

Avoiding Interruptions - QoE Trade-offs in Block-coded Streaming Media Applications

Ali ParandehGheibi[†], Muriel Médard[‡], Srinivas Shakkottai[‡], Asuman Ozdaglar[†]
 parandeh@mit.edu, medard@mit.edu, sshakkot@tamu.edu, asuman@mit.edu

Abstract—We take an analytical approach to study Quality of user Experience (QoE) for media streaming applications. We use the fact that random linear network coding applied to blocks of video frames can significantly simplify the packet requests at the network layer and avoid duplicate packet reception. We model the receiver's buffer as a queue with Poisson arrivals and deterministic departures. We consider the probability of interruption in video playback (buffer underflow) as well as the number of initially buffered packets (initial waiting time) as the QoE metrics. We explicitly characterize the optimal trade-off between these metrics by providing upper and lower bounds on the minimum initial buffering required to achieve certain level of interruption probability for different regimes of the system parameters. Our bounds are asymptotically tight as the file size goes to infinity. Further, we show that for arrival rates slightly larger than the play rate, the minimum initial buffering remains bounded as the file size grows. This is not the case when the arrival rate and the play rate match.

I. INTRODUCTION

Peer-to-peer networks (P2P) are a fast-growing means of video delivery. It has been estimated that between 35-90% of Internet bandwidth is consumed by P2P applications [1], [2]. Today, P2P file-sharing networks are seeing a drop in popularity [3], but the original file sharing ideas are being used for video streaming in networks such as PPLive [4] and QQLive [5]. As smart phones become the medium of choice for Internet media access, P2P video distribution over the wireless medium is likely to gain significance.

P2P video streaming is generally accomplished by dividing the video file into *blocks*, which are then further divided into packets for transmission. After each block is received, it can be played out by the receiver. In order to ensure smooth sequential playback, a fresh block must be received before the current block has been played. If such a fresh block is not available the playback freezes, causing a negative user experience. Blocks may be buffered in advance of playing out in order to provide a level of protection against a playback freeze, with more initial buffering providing a lower likelihood of playback interruption. Hence, there is a trade-off between the initial waiting time and playback interruptions.

In this paper, our main objective is to characterize the amount of buffering needed for a target probability of playback interruption over the duration of the video. We consider a model in which network coding is used across the packets of

each block. A wireless user can obtain coded packets from multiple sources (other users and servers). However, since the wireless channel is unreliable, packets cannot be obtained deterministically. Thus, our question is *how much should we buffer prior to playback in order to account for wireless channel variations?*

We first show how to model the receiver's buffer as an M/D/1 queue. We then provide upper and lower bounds on the minimum initial buffering required so that the playback interruption probability is below a desired level. The optimal trade-off between the initial buffering and the interruption probability depends on the file size as well as the arrival rate of the packets as compared to the playback rate. We show that our bounds are asymptotically tight as the file size goes to infinity. Moreover, if the arrival rate is slightly larger than the play rate, the minimum initial buffering for a given interruption probability remains bounded as the file size grows. However, when the arrival rate and the play rate match, the minimum initial buffer size grows as the square-root of the file size.

There is significant work in the space of P2P streaming. Close to our work, [6], [7], [8], [9] develop analytical models on the trade-off between the steady state probability of missing a block, and buffer size under different block selection policies for live streaming in a full mesh P2P network with deterministic channels. A further modification is to use random linear network coding techniques [10] to make block selection simpler [11], [12], [13], [14] in the wired and wireless context. In contrast, we focus on a very different scenario of streaming of pre-prepared content over unreliable wireless channels using network coding. Further, our analysis is on transient effects—we are interested in the first time that video playback is interrupted as a function of the initial amount of buffering.

II. SYSTEM OVERVIEW

We consider a media streaming system as follows. Media files are usually divided into blocks consisting of multiple frames. The video coding is such that all the frames in the block need to be available before any frames can be played. Blocks are requested in sequence by the playback application from the user-end. The server (or other peers) packetize the requested block and transmit them to the user as in Figure 1. Obtaining the packets of a block from various peers in a P2P system requires the receiver to keep track of missing packets in a block, and request them from different peers. However, since packet transmission is unreliable in a wireless context, requesting each packet from only one peer might

[†] Department of Electrical Engineering and Computer Science, MIT.

[‡] ECE Department, Texas A&M University.

This material is based upon work under subcontract # 18870740-37362-C issued by Stanford University and supported by DARPA.

cause unreasonable delays, while requesting a particular packet from multiple peers can result in inefficient resource usage.

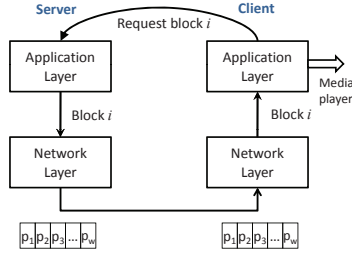


Fig. 1. The media player (application layer) requires complete blocks. At the network layer each block is divided into packets and delivered.

Random linear codes can be used to alleviate such inefficiencies. Here, instead of requesting a particular packet from block i , the receiver simply requests a *degree of freedom* of block i . The server in turn responds with a random linear combination of all packets that it has in block i . The coefficients of each combination are chosen uniformly at random from a Galois field of size q . The coded packets delivered to the receiver can be thought of as linear equations, where the unknowns are the original packets in block i . Block i can be fully recovered by solving a system of linear equations if it is full rank. It can be shown that if the field size q is large enough, the received linear equations are linearly independent with very high probability [10]. Therefore, for recovering a block of W packets, it is sufficient to receive W coded packets from different servers. In a P2P system, it is unlikely that a randomly contacted peer would have all packets corresponding to a particular block. However, storing blocks in a random linear coded fashion at all peers ensures that with high probability, the selected peer has a new degree of freedom to offer (see [13] for further discussion). Thus, each received coded packet is likely to be independent of previous ones with probability $1 - \delta(q)$, where $\delta(q) \rightarrow 0$ as $q \rightarrow \infty$.

Note that such random linear coding does not introduce additional decoding delay for each block, since the frames in a block can only be played out when the whole block is received. So there is no difference in delay whether the end-user received W uncoded packets of the block or W independent coded packets that can then be decoded.

III. SYSTEM MODEL AND QOE METRICS

Consider a single user receiving a media file from various peers it is connected to. Each peer could be a wireless access point or another wireless user operating as a server. We assume that the video file consists of T packets that are divided into blocks of W packets. Each server sends random linear combinations of the packets within the current block to the receiver. We assume that the linear combination coefficients are selected from a Galois field of size q . We assume the block size W is small compared to the total length of the file, but large enough to ignore the boundary effects of moving from one block to the next. Time is continuous, and the arrival process of packets from each peer is a Poisson process

independent of other arrival processes. Since no redundant packet is delivered from different peers, we can combine the arrival processes into one Poisson process of rate R_s . We assume that each received coded packet is linearly independent from the previous ones with probability $1 - \delta(q)$. Hence, the effective arrival process of *useful* packets is Poisson with rate $R = R_s(1 - \delta(q))$. Note that R approaches R_s for large enough field size. We normalize the playback rate to one, i.e., it takes one unit of time to play a single packet. Thus, our simplified model is just a single-server-single-receiver system. We also assume that the parameter R is known at the receiver, which first buffers D packets from the beginning of the file, and then starts the playback.

The presence of some packets in the buffer does not guarantee that there will be no interruption since we require W packets corresponding to a block before it can be decoded and played out. However, if there are at least W packets in the buffer, there is at least one playable packet. This is so since either the first W packets in the buffer belong to the same block, or they belong to two different blocks. In the former case, the packets of the block can be decoded, and in the latter case, the first block of the two must be already decoded; otherwise, the next block would not be sent from the server. Therefore, the dynamics of the receiver's buffer size $Q(t)$ can be described as follows

$$Q(t) = D + A(t) - t, \quad (1)$$

where D is the initial buffer size and $A(t)$ is a Poisson process of rate R . We declare an interruption in playback when the buffer size decreases to the threshold W . For simplicity of notation, we assume that an *extra* block is initially buffered (not taken into account in D). Hence, we can declare an interruption in playback when the buffer size reaches zero before reaching the end of the file. More precisely, let

$$\tau_e = \inf\{t : Q(t) \leq 0\}, \quad \tau_f = \inf\{t : Q(t) \geq T - t\}, \quad (2)$$

where τ_f corresponds to time of completing the file download, because we have already played τ_f packets and the buffer contains the remaining $T - \tau_f$ packets to be played. The video streaming is interrupted if and only if $\tau_e < \tau_f$.

We consider the following metrics to quantify Quality of user Experience (QoE). The first metric is the initial waiting time before the playback starts. This is directly captured by the initial buffer size D . Another metric that affects QoE is the probability of interruption during the playback denoted by

$$p(D) = \Pr\{\tau_e < \tau_f\}, \quad (3)$$

where τ_e and τ_f are defined in (2). In our model, the user expects to have an interruption-free experience with probability higher than a desired level $1 - \epsilon$. Note that there is a fundamental trade-off between the interruption probability ϵ and the initial buffer size D . For example, owing to the randomness of the arrival process, in order to have zero probability of interruption, it is necessary to fully download the file, i.e., $D = T$. Nevertheless, we need to buffer only a small fraction of the file if user tolerates a positive probability of interruption. These trade-offs and their relation to system parameters R and T are addressed in the following section.

IV. OPTIMAL QOE TRADE-OFFS

We would like to obtain the smallest initial buffer size so that the interruption probability is below a desired level ϵ , which is denoted by

$$D^*(\epsilon) = \min\{D \geq 0 : p(D) \leq \epsilon\}, \quad (4)$$

where $p(D)$ is the interruption probability defined in (3). Note that in general $p(D)$ and hence $D^*(\epsilon)$ depend on the arrival rate R and the file size T which are assumed to be known and constant. In the following we characterize the optimal trade-off between the initial buffer size and the interruption probability by providing bounds on $D^*(\epsilon)$. An upper bound (achievability) on $D^*(\epsilon)$ is particularly useful, since it provides a sufficient condition for desirable user experience. A lower bound (converse) of $D^*(\epsilon)$ provides a necessary condition on the initial buffer size for a desirable level ϵ of interruption probability. Let us first introduce some useful lemmas.

Lemma 1. Let $X(t) = e^{-rQ(t)}$, where $Q(t)$ is given by (1), and define

$$\gamma(r) = r + R(e^{-r} - 1). \quad (5)$$

Then for every $r \geq 0$ such that $\gamma(r) \geq 0$, $X(t)$ is a sub-martingale with respect to the canonical filtration $\mathcal{F}_t = \sigma(X(s), 0 \leq s \leq t)$, i.e., the smallest σ -field containing the history of the stochastic process X up to time t .

Proof: For every t , $|X(t)| \leq 1$. Hence, $X(t)$ is uniformly integrable. It remains to show that for every $t \geq 0$ and $h > 0$,

$$\mathbf{E}[X(t+h)|\mathcal{F}_t] \geq X(t) \quad \text{a.s.} \quad (6)$$

The left-hand side of (6) can be expressed as

$$\begin{aligned} \mathbf{E}[X(t+h)|\mathcal{F}_t] &= \mathbf{E}\left[e^{-r(Q(t+h)-Q(t))} \middle| \mathcal{F}_t\right] X(t) \\ &= \mathbf{E}\left[e^{-r(A(t+h)-A(t))} \middle| \mathcal{F}_t\right] e^{rh} X(t) \\ &\stackrel{(a)}{=} \mathbf{E}\left[e^{-rA(h)}\right] e^{rh} X(t) \\ &\stackrel{(b)}{=} e^{h(r+R(e^{-r}-1))} X(t) = e^{h\gamma(r)} X(t), \end{aligned}$$

where (a) follows from independent increment property of the Poisson process, and (b) follows from the fact that $A(t)$ is a Poisson random variable. Now, it is immediate to verify (6) for any r with $\gamma(r) \geq 0$. ■

Next, we use Doob's maximal inequality [15] to bound the interruption probability.

Lemma 2. Let $p(D)$ be the interruption probability defined in (3), and $\gamma(r)$ be given by (5). Then, for any $r \geq 0$ with $\gamma(r) \geq 0$

$$p(D) \leq e^{-rD+T\gamma(r)}, \quad \text{for all } D, T, R \geq 0. \quad (7)$$

Proof: By definition of $p(D)$ in (3), we have

$$\begin{aligned} p(D) &= \Pr\{\tau_e < \tau_f\} \\ &\leq \Pr\{\tau_e \leq T\} = \Pr\left\{\inf_{0 \leq t \leq T} Q(t) \leq 0\right\} \\ &= \Pr\left\{\sup_{0 \leq t \leq T} e^{-rQ(t)} \geq 1\right\} \\ &\stackrel{(a)}{\leq} \mathbf{E}[e^{-rQ(T)}] = \mathbf{E}[e^{-r(D+A(T)-T)}] \\ &= e^{-r(D-T)} e^{RT(e^{-r}-1)} = e^{-rD+T\gamma(r)}, \end{aligned}$$

where (a) holds by applying Doob's maximal inequality [15] to the non-negative sub-martingale $X(t) = e^{-rQ(t)}$. Note that $X(t)$ is a sub-martingale for all r with $\gamma(r) \geq 0$ by Lemma 1. ■

Lemma 3. Let $\gamma(r)$ be as defined in (5). Define $\bar{r}(R)$ as the largest root of $\gamma(r)$, i.e.,

$$\bar{r}(R) = \sup\{r : \gamma(r) = 0\}. \quad (8)$$

The following relations hold:

$$\bar{r}(R) = 0, \quad \text{if } 0 \leq R \leq 1, \quad (9)$$

$$\frac{2(R-1)}{R} \leq \bar{r}(R) \leq 2(R-1), \quad \text{if } 1 \leq R \leq 2, \quad (10)$$

$$R-1 \leq \bar{r}(R) \leq R \leq 2(R-1), \quad \text{if } R \geq 2. \quad (11)$$

Proof: We omit the proof for brevity. See [16]. ■

Next, we provide sufficient conditions on the initial buffer size to avoid interruptions with high probability for different regimes of the arrival rate.

Theorem 1. [Achievability] Let $D^*(\epsilon)$ be defined as in (4), and $\bar{r}(R)$ be given by (8). Then

(a) For all $R > 1$,

$$D^*(\epsilon) \leq \frac{1}{\bar{r}(R)} \log\left(\frac{1}{\epsilon}\right). \quad (12)$$

(b) For all $0 \leq R \leq 1 + \left(\frac{1}{2T} \log\left(\frac{1}{\epsilon}\right)\right)^{\frac{1}{2}}$,

$$D^*(\epsilon) \leq \min\left\{\frac{1}{\bar{r}(R)} \log\left(\frac{1}{\epsilon}\right), T(1-R) + \left(2TR \log\left(\frac{1}{\epsilon}\right)\right)^{\frac{1}{2}}\right\}. \quad (13)$$

Proof: First, note that for any upper bound $\bar{p}(D)$ of the interruption probability $p(D)$, any feasible solution of

$$\bar{D}(\epsilon) = \min\{D \geq 0 : \bar{p}(D) \leq \epsilon\} \quad (14)$$

provides an upper bound on $D^*(\epsilon)$. This is so since the optimal solution of the above problem is feasible in the minimization problem (4). If the problem in (14) is infeasible, we use the convention $\bar{D}(\epsilon) = \infty$, which is a trivial bound on $D^*(\epsilon)$. The rest of the proof involves finding the tightest bounds on $p(D)$ and solving (14).

Part (a): By Lemma 2, for $r = \bar{r}(R)$, we can write

$$p(D) \leq \bar{p}_a(D) = e^{-\bar{r}(R)D}, \quad \text{for all } D, T, R \geq 0.$$

Solving $\bar{p}_a(D) = \epsilon$ for D gives the result of part (a). Since $\bar{r}(R) = 0$ for $R \leq 1$ (cf. Lemma 3), this bound is not useful in that range.

Part (b): First, we claim that for all $D \geq T(1-R+\bar{r}(R))$,

$$p(D) \leq \bar{p}_b(D) = e^{-\frac{1}{2}TRz^2},$$

where $z = 1 - \frac{1}{R}(1 - \frac{D}{T})$. We use Lemma 2 with $r = r^* = -\log\left(\frac{1}{R}(1 - \frac{D}{T})\right)$ to prove the claim. Note that $r^* \geq 0$,

because $D \geq T(1 - R)$. In order to verify the second hypothesis of Lemma 2, consider the following

$$\begin{aligned} R(e^{-r^*} - e^{-\bar{r}(R)}) &= \bar{r}(R) + R(e^{-r^*} - 1) - \gamma(\bar{r}(R)) \\ &\stackrel{(a)}{=} \bar{r}(R) - R + (1 - \frac{D}{T}) \\ &\stackrel{(b)}{=} \frac{1}{T} [T(1 - R + \bar{r}(R)) - D] \stackrel{(c)}{\leq} 0, \end{aligned}$$

where (a) and (b) follow from the definition of $\bar{r}(R)$ and r^* , respectively, and (c) holds by the hypothesis of the claim. Thus, $r^* \geq \bar{r}(R)$. Using the facts that $\bar{r}(R)$ is the largest root of $\gamma(r)$, and $\gamma(r) \rightarrow +\infty$ as $r \rightarrow \infty$, we conclude that $\gamma(r^*) \geq 0$. Now, we apply Lemma 2 to get

$$\begin{aligned} p(D) &\leq e^{-r^*D + T\gamma(r^*)} \\ &\stackrel{(a)}{=} e^{TR(\frac{1}{R}(1 - \frac{D}{T})r^* - (1 - e^{-r^*}))} \\ &\stackrel{(b)}{=} e^{TR(-(1-z)\log(1-z) - z)} \\ &\stackrel{(c)}{\leq} e^{-\frac{1}{2}TRz^2}, \end{aligned}$$

where (a) and (b) follow from the definition of $\gamma(r)$ and z . We skip the proof of (c) for brevity (cf. Appendix of [16]). Therefore, the claim holds.

Now, let $\bar{D} = T(1 - R) + \left(2TR \log\left(\frac{1}{\epsilon}\right)\right)^{\frac{1}{2}}$. Using the claim that we just proved, we may verify that $p(\bar{D}) \leq \bar{p}_b(\bar{D}) = \epsilon$, if $\bar{D} \geq T(1 - R + \bar{r}(R))$. In order to check the hypothesis of the claim, note that for $R \leq 1$, $\bar{r}(R) = 0$ (cf. Lemma 3), and for all $1 \leq R \leq 1 + \left(\frac{1}{2T} \log\left(\frac{1}{\epsilon}\right)\right)^{\frac{1}{2}}$, we have

$$\begin{aligned} \bar{D} - T(1 - R) &= \left(2TR \log\left(\frac{1}{\epsilon}\right)\right)^{\frac{1}{2}} \\ &\geq 2T\left(\frac{1}{2T} \log\left(\frac{1}{\epsilon}\right)\right)^{\frac{1}{2}} \\ &\stackrel{(d)}{\geq} 2T(R - 1) \stackrel{(e)}{\geq} T\bar{r}(R), \end{aligned}$$

where inequality (d) follows from the hypothesis of Part (b), and inequality (e) is true by Lemma 3. Therefore, $D^*(\epsilon) \leq \bar{D}$ for all $R \leq 1 + \left(\frac{1}{2T} \log\left(\frac{1}{\epsilon}\right)\right)^{\frac{1}{2}}$. Note that, the upper bound that we obtained in Part (a) is also valid for all R . Hence, the minimum of the two gives the tightest bound. ■

When the arrival rate R is smaller than one (the playback rate), the upper bound in Theorem 1 consists of two components. The first term, $T(1 - R)$, compensates the expected number of packets that are required by the end of $[0, T]$ period.

The second component, $\left(2TR \log\left(\frac{1}{\epsilon}\right)\right)^{\frac{1}{2}}$, compensates the randomness of the arrivals to avoid interruptions with high probability. Note that this term increases by decreasing the maximum allowed interruption probability, and it would be zero for a deterministic arrival process. For the case when the arrival rate is larger than the playback rate, the minimum required buffer size does not grow with the file size. By continuity of the probability measure, we can show that the upper bound in Theorem 1 remains bounded for infinite file sizes. This is so since the buffer size in (1) has a positive drift. Hence, if there is no interruption at the beginning of the playback period, it becomes more unlikely to happen later.

In the following, we show that the upper bounds presented in Theorem 1 are *asymptotically tight*, by providing lower bounds on the minimum required buffer size $D^*(\epsilon)$, for different regimes of the arrival rate R . Let us first define the notion of a tight bound.

Definition 1. Let $\hat{D}$ be a lower or upper bound of the minimum buffer size $D^*(\epsilon)$ that depends on the file size T . The bound $\hat{D}$ is an *asymptotically tight* bound if $\frac{|\hat{D} - D^*(\epsilon)|}{D^*(\epsilon)}$ vanishes as T goes to infinity.

Theorem 2. [Converse] Let $D^*(\epsilon)$ be defined as in (4), and $\bar{r}(R)$ be given by (8). Then

(a) For all $R > 1$,

$$D^*(\epsilon) \geq -\frac{1}{\bar{r}(R)} \log\left(\epsilon + 2e^{-\frac{(R-1)^2}{4(R+1)}T}\right). \quad (15)$$

(b) For each $R \leq 1$ and $\epsilon \leq \frac{1}{16}$, if $T \geq C \log\left(\frac{1}{\epsilon}\right)$ then

$$D^*(\epsilon) \geq T(1 - R) + \frac{1}{2} \left(2TR \log\left(\frac{1}{\epsilon}\right)\right)^{\frac{1}{2}}, \quad (16)$$

where C is a constant that only depends on R .

Proof: We only sketch the proof of part (a) owing to space limitation (cf. [16] for a complete proof). Using a similar argument as in the proof of Theorem 1, it is sufficient to provide a lower bound on $p(D)$ defined in (3). Define τ_B as the first time that $Q(t)$ crosses a threshold B . We have

$$p(D) = \Pr\{\tau_e < \tau_f\} \geq \Pr\{\tau_e < \min(\tau_B, T - B)\}. \quad (17)$$

By Doob's optional stopping theorem for a properly defined martingale, the right-hand side of (17) is bounded below by

$$e^{-\bar{r}(R)D} - e^{-\bar{r}(R)B} - \Pr\{0 \leq Q(T - B) \leq B\}. \quad (18)$$

Now, we choose B proportional to T such that both of the second and third terms of (18) are bounded by $e^{-\frac{(R-1)^2}{4(R+1)}T}$. The result in (15) follows immediately from this. ■

Note that the result in part (b) of Theorem 2 does not hold for all ϵ . In fact, we can show that $D^*(\epsilon) < T(1 - R)$ for a large interruption probability ϵ . In the extreme case $\epsilon = 1$, it is clear that $D^*(\epsilon) = 0$. Nevertheless, since we are interested in *avoiding* interruptions, we do not study this regime of the interruption probabilities. Comparing the lower bounds obtained in Theorem 2 with the upper bounds obtained in Theorem 1, we observe that they demonstrate a similar behavior as a function of the parameters T and R . Now, we can show that the obtained bounds are asymptotically tight.

Corollary 1. The upper bounds and lower bounds of $D^*(\epsilon)$ given by Theorems 1 and 2 are asymptotically tight, if $R > 1$, or $R < 1$ and $\epsilon \leq \frac{1}{16}$.

Proof: Let D_l and D_u be lower and upper bounds of $D^*(\epsilon)$, respectively. By Definition 1, for D_l or D_u to be asymptotically tight, it is sufficient to show $\frac{D_u - D_l}{D_l}$ goes to zero as T grows. We may verify this claim by using the upper and lower bounds presented in Theorem 1 and Theorem 2, and taking the limit as T goes to infinity. ■

Next, we numerically obtain the optimal trade-off curve between the interruption probability and initial buffer size, and compare the results with the bounds derived earlier.

V. NUMERICAL RESULTS

We use MATLAB simulations to compute the minimum initial buffer size $D^*(\epsilon)$ for a given interruption probability ϵ in various scenarios. We start from a small initial buffer size D , and for each D we compute the interruption probability $p(D)$ via Monte-Carlo method. We increase D until the constraint $p(D) \leq \epsilon$ is satisfied. Since $p(D)$ is monotonically decreasing in D , this gives the minimum required buffer size. Here, we restrict D to take only integer values, and round each upper bound value up to the nearest integer, and each lower bound value down to the nearest integer.

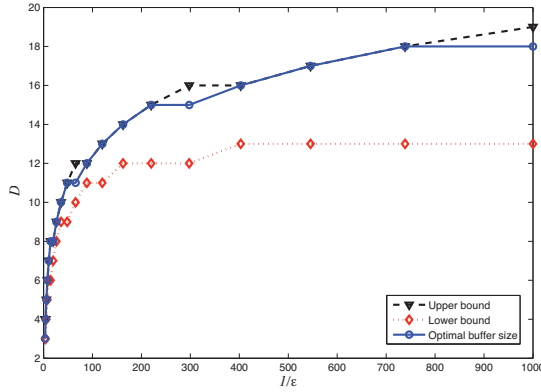


Fig. 2. The minimum buffer size $D^*(\epsilon)$ as a function of the interruption probability.

Figure 2 shows the minimum required buffer size $D^*(\epsilon)$ as well as the upper and lower bounds given by Theorems 1 and 2 as a function of $\frac{1}{\epsilon}$, where the arrival rate is fixed to $R = 1.2$ and the file size $T = 500$. We observe that the numerically computed trade-off curve closely matches our analytical results.

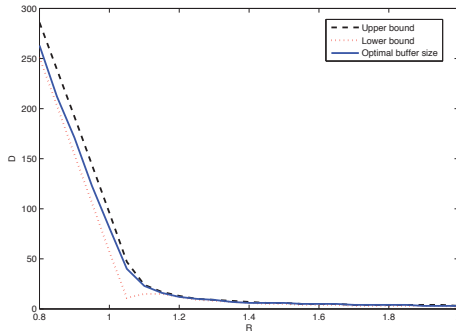


Fig. 3. The minimum buffer size $D^*(\epsilon)$ as a function of the arrival rate R .

Figure 3 plots the minimum required buffer size $D^*(\epsilon)$ as well as the upper and lower bounds given by Theorems 1 and 2 versus the arrival rate R , where $\epsilon = 10^{-2}$ and the file size is fixed to $T = 10^3$. Note that when the arrival rate is almost equal or less than the playback rate, increasing the arrival rate can significantly reduce the initial buffering delay. However, for larger arrival rates $D^*(\epsilon)$ is small enough such that increasing R does not help anymore.

VI. CONCLUSIONS

In this paper, we studied the problem of media streaming with focus on the trade-offs between the two QoE metrics—probability of interruption in media playback, and the initial waiting time before starting the playback. In our system, the user can receive packets of the media stream from multiple sources by requesting packets in each block of the file. We used the fact that sending random linear combinations of the packets within each block of the media file simplifies the packet selection strategies of P2P systems. This fact allowed us to describe the receiver's buffer dynamics as an M/D/1 queue, and explicitly characterize the trade-off between the QoE metrics for different ranges of the system parameters. We observed that the minimum initial buffer size to attain a desired level of interruption probability remains bounded as the file size grows if the arrival rate is slightly larger than the play rate. Further, when the arrival rate and the play rate match, the initial buffer size needs to scale as the square root of the file size to account for randomness of the arrivals. Finally, our numerical results confirmed that the optimal trade-off curves demonstrate a similar behavior to that predicted by our bounds.

This work is the first step in analytical characterization of QoE trade-offs in wireless media streaming applications. An interesting extension to this work would be to obtain optimal resource allocation policies to satisfy users who have different interruption probability and initial waiting time targets.

REFERENCES

- [1] C. Fraleigh, S. Moon, B. Lyle, C. Cotton, M. Khan, D. Moll, R. Rockell, T. Seely, and C. Diot. Packet-level traffic measurements from the Sprint IP backbone. *IEEE Network Magazine*, 17(6):6–16, 2003.
- [2] K. P. Gummadi, R. J. Dunn, S. Saroiu, S. D. Gribble, H. M. Levy, and J. Zahorjan. Measurement, modeling, and analysis of a peer-to-peer file-sharing workload. In *Proc. SOSP*.
- [3] C. Labovitz, D. McPherson, and S. Iekel-Johnson. 2009 Internet Observatory report. In *NANOG-47*, October 2009.
- [4] PPLive. <http://www.pplive.com/>, 2009.
- [5] QQLive. <http://www.qqlive.com/>, 2009.
- [6] Y. P. Zhou, D. M. Chiu, and J. C. S. Lui. A simple model for analyzing P2P streaming protocols. In *Proc. IEEE ICNP 2007*.
- [7] T. Bonald, L. Massoulié, F. Mathieu, D. Perino, and A. Twigg. Epidemic live streaming: optimal performance trade-offs. *SIGMETRICS Perform. Eval. Rev.*, 36(1):325–336, 2008.
- [8] Bridge Q. Zhao, John C.S. Lui, and Dah-Ming Chiu. Exploring the optimal chunk selection policy for data-driven P2P streaming systems. In *The 9th International Conference on Peer-to-Peer Computing*, 2009.
- [9] L. Ying, R. Srikant, and S. Shakkottai. The Asymptotic Behavior of Minimum Buffer Size Requirements in Large P2P Streaming Networks. In *Proc. of the Information Theory and Applications Workshop (to appear)*, San Diego, CA, February 2010.
- [10] T. Ho, R. Koetter, M. Médard, M. Effros, J. Shi, and D. Karger. A random linear network coding approach to multicast. *IEEE Transactions on Information Theory*, 52:4413–4430, 2006.
- [11] S. Acedanski, S. Deb, M. Médard, and R. Koetter. How good is random linear coding based distributed networked storage. In *NetCod*, 2005.
- [12] C. Gkantsidis, J. Miller, and P. Rodriguez. Comprehensive view of a live network coding p2p system. In *Proc. ACM SIGCOMM*, 2006.
- [13] M. Wang and B. Li. R²: Random push with random network coding in live peer-to-peer streaming. *IEEE JSAC, Special Issue on Advances in Peer-to-Peer Streaming Systems*, 25:1655–1666, 2007.
- [14] H. Chi and Q. Zhang. Deadline-aware network coding for video on demand service over p2p networks. In *PacketVideo*, 2006.
- [15] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, 1997.
- [16] A. ParandehGheibi, M. Médard, S. Shakkottai, and A. Ozdaglar. Avoiding interruptions - QoE trade-offs in block-coded streaming media applications. submitted to ISIT 2010, arXiv:1001.1937 [cs.MM].