2\pi(t-1)K}c\sigma} \exp\big(-\frac{(t-1-x_r - c(t-1)K\mu)^2}{2(t-1)K(c\sigma)^2}\big) \\ &= \frac{x_r}{\sqrt{2\pi(t-1)^3K}c\sigma} \exp\big(-\frac{((t-1)(1-cK\mu) - x_r)^2}{2(t-1)K(c\sigma)^2}\big). \qquad (18) \end{aligned}$$

The approximation in eq.(17) means that the actual frame arrival process is replaced by the corresponding Brownian motion. Eq.(18) yields the c.d.f. of continuous playback interval

$$\mathbb{P}\{T_I \leq l\} = \sum_{t=x_r+1}^{l} P_s(t) = \sum_{t=x_r+1}^{l} \frac{x_r}{\sqrt{2\pi(t-1)^3K}c\sigma} \exp\big(-\frac{((t-1)(1-cK\mu) - x_r)^2}{2(t-1)K(c\sigma)^2}\big). \tag{19}$$

The starvation probability of an infinite file size is equivalent to $\mathbb{P}\{T_I \leq \infty\}$. According to Lemma 1 in [5], eq.(19) with $l = \infty$ is further approximated by

$$P_s^{tot}(\infty) := \begin{cases} 1 & \text{if } \rho < 1; \\ \exp\big(\frac{2x_r(1-cK\mu)}{K(c\sigma)^2}\big) & \text{otherwise }. \end{cases} \tag{20}$$

The probability of starvation is determined by the start-up threshold x_r and the traffic load ρ. If ρ is less than one, the starvation is inevitable. For $\rho > 1$ and x_r is large enough, the starvation probability is negligible. The probability of no starvation P_{ns} is the complement of P_s^{tot}. Next, we compute the average playback interval by

$$E[T_I] = \sum_{t=x_r+1}^{\infty} (t-1) \cdot P_s(t) \approx \frac{x_r}{(1-cK\mu)}, \quad \text{if } \ cK\mu < 1. \tag{21}$$

The analysis of above approximation can be found in the technical report [12]. If $E[T_I]$ is infinitely large, the starvation event takes place only in the beginning.

3.4 Computing Start-Up/Rebuffering Delays

The start-up delay T_p denotes the number of minimum required P-slots to have $Y_{T_p} \geq x_p$. Its cumulative probability distribution is expressed as

$$\mathbb{P}\{T_p < t\} = \mathbb{P}\{Y_t \geq x_p\} \approx \int_0^{x_p} \frac{1}{\sqrt{2\pi t K} c\sigma} \exp\big(\frac{(x - cK\mu t)^2}{2tK(c\sigma)^2}\big) dx. \tag{22}$$

The expected start-up delay, $E[T_p]$, is easily obtained as the busy period of GI/G/1 queue,

$$E[T_p] = \frac{x_p}{E[X]} = \frac{x_p}{cK\mu}. \tag{23}$$

Similarly, we can compute the distribution of rebuffering time T_r.

3.5 Prediction of the Starvation Behavior by the User

The computation of eq.(19),(20) and (21) relies on μ and σ that are only accessible to the BS. An important question is how a user predicts the starvation behavior based on local measurement of arrival process. Define a time window T_W in P-slots. A user records the complete frame arrival at each P-slot of the time window T_W. We construct a r.v. $\bar{X}$ and a stochastic process $\{\bar{Y}_k : k > 0\}$. The r.v. $\bar{X}$ characterizes the frame arrival to the playout measured by the user, and the r.v. $\bar{Y}_k := \sum_{t=1}^{k} \bar{X}(t)$. Let $\bar{\mu}$ and $\bar{\sigma}^2$ be the mean and the variance of $\bar{X}$.Therefore, $\Phi_{\tilde{Y}}(t-1, t-1-x_r)$ in eq.(18) is substituted by $\Phi_{\bar{Y}}(t-1, t-1-x_r)$. Given the i.i.d. assumption of $\tilde{X}$, $\tilde{Y}_k$ can be approximated by the Brownian motion $\bar{Y}_k \approx k\bar{\mu} + \sqrt{k}\bar{\sigma} \cdot \mathcal{B}(k)$ for large k. The media player can use eq.(20) and (21) to predict the starvation probability and the expected playback interval respectively. We only need to substitute $cK\mu$ by $\bar{\mu}$, and substitute $\sqrt{K}c\sigma$ by $\bar{\sigma}$.

4 Simulation

The main goal of this section is to validate the accuracy of QoE model. Due to the page limit, our simulation is confined to the network where the users have same average SNR and the BS adopts PF scheduler. (More simulation studies can be found in the technical report [12].) The frequency width of wireless channel is set to 2MHz. The base station adopts the proportional fair algorithm to schedule users every 2ms. 4-QAM, 16-QAM and 64-QAM modulation schemes are adopted. We assume that the durations of all video streams are large enough. Each stream has a playback rate of 300Kbps and a frame rate of 25fps. For each scenario, we conduct three hundreds of simulation runs.

The first set of experiments demonstrate the c.d.f. of playback interval for different values of user population and average SNR. For each set of parameters, we compare the c.d.f. curves obtained from the model at the BS side, the model at the receiver side, and the simulation. The model at the BS side is computed by eq.(19). The model at the user side is also obtained from eq.(19), but with substitutions of $cK\mu$ by $\bar{\mu}$ and $\sqrt{K}c\sigma$ by $\bar{\sigma}$. The time window of measuring the arrival process is set to 200 P-slots (equivalent to 8s). We plot the c.d.f. of playback intervals with $x_r = 100$ and $x_r = 200$ in Fig. 2. The wireless channel is shared by 20 homogeneous users, and the average SNR of a user is 10 dB. The average number of frame arrival is 0.73 per P-slot so that the starvation event happens for sure. Fig. 2 exhibits a very good matching between the playback interval model computed by the BS and the simulation. However, the model based on user measurements is far away from accuracy. Our simulations provide an important insight that the QoE prediction at the user side might not work very well.

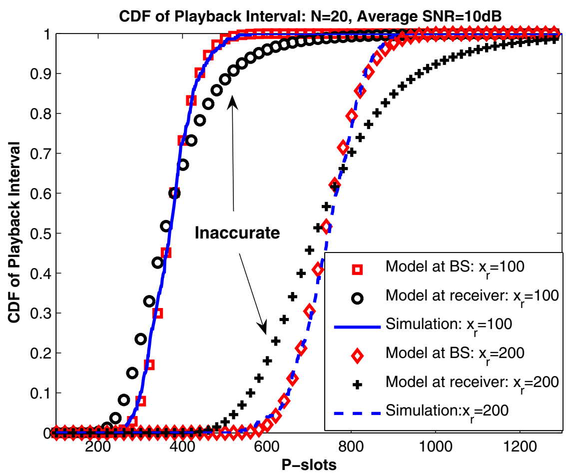

Fig. 2. CDF of playback interval: $N = 20$ and $\bar{\gamma} = 10dB$

The second set of experiments evaluate the average playback interval when the user number and the average SNR change. In Fig.3, the number of users increases from 16 to 40, the average playback interval reduces. We have two observations here. One is that the slop of decrease is much more prominent in the range $16 \leq N \leq 24$ than that in the range $26 \leq N \leq 40$. Here, the expected frame arrival rate is 0.619 frame per P-slot at $N = 24$ and 0.897 frame per P-slot at $N = 16$. The other observation is that the average playback interval predicted by a user is not accurate at $N = 16$. The predicted value is very sensitive to the measurement error when the frame arrival rate is close to 1.

The third set of experiments explore the behavior with respect to the rebuffering threshold. We only compare the model at the BS side and the simulation. Fig.4 illustrates the probability of future starvation after the current rebuffering. Since the starvation happens certainly with $\rho \leq 1$, we only consider the case where ρ is slightly greater than 1. Suppose that a cell has 19 homogeneous users

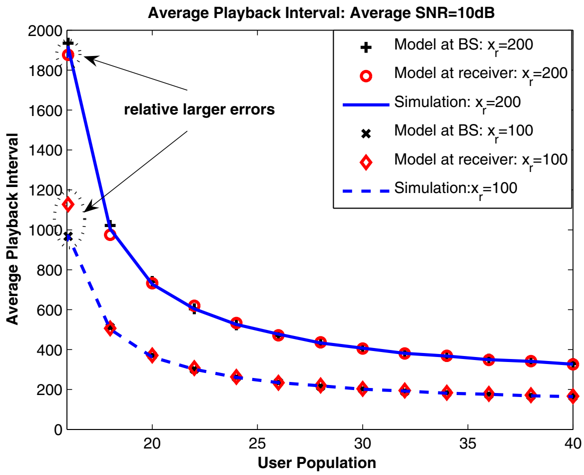

Fig. 3. Average playback interval with different user numbers: $\bar{\gamma} = 10dB$

with average SNR of 12dB. The frame arrival rate computed from the model is 1.0492 frames per P-slot. We compare the probability of starvation computed from (20) and that of simulation. Our experiments validate the accuracy of the asymptotic starvation probability and startup delay. The figure 4 also shows that the average startup delay increases linearly with the rebuffering threshold, with a perfect match between simulation and model.

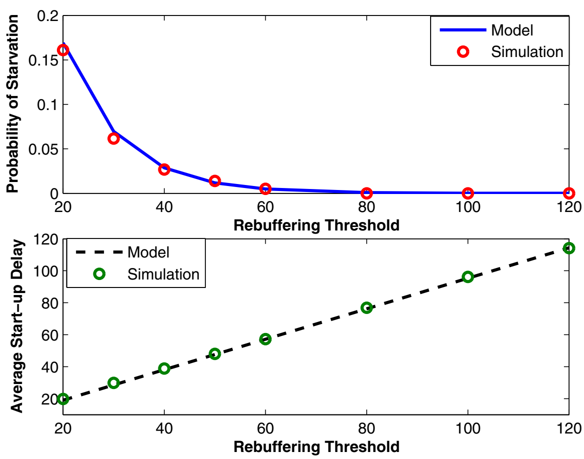

Fig. 4. Probability of Starvation and Average Start-up Delay

5 Conclusion and Future Work

We analyzed the starvation behavior for users experiencing Rayleigh fading in a wireless channel. The arrival and the service processes of a playout buffer do not function at the same job size, nor at the same time scale. We use Takács Ballot Theorem to compute the metrics related to buffer starvation in a very simple way. When the knowledge of arrival process is complete (e.g. at the BS side), our model matches the simulation result very well. When the arrival process is measured by the playout buffer, the prediction of starvation behavior is inaccurate. Hence, our study suggests that the BS, instead of autonomous users, should be responsible for the QoE optimization. In the future, we aim at assessing the impact of flow dynamics on finite video streams and at developing QoE optimization schemes that take into account the dynamic flow arrivals and departures.

Acknowledgements. This work was supported by a contract with Orange Lab, Issy Les Moulineaux. The authors would like to thank the anonymous reviewers for their valuable comments to improve the quality of the paper.

References

1. Takács, L.: Ballot Problems. Probability Theory and Related Fields 1(2), 154–158 (1962)
2. Luan, H., Cai, L.X., Shen, X.: Impact of network dynamics on users' video quality: analytical framework and QoS provision. IEEE Trans. on Multimedia 12(1), 64–78 (2010)
3. Liang, G., Liang, B.: Effect of delay and buffering on jitter-free streaming over random VBR channels. IEEE Trans. Multimedia 10(6), 1128–1141 (2008)
4. ParandehGheibi, A., Medard, M., Ozdaglar, A., Shakkottai, S.: Avoiding Interruptions a QoE Reliability Function for Streaming Media Applications. IEEE Journal on Selected Area in Communications 29(5), 1064–1074 (2011)
5. Xu, Y., Altman, E., Elzouzi, R., Haddad, M., Elayoubi, S., Jimenez, T.: Probabilistic Analysis of Buffer Starvation in Markovian Queues. In: IEEE Infocom (2012)
6. Combes, R., Elayoubi, S., Altman, Z.: Cross-layer analysis of scheduling gains: Application to LMMSE receivers in frequency-selective Rayleigh-fading channels. In: Proc. of ICST WiOpt 2011, pp. 133–139 (2011)
7. Choi, J.G., Bahk, S.: Cell-throughput analysis of the proportional fair scheduler in the single-cell environment. IEEE Trans. Veh. Tech. 56, 766–778 (2007)
8. Bonald, T., Muscariello, L.: Opportunistic Scheduling of Voice and Data Traffic in Wireless Networks. In: EuroFGI Workshop on IP QoS and Traffic Control (2007)
9. Dobrian, F., Awan, A., Stoica, I., Sekar, V., Ganjam, A., Joseph, D., Zhan, J., Zhang, H.: Understanding the Impact of Video Quality on User Engagement. In: Proc. of ACM Sigcomm 2011, Toronto (2011)
10. Chang, Y.J., Chien, F.T., Kuo, C.C.: Cross-layer QoS Analysis of Opportunistic OFDM-TDMA and OFDMA Networks. IEEE JSAC 25, 657–666 (2007)
11. Donsker, M.D.: Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. Annals of Mathematical Statistics 23 (1952)
12. Xu, Y.: QoE Analysis of Media Streaming in Wireless Data Networks (2011), http://sites.google.com/site/yuedongxu/