

# Access-Network Association Policies for Media Streaming in Heterogeneous Environments

Ali ParandehGheibi<sup>†</sup>, Muriel Médard<sup>†</sup>, Asuman Ozdaglar<sup>†</sup>, Srinivas Shakkottai<sup>‡</sup>  
 parandeh@mit.edu, medard@mit.edu, asuman@mit.edu, sshakkot@tamu.edu

**Abstract**—We study the design of media streaming applications in the presence of multiple heterogeneous wireless access methods with different throughputs and costs. Our objective is to analytically characterize the trade-off between the usage cost and the Quality of user Experience (QoE), which is represented by the probability of interruption in media playback and the initial waiting time. We model each access network as a server that provides packets to the user according to a Poisson process with a certain rate and cost. Blocks are coded using random linear codes to alleviate the duplicate packet reception problem. Users must take decisions on how many packets to buffer before playout, and which networks to access during playout.

We design, analyze and compare several control policies with a threshold structure. We formulate the problem of finding the optimal control policy as an MDP with a probabilistic constraint. We present the HJB equation for this problem by expanding the state space, and exploit it as a verification method for optimality of the proposed control law.

## I. INTRODUCTION

Media streaming is fast becoming the dominant application on the Internet [1]. The popularity of such media transfers has been accompanied by the growing usage of wireless handheld devices as the preferred means of media access. It is expected that such media streaming would happen in both a device to device (D2D) as well as in a base-station to device fashion, and both the hardware and applications needed for such communication schemes are already making an appearance [2], [3].

Media streaming is achieved by dividing a file into *blocks*, which are then further divided into packets for transmission. After each complete block is received, it can be decoded and played out. Since we consider a streaming application, blocks inherently have a sequence associated with them, and each block must be received by the time the previous one has been played out. The absence of a block at the time of playout would cause a frame freeze, which is to be avoided if possible. When there are multiple networks that can be used to access a particular piece of content (e.g. from a base station or a peer device) each device must take decisions on associating with one or more such access networks. However, the costs of different access methods might be different. For example, accessing the base station of a cellular network can result in additional charges per packet, while it might be possible to receive the same packets from the access point of a local WLAN or another device with a lower cost or possibly for

free. Further, the cost of communication might be mitigated by the initial amount of buffering before playout. Hence, there are trade-offs between the probability of frame skipping, the initial waiting time, and the cost of different access methods available.

The objective of this paper is to understand the trade-off between initial buffering, and the usage of low and costly communication methods for attaining a target probability of skip-free playout. We consider a system wherein network coding is used to ensure that packet identities can be ignored, and packets may potentially be obtained from two sources (servers) that have different rates of transmission. The wireless channel is unreliable, and we assume that each server can deliver packets according to a Poisson process with a known rate. Further, the costs of accessing the two servers are different; for simplicity we assume that one of the servers is free. Thus, *our goal is to develop an algorithm that switches between the free and the costly servers in order to attain a target probability of skipping at lowest cost.*

Our contributions are as follows. We first develop an analytical characterization of the interruption probability for the single server case. Using this result, we obtain a lower bound on the cost of offline policies that do not observe the trajectory of packets received. We show that such policies have a threshold form in terms of the time of association with the costly server. Using the offline algorithm as a starting point, we develop an *online* algorithm with lower cost that has a threshold form – both free and costly servers are used until the queue length reaches a threshold, followed by only free server usage. We then develop an online algorithm in which the risk of interruption is spread out across the trajectory. Here, only the free server is used whenever the queue length is above a certain threshold, while both servers are used when the queue length is below the threshold. The threshold is designed as a function of the initial buffer size and the desired interruption probability.

We formulate the problem of finding the optimal network association policy as a Markov Decision Process with a probabilistic constraint. Similarly to the Bellman equation proposed by Chen [4] for a discrete time MDP with probabilistic constraints, we write the Hamilton-Jacobi-Bellman equation for the problem. Using a guess and check approach, we derive an approximate solution of the HJB equation, and show that the optimal policy given by the approximate value function takes a threshold form.

Media streaming, particularly in the area of P2P networks has attracted significant recent interest. For example, work

<sup>†</sup> Dept. of Electrical Engineering and Computer Science, MIT.

<sup>‡</sup> Dept. of Electrical and Computer Engineering, Texas A&M University.

such as [5], [6], [7] develop analytical models on the trade-off between the steady state probability of missing a block and buffer size under different block selection policies. Unlike our model, they consider live streaming with deterministic channels. The use of random linear codes considerably simplifies packet selection [8], [9], [10], [11], and we can use the same idea to ensure that packets can be received from multiple sources without the need to coordinate the exact identities of the packets from each. However, we focus on content that is already cached at multiple locations, and must be streamed over one or more unreliable channels. Related to our work is [12], which considers two possible wireless access methods (WiFi and UMTS) for file delivery, assuming particular throughput models for each access method. In contrast to this work, packet arrivals are stochastic in our model, and our streaming application requires hard constraints on quality of user experience.

## II. SYSTEM MODEL AND QOE METRICS

We consider a media streaming system as follows. A single user is receiving a media file of size  $F$  from various servers it is connected to. Each server could be a wireless access point or another wireless user operating as a server. The receiver first buffers  $D$  packets from the beginning of the file, and then starts the playback.

We assume that time is continuous, and the arrival process of packets from each server is a Poisson process independent of other arrival processes. Further, we assume that no redundant packet is delivered from different servers. This assumption can be justified if there is no delay in the feedback to the servers, or by sending random linear combination of the packets in the server (see [13] and [14] for more details). Therefore, we can combine the arrival processes of any subset  $S$  of the servers into one Poisson process of rate  $R_S$  equal to the summation of the rates from the corresponding servers.

There are two types of servers in the system: free servers and the costly ones. There is no cost associated with receiving packets from a free server, but a unit cost is incurred for each (coded) packet delivered by any costly server. As described above, we can combine all the free servers into one free server from which packets arrive according to a Poisson process of rate  $R_0$ . Similarly, we can merge all of the costly servers into one costly server with effective rate of  $R_c$ . At any time  $t$ , the user has the option to use packets only from the free server or from both the free and the costly servers. In the latter case, the packets arrive according to a Poisson process of rate  $R_1 = R_0 + R_c$ . The user's action at time  $t$  is denoted by  $u_t \in \{0, 1\}$ , where  $u_t = 0$  if only the free server is used at time  $t$ , while  $u_t = 1$  if both free and costly servers are used. We normalize the playback rate to one, i.e., it takes one unit of time to play a single packet. We also assume that the parameters  $R_0$  and  $R_1$  are known at the receiver.

The dynamics of the receiver's buffer size (queue-length)  $x_t$  can be described as follows

$$x_t = D + N_t + \int_0^t u_\tau dN_\tau^c - t, \quad (1)$$

where  $D$  is the initial buffer size,  $N_t$  Poisson processes of rate  $R_0$  and  $N_t^c$  is a Poisson counter of rate  $R_c$  which is independent of the process  $N_t$ . The last term correspond to the unit rate of media playback.

The user's association (control) policy is formally defined below.

**Definition 1.** [Control Policy] Let  $h_t = \{x_s : 0 \leq s \leq t\} \cup \{u_s : 0 \leq s < t\}$  denote the history of the buffer sizes and actions up to time  $t$ , and  $\mathcal{H}$  be the set of all histories for all  $t$ . A *deterministic association policy* denoted by  $\pi$  is a mapping  $\pi : \mathcal{H} \mapsto \{0, 1\}$ , where at any time  $t$

$$\pi(h_t) = \begin{cases} 0, & \text{if only the free server is chosen,} \\ 1, & \text{if both servers are chosen.} \end{cases}$$

Denote by  $\Pi$  the set of all such control policies.

We can declare an interruption in playback when the buffer size decreases to zero before reaching the end of the file, i.e., when there is no packet at the receiver to be played but the file is not completely downloaded. More precisely, let

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

where  $\tau_f$  corresponds to time of completing the file download, because we have already played  $\tau_f$  packets and the buffer contains the remaining  $F - \tau_f$  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 for a particular control policy  $\pi$  denoted by

$$p^\pi(D) = \mathbf{P}\{\tau_e < \tau_f\}, \quad (3)$$

where  $\tau_e$  and  $\tau_f$  are defined in (2).

**Definition 2.** The policy  $\pi$  is defined to be  $(D, \epsilon)$ -feasible if  $p^\pi(D) \leq \epsilon$ . The set of all such feasible policies is denoted by  $\Pi(D, \epsilon)$ .

The third metric that we consider in this work is the expected cost of using the costly server which is proportional to the expected usage time of the costly server. For any  $(D, \epsilon)$ , the usage cost of a  $(D, \epsilon)$ -feasible policy  $\pi$  is given by<sup>1</sup>

$$J^\pi(D, \epsilon) = \mathbf{E} \left[ \int_0^F u_t dt \right]. \quad (4)$$

The value function or optimal cost function  $V$  is defined as

$$V(D, \epsilon) = \min_{\pi \in \Pi(D, \epsilon)} J^\pi(D, \epsilon), \quad (5)$$

and the optimal policy  $\pi^*$  is defined as the optimal solution of the minimization problem in (5).

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  $\epsilon$ , the initial buffer size  $D$ , and the

<sup>1</sup>Throughout this work, we use the convention that the cost of an infeasible policy is infinite.

usage cost. These trade-offs depend on the association policy as well as the system parameters  $R_0$ ,  $R_c$  and  $F$ .

We first characterize the trade-offs between the QoE metrics for degenerate control policies. Next, we use these results to design association policies.

### III. QOE TRADE-OFFS FOR THE SINGLE-SERVER PROBLEM

Consider a single-server problem where the receiver receives the packets according to a Poisson process of rate  $R$ . The user's only decision in this case is the initial buffer size  $D$ . We would like to characterize the optimal trade-off between the initial buffer size and the interruption probability  $p(D)$  by providing bounds on the interruption probability as a function of the system parameters  $R$  and  $F$ . An upper bound (achievability) on  $p(D)$  is particularly useful, since it provides a sufficient condition for desirable user experience. A lower bound (converse) on  $p(D)$  demonstrates how tight the upper bound is.

**Theorem 1.** *For the initial buffer size  $D$ , let  $p(D)$  be the interruption probability of a single-server system defined as in (3). Define  $\gamma(r)$  as*

$$\gamma(r) = r + R(e^{-r} - 1), \quad (6)$$

and  $\bar{r}(R)$  as the largest root of  $\gamma(r)$ , i.e.,

$$\bar{r}(R) = \sup\{r : \gamma(r) = 0\}. \quad (7)$$

Then for all  $R > 1$ ,

$$e^{-\bar{r}(R)D} - 2e^{-\frac{(R-1)^2}{4(R+1)}F} \leq p(D) \leq e^{-\bar{r}(R)D}. \quad (8)$$

*Proof:* We do not include the proof owing to space limitations. See [14] for a complete proof. ■

Note that the upper bounds and lower bounds of  $p(D)$  given by Theorem 1 are asymptotically tight as  $F$  goes to infinity. Therefore, for  $F = \infty$ , by continuity of the probability measure we get

$$p(D) = \mathbf{P}\left(\min_{t \geq 0} x_t \leq 0 \mid x_0 = D\right) = e^{-\bar{r}(R)D}. \quad (9)$$

Using this characterization, we can identify the ranges of the QoE metrics for which there exists no feasible policy or the costly server is not required.

**Corollary 1.** (a) *For any  $(D, \epsilon)$  such that  $D \geq \frac{1}{\bar{r}(R_0)} \log\left(\frac{1}{\epsilon}\right)$ ,*

$$\min_{\pi \in \Pi} J^\pi(D, \epsilon) = 0.$$

(b) *For any  $(D, \epsilon)$  such that  $D < \frac{1}{\bar{r}(R_1)} \log\left(\frac{1}{\epsilon}\right)$ ,*

$$\min_{\pi \in \Pi} J^\pi(D, \epsilon) = \infty.$$

*Proof:* Consider the degenerate policy  $\pi_0 \equiv 0$ . This policy is equivalent to a single-server system with arrival rate  $R = R_0$ . By Definition 2, and (9), the policy  $\pi_0$  is  $(D, \epsilon)$ -feasible for all  $D \geq \frac{1}{\bar{r}(R_0)} \log\left(\frac{1}{\epsilon}\right)$ . Note that by (4) this policy does not incur any cost, which results in part (a).

Moreover, for all  $(D, \epsilon)$  with  $D < \frac{1}{\bar{r}(R_1)} \log\left(\frac{1}{\epsilon}\right)$ , there is no  $(D, \epsilon)$ -feasible policy. This is so since the buffer size under

any policy  $\pi$  is stochastically dominated by the one governed by the degenerate policy  $\pi_1 \equiv 1$ . Hence,

$$p^\pi(D) \geq p^{\pi_1}(D) = \exp(-\bar{r}(R_1)D) > \epsilon.$$

Using the convention of infinite cost for infeasible policies, we obtain the result in part (b). ■

For simplicity of notation, let  $\alpha_0 = \bar{r}(R_0)$ , and  $\alpha_1 = \bar{r}(R_1)$ . Throughout the rest of this paper, we study the case that the file size  $F$  is infinite, since the control policies in this case take simpler forms and the cost of such control policies provide an upper bound for the finite file size case. Further, by Corollary 1 we focus on the region

$$\mathcal{R} = \left\{ (D, \epsilon) : \frac{1}{\alpha_1} \log\left(\frac{1}{\epsilon}\right) \leq D \leq \frac{1}{\alpha_0} \log\left(\frac{1}{\epsilon}\right) \right\} \quad (10)$$

to analyze the expected cost of various classes of control policies.

### IV. DESIGN AND ANALYSIS OF ASSOCIATION POLICIES

In this section, we propose several classes of parameterized control policies. We first characterize the range of the parameters for which the association policy is feasible for a given initial buffer size  $D$  and the desired level of interruption probability  $\epsilon$ . Then, we try to choose the parameters such that the expected cost of the policy is minimized.

#### A. Off-line Policy

Consider the class of policies where the decisions are made off-line before starting media streaming. In this case, the arrival process is not observable by the decision maker. Therefore, the user's decision space reduces to the set of deterministic functions  $u : \mathbb{R} \rightarrow \{0, 1\}$ , that maps time into the action space.

**Theorem 2.** *Let the cost of a control policy be defined as in (4). In order to find a minimum-cost off-line policy, it is sufficient to consider policies of the following form:*

$$\pi(h_t) = u_t = \begin{cases} 1, & \text{if } t \leq t_s \\ 0, & \text{if } t > t_s. \end{cases} \quad (11)$$

*Proof:* In general any off-line policy  $\pi$  consists of multiple intervals in which the costly server is used. Consider an alternative policy  $\pi'$  of the form of (11) where  $t_s = J^\pi$ . By definition of the cost function in (4) the two policies incur the same cost. Moreover, the buffer size process under policy  $\pi$  is stochastically dominated by the one under policy  $\pi'$ , because the policy  $\pi'$  counts the arrivals from the costly server earlier, and the arrival process is stationary. Hence, the interruption probability of  $\pi'$  is not larger than that of  $\pi$ . Therefore, for any off-line policy, there exists another off-line policy of the form given by (11). ■

**Theorem 3.** *Consider the class of off-lines policies of the form (11). For any  $(D, \epsilon) \in \mathcal{R}$ , the policy  $\pi$  defined in (11) is feasible if*

$$t_s \geq t_s^* = \frac{R_0}{R_1 - R_0} \left[ \frac{1}{\alpha_0} \log\left(\frac{1}{\epsilon - e^{-\alpha_1 D}}\right) - D \right]. \quad (12)$$

*Proof:* By Definition 2, we need to show that  $p^\pi(D) \leq \epsilon$ . By a union bound on the interruption probability, it is sufficient to verify

$$\mathbf{P}\left(\min_{0 \leq t \leq t_s} x_t \leq 0 \mid x_0 = D\right) + \mathbf{P}\left(\min_{t > t_s} x_t \leq 0 \mid x_0 = D\right) \leq \epsilon. \quad (13)$$

In the interval  $[0, t_s]$ ,  $x_t$  behaves as in a single-server system with rate  $R_1$ . Hence, by Theorem 1 we get

$$\mathbf{P}\left(\min_{0 \leq t \leq t_s} x_t \leq 0 \mid x_0 = D\right) \leq e^{-\alpha_1 D}. \quad (14)$$

For the second term in (13), we have

$$\begin{aligned} & \mathbf{P}\left(\min_{t > t_s} x_t \leq 0 \mid x_0 = D\right) \\ &= \sum_{q=D-t_s}^{\infty} \mathbf{P}\left(\min_{t > t_s} x_t \leq 0 \mid x_{t_s} = q\right) \mathbf{P}(x_{t_s} = q) \\ &\stackrel{(a)}{\leq} \sum_{q=D-t_s}^{\infty} e^{-\alpha_0 q} \mathbf{P}(x_{t_s} = q) \\ &= \sum_{k=0}^{\infty} e^{-\alpha_0(D+k-t_s)} \mathbf{P}(N_{t_s} + N_{t_s}^c = k) \\ &\stackrel{(b)}{=} \sum_{k=0}^{\infty} e^{-\alpha_0(D+k-t_s)} \frac{e^{-R_1 t_s} (R_1 t_s)^k}{k!} \\ &= e^{-\alpha_0(D-t_s) + R_1 t_s (e^{-\alpha_0} - 1)} \sum_{k=0}^{\infty} \frac{e^{-R_1 t_s e^{-\alpha_0}} (R_1 t_s e^{-\alpha_0})^k}{k!} \\ &= \exp\left(-\alpha_0(D-t_s) + R_1 t_s (e^{-\alpha_0} - 1)\right) \cdot 1 \\ &\stackrel{(c)}{=} \exp\left(-\alpha_0(D-t_s) + R_1 t_s \left(-\frac{\alpha_0}{R_0}\right)\right) \\ &\stackrel{(d)}{\leq} \epsilon - e^{-\alpha_1 D}, \end{aligned}$$

where (a) follows from Theorem 1 and the fact that  $u_t = 0$ , for  $t \geq t_s$ . (b) is true because  $N_{t_s} + N_{t_s}^c$  is a Poisson random variable with mean  $R_1 t_s$ . (c) holds since  $\alpha_0$  is the root of  $\gamma(r)$  defined in (6) for  $R = R_0$ . Finally, (d) follows from the hypothesis of the theorem.

By combining the above bounds, we may verify (13) which in turns proves feasibility of the proposed control policy. ■

Note that obtaining the optimal off-line policy is equivalent to finding the smallest  $t_s$  for which the policy is still feasible. Therefore,  $t_s^*$  given in (12) provides an upper bound on the minimum cost of an off-line policy. Observe that  $t_s^*$  is almost linear in  $D$  for all  $(D, \epsilon)$  that is not too close to the lower boundary of region  $\mathcal{R}$ . As  $(D, \epsilon)$  gets closer to the boundary,  $t_s^*$  and the expected cost grows to infinity, which is in agreement with Corollary 1. In this work we pick  $t_s^*$  as a benchmark for comparison to other policies that we present next.

## B. Online Safe Policy

Let us now consider the class of online policies where the decision maker can observe the buffer size history. Inspired by the structure of the optimal off-line policies, we first focus on a *safe* control policy in which in order to avoid interruptions, the costly server is used at the beginning until the buffer size

reaches a certain threshold after which the costly server is never used. This policy is formally defined below.

**Definition 3.** The online safe policy  $\pi^S$  parameterized by the threshold value  $S$  is given by

$$\pi^S(h_t) = \begin{cases} 1, & \text{if } t \leq \tau_S \\ 0, & \text{if } t > \tau_S, \end{cases} \quad (15)$$

where  $\tau_S = \inf\{t \geq 0 : x_t \geq S\}$ .

**Theorem 4.** Let  $\pi^S$  be the safe policy defined in Definition 3. For any  $(D, \epsilon) \in \mathcal{R}$ , the safe policy is feasible if

$$S \geq S^* = \frac{1}{\alpha_0} \log\left(\frac{1}{\epsilon - e^{-\alpha_1 D}}\right). \quad (16)$$

Moreover,

$$\begin{aligned} \min_{S \geq S^*} J^{\pi^S}(D, \epsilon) &= J^{\pi^{S^*}}(D, \epsilon) \\ &= \frac{1}{R_1 - 1} \left[ \frac{1}{\alpha_0} \log\left(\frac{1}{\epsilon - e^{-\alpha_1 D}}\right) - D + \xi \right], \end{aligned}$$

where  $\xi \in [0, 1)$ .

*Proof:* Similar to the proof of Theorem 3, we need to show that the total probability of interruption before and after crossing the threshold  $S$  is bounded from above by  $\epsilon$ . Observe that for any realization of  $\tau_S$  the bound in (14) still holds. Further, since the costly server is not used after crossing the threshold and  $x_{\tau_S} \geq S$ , Theorem 1 implies

$$\mathbf{P}\left(\min_{t > \tau_S} x_t \leq 0 \mid x_0 = D\right) \leq e^{-\alpha_0 S} \leq \epsilon - e^{-\alpha_1 D}, \quad (17)$$

where the second inequality follows from (16). Finally, combining (14) and (17) gives  $p^{\pi^S}(D) \leq \epsilon$ , which is the desired feasibility result.

For the second part, first observe that  $J^{\pi^S}(D, \epsilon) = \mathbf{E}[\tau_S]$ . In order to cross a threshold  $S \geq S^*$ , the threshold  $S^*$  must be crossed earlier, because  $x_0 = D \leq S^*$ . Hence,  $\tau_S$  stochastically dominates  $\tau_{S^*}$ , implying

$$J^{\pi^S}(D, \epsilon) = \mathbf{E}[\tau_S] \geq \mathbf{E}[\tau_{S^*}] = J^{\pi^{S^*}}(D, \epsilon), \quad \text{for all } S \geq S^*.$$

It only remains to compute  $\mathbf{E}[\tau_{S^*}]$ . It follows from Wald's identity or Doob's optional stopping theorem [15] that

$$D + (R_1 - 1)\mathbf{E}[\tau_{S^*}] = \mathbf{E}[x_{\tau_{S^*}}] = S^* + \xi, \quad (18)$$

where  $\xi \in [0, 1)$  because the jumps of a Poisson process are of units size, and hence the overshoot size when crossing a threshold is bounded by one, i.e.,  $S^* \leq x_{\tau_{S^*}} < S^* + 1$ . Rearranging the terms in (18) and plugging the value of  $S^*$  from (16) immediately gives the result. ■

Let us now compare the online safe policy  $\pi^{S^*}$  with the off-line policy defined in (11) with parameter  $t_s^*$  as in (12). We observe that the cost of the online safe policy is almost proportional to that of the off-line policy, where the cost ratio of the off-line policy to that of the online safe policy is given by

$$\frac{R_0(R_1 - 1)}{R_1 - R_0} = 1 + \frac{R_1(R_0 - 1)}{R_1 - R_0} > 1.$$

Note that the structure of both policies is the same, i.e., both policies use the costly server for a certain period of time

and then switch back to the free server. As suggested here, the advantage of observing the buffer size allows the online policies to avoid excessive use of the costly server when there are sufficiently large number of arrivals from the free server. In the following, we present another class of online policies.

### C. Online Risky Policy

In this part, we study a class of online policies where the costly server is used only if the buffer size is below a certain threshold. We call such policies “risky” as the risk of interruption is spread out across the whole trajectory unlike the “safe” policies. Further, we constrain risky policies to possess the property that the action at a particular time should only depend on the buffer size at that time, i.e., such policies are *stationary Markov* with respect to buffer size as the state of the system. The risky policy is formally defined below.

**Definition 4.** The online risky policy  $\pi^T$  parameterized by the threshold value  $T$  is given by

$$\pi^T(h_t) = \pi^T(x_t) = \begin{cases} 1, & \text{if } 0 < x_t < T \\ 0, & \text{otherwise.} \end{cases} \quad (19)$$

**Lemma 1.** Let  $x_t$  be the buffer size of a single-server system with arrival rate  $R > 1$ . Let the initial buffer size be  $D$  and for any  $T \geq D > 0$  define the following stopping times

$$\tau_T = \inf\{t > 0 : x_t \geq T\}, \quad \tau_e = \inf\{t \geq 0 : x_t \leq 0\}. \quad (20)$$

Then

$$\mathbf{P}(\tau_e > \tau_T) = \frac{1 - e^{-\bar{r}(R)D}}{1 - \mathbf{E}[e^{-\bar{r}(R)x_{\tau_T}} | \tau_e > \tau_T]}, \quad (21)$$

where  $\bar{r}(R)$  is defined in (7).

*Proof:* Let  $Y(t) = e^{-\bar{r}(R)x_t}$ . We may verify that  $Y(t)$  is a martingale and uniformly integrable. Also, define the stopping time  $\tau = \min\{\tau_T, \tau_e\}$ . Since  $R > 1$ , we have  $\mathbf{P}(\tau \geq t) \leq \mathbf{P}(0 < x_t < T) \rightarrow 0$ , as  $t \rightarrow \infty$ . Hence,  $\tau < \infty$  almost surely. Therefore, we can employ Doob’s optional stopping theorem [15] to write

$$\begin{aligned} e^{-\bar{r}(R)D} &= \mathbf{E}[Y(0)] = \mathbf{E}[Y(\tau)] \\ &= \mathbf{P}(\tau_e \leq \tau_T) \cdot 1 \\ &\quad + \mathbf{P}(\tau_e > \tau_T) \mathbf{E}[e^{-\bar{r}(R)x_{\tau_T}} | \tau_e > \tau_T]. \end{aligned}$$

The claim immediately follows from the above relation after rearranging the terms.  $\blacksquare$

**Theorem 5.** Let  $\pi^T$  be the risky policy defined in Definition 4. For any  $(D, \epsilon) \in \mathcal{R}$ , the policy  $\pi^T$  is feasible if the threshold  $T$  satisfies

$$T \geq T^* = \begin{cases} \frac{1}{\alpha_1 - \alpha_0} \left[ \log\left(\frac{\beta}{\epsilon}\right) - \alpha_0 D \right], & \text{if } D \geq \bar{D}, \\ \frac{1}{\alpha_1} \log\left(\frac{\epsilon + \beta(1 - e^{-\alpha_1 D}) - 1}{\epsilon - e^{-\alpha_1 D}}\right), & \text{if } D \leq \bar{D}, \end{cases} \quad (22)$$

where  $\beta = \frac{\alpha_1}{\alpha_0(1 - \frac{\alpha_0}{2})}$  and  $\bar{D} = \frac{1}{\alpha_1} \log\left(\frac{\beta}{\epsilon}\right)$ .

*Proof:* Let us first characterize the interruption probability of the policy  $\pi^T$  when the initial buffer size is  $D = T$ . In this

case, by definition of  $\pi^T$  the behavior of  $x_t$  is initially the same as a single-server system with rate  $R_1$  until the threshold  $T$  is crossed. Hence, by Lemma 1 we have

$$\begin{aligned} p^{\pi^T}(T) &= \mathbf{P}\left(\min_{t \geq 0} x_t \leq 0 | x_0 = T\right) \\ &= \mathbf{P}(\tau_e < \tau_T) \cdot 1 \\ &\quad + \mathbf{P}(\tau_T < \tau_e) \mathbf{P}\left(\min_{t \geq \tau_T} x_t \leq 0 | \tau_T < \tau_e, x_0 = T\right) \\ &= \frac{e^{-\alpha_1 T} - \mathbf{E}[e^{-\alpha_1 x_{\tau_T}} | \tau_e > \tau_T]}{1 - \mathbf{E}[e^{-\alpha_1 x_{\tau_T}} | \tau_e > \tau_T]} \\ &\quad + \frac{(1 - e^{-\alpha_1 T}) \mathbf{P}\left(\min_{t \geq \tau_T} x_t \leq 0 | \tau_T < \tau_e, x_0 = T\right)}{1 - \mathbf{E}[e^{-\alpha_1 x_{\tau_T}} | \tau_e > \tau_T]}. \end{aligned} \quad (23)$$

Further,

$$\begin{aligned} &\mathbf{P}\left(\min_{t \geq \tau_T} x_t \leq 0 | \tau_T < \tau_e, x_0 = T\right) \\ &= \int_T^{T+1} \mathbf{P}\left(\min_{t \geq \tau_T} x_t \leq 0 | x_{\tau_T}\right) d\mu(x_{\tau_T}) \\ &\stackrel{(a)}{=} \int_T^{T+1} \mathbf{P}\left(\min_{t \geq 0} x_t \leq 0 | x_0\right) d\mu(x_0) \\ &\stackrel{(b)}{=} \int_T^{T+1} \mathbf{P}\left(\min_{t \geq 0} x_t \leq 0 | \min_{t \geq 0} x_t \leq T, x_0\right) \\ &\quad \times \mathbf{P}\left(\min_{t \geq 0} x_t \leq T | x_0\right) d\mu(x_0) \\ &\stackrel{(c)}{=} \int_T^{T+1} p^{\pi^T}(T) e^{-\alpha_0(x_0 - T)} d\mu(x_0) \\ &= \mathbf{E}[e^{-\alpha_0(x_{\tau_T} - T)} | \tau_T < \tau_e] p^{\pi^T}(T), \end{aligned} \quad (24)$$

where  $\mu$  denotes the conditional distribution of  $x_{\tau_T}$  given  $\tau_T < \tau_e$ . Note that  $x_{\tau_T} \in [T, T+1]$  because the size of the overshoot is bounded by one. Further, (a) follows from stationarity of the arrival processes and the control policy, (b) holds because a necessary condition for the interruption event is to cross the threshold  $T$  when starting from a point  $x_0 \geq T$ . Finally (c) follows from (9) and the definition of the risky policy. The relations (23) and (24) together result in

$$p^{\pi^T}(T) = \frac{e^{-\alpha_1 T} (1 - \mathbf{E}_\mu[e^{-\alpha_1(x_{\tau_T} - T)])}{1 - \mathbf{E}_\mu[e^{-\alpha_0(x_{\tau_T} - T)]} + \kappa}, \quad (25)$$

where  $\kappa = \mathbf{E}_\mu[e^{-\alpha_0 x_{\tau_T} - (\alpha_1 - \alpha_0)T}] - \mathbf{E}_\mu[e^{-\alpha_1 x_{\tau_T}}] \geq 0$ . Therefore, using the fact that

$$1 - x \leq e^{-x} \leq 1 - x + \frac{x^2}{2}, \quad \text{for all } x \geq 0, \quad (26)$$

we can provide the following bound

$$\begin{aligned} p^{\pi^T}(T) &\leq \frac{e^{-\alpha_1 T} (\alpha_1 \mathbf{E}_\mu[x_{\tau_T} - T])}{\alpha_0 \mathbf{E}_\mu[x_{\tau_T} - T] \left(1 - \frac{\alpha_0}{2} \cdot \frac{\mathbf{E}_\mu[(x_{\tau_T} - T)^2]}{\mathbf{E}_\mu[x_{\tau_T} - T]}\right)} \\ &\leq \frac{\alpha_1}{\alpha_0(1 - \frac{\alpha_0}{2})} e^{-\alpha_1 T} = \beta e^{-\alpha_1 T}, \end{aligned} \quad (27)$$

where the last inequality holds since  $0 \leq x_{\tau_T} - \bar{D} \leq 1$ .

Now we prove feasibility of the risky policy  $\pi^{T^*}$  when  $D > \bar{D}$ . Observe that by (22),  $D > T^*$ , hence the behavior of the

buffer size  $x_t$  is the same as the one in a single-server system with rate  $R_0$  until the threshold  $T^*$  is crossed. Thus

$$\begin{aligned}
p^{\pi^{T^*}}(D) &= \mathbf{P}\left(\min_{t \geq 0} x_t \leq 0 \mid x_0 = D\right) \\
&= \mathbf{P}\left(\min_{t \geq 0} x_t \leq 0 \mid \min_{t \geq 0} x_t \leq T^*, x_0 = D\right) \\
&\quad \times \mathbf{P}\left(\min_{t \geq 0} x_t \leq T^* \mid x_0 = D\right) \\
&= p^{\pi^{T^*}}(T^*) e^{-\alpha_0(D-T^*)} \\
&\leq \beta e^{-(\alpha_1 - \alpha_0)T^* - \alpha_0 D} = \epsilon,
\end{aligned}$$

where the inequality follows from (27), and the last equality holds by (22).

Next we verify the feasibility of the policy  $\pi^{T^*}$  for  $D \leq \bar{D}$ . In this case,  $D \leq T^*$  and by definition of the risky policy the system behaves as a single-server system with arrival rate  $R_1$  until the threshold  $T^*$  is crossed or the buffer size hits zero (interruption). Hence, we can bound the interruption probability as follows

$$\begin{aligned}
p^{\pi^{T^*}}(D) &= \mathbf{P}(\tau_e < \tau_{T^*}) \cdot 1 \\
&\quad + \mathbf{P}(\tau_{T^*} < \tau_e) \mathbf{P}\left(\min_{t \geq \tau_{T^*}} x_t \leq 0 \mid \tau_{T^*} < \tau_e, x_0 = D\right) \\
&\stackrel{(a)}{=} 1 - \mathbf{P}(\tau_{T^*} < \tau_e) \left(1 - \mathbf{E}_\mu[e^{-\alpha_0(x_{\tau_{T^*}} - T^*)}] p^{\pi^{T^*}}(T^*)\right) \\
&\stackrel{(b)}{\leq} \frac{(\beta - 1)(1 - e^{-\alpha_1 D})}{1 - \mathbf{E}_\mu[e^{-\alpha_1 x_{\tau_{T^*}}}] + 1 - \beta(1 - e^{-\alpha_1 D})} \\
&\stackrel{(c)}{\leq} \frac{(\beta - 1)(1 - e^{-\alpha_1 D})}{1 - e^{-\alpha_1 T^*}} + 1 - \beta(1 - e^{-\alpha_1 D}) \stackrel{(d)}{=} \epsilon,
\end{aligned}$$

where (a) follows from (24), (b) can be verified after some manipulations by combining the result of Lemma 1 and (25), and (c) holds since  $\beta \geq 1$  and  $x_{\tau_{T^*}} \geq T^*$ . Finally, (d) immediately follows from plugging in the definition of  $T^*$  from (22).

Therefore, the risky policy  $\pi^{T^*}$  is feasible by Definition 2. Observe that the buffer size under any policy  $\pi^T$  of the form (19) with  $T \geq T^*$  stochastically dominates that of policy  $\pi^{T^*}$ , because  $\pi^T$  switches to the costly server earlier, and stays in that state longer. Hence,  $\pi^T$  is feasible for all  $T \geq T^*$ . ■

Theorem 5 facilitates the design of risky policies with a single-threshold structure, for any desired initial buffer size  $D$  and interruption probability  $\epsilon$ . For a fixed  $\epsilon$ , when  $D$  increases,  $T^*$  (the design given by Theorem 5) decreases to zero. On the other hand, if  $D$  decreases to  $\frac{1}{\alpha_1} \log\left(\frac{1}{\epsilon}\right)$  (the boundary of  $\mathcal{R}$ ), the threshold  $T^*$  quickly increases to infinity, i.e., the policy does not switch back to the free server unless a sufficiently large number of packets is buffered. Figure 1 plots  $T^*$  and  $D$  as a function of  $D$  for a fixed  $\epsilon$ . Observe that for large range of  $D$ ,  $T^* \leq D$ , i.e., the costly server is not initially used. In this range, owing to the positive drift of  $Q_t$ , the probability of ever using the costly server exponentially decreases in  $(D - T^*)$ .

Next we compute relatively tight bounds on the expected cost of the online risky policy and compare with the previously proposed policies.

**Theorem 6.** For any  $(D, \epsilon) \in \mathcal{R}$ , consider an online risky policy  $\pi^{T^*}$  defined in Definition 4, where the threshold  $T^*$  is

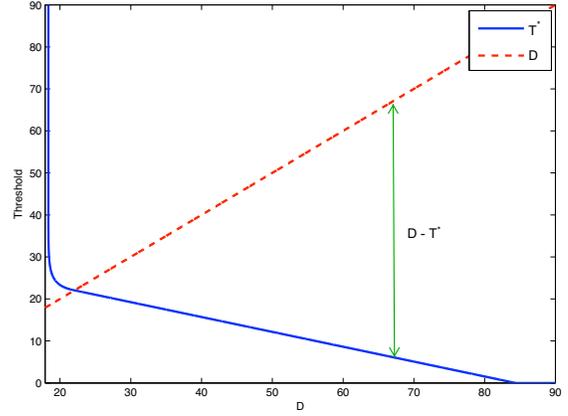


Fig. 1. The switching threshold of the online risky policy as a function of the initial buffer size for  $\epsilon = 10^{-3}$  (See Theorem 5).

given by (22) as function of  $D$  and  $\epsilon$ . If  $D \geq \bar{D}$  then

$$J^{\pi^{T^*}}(D, \epsilon) \leq \frac{\beta}{\alpha_1(R_1 - 1)} e^{-\alpha_0(D-T^*)}, \quad (28)$$

and if  $D \leq \bar{D}$

$$J^{\pi^{T^*}}(D, \epsilon) \leq \frac{1 - e^{-\alpha_1 D}}{(R_1 - 1)(1 - e^{-\alpha_1 T^*})} \left(T^* + 1 + \frac{\beta}{\alpha_1}\right) - \frac{D}{R_1 - 1}, \quad (29)$$

where  $\beta = \frac{\alpha_1}{\alpha_0(1 - \frac{\alpha_0}{2})}$  and  $\bar{D} = \frac{1}{\alpha_1} \log\left(\frac{\beta}{\epsilon}\right)$ .

*Proof:* Similarly to the proof of Theorem 5, we first consider the risky policy  $\pi^T$  with the initial buffer size  $T$ . By definition of  $\pi^T$ , the costly server is used until the threshold  $T$  is crossed. Thus the expected cost of this policy is bounded by the expected time until crossing the threshold plus the expected cost given that the threshold is crossed, i.e.,

$$J^{\pi^T}(T, \epsilon) \leq \frac{\mathbf{E}[x_{\tau_T}] - T}{R_1 - 1} + E[e^{-\alpha_0(x_{\tau_T} - T)}] J^{\pi^T}(T, \epsilon),$$

where  $\tau_T$  is defined in (20). The above relation implies

$$\begin{aligned}
J^{\pi^T}(T, \epsilon) &\leq \frac{1}{R_1 - 1} \cdot \frac{\mathbf{E}[x_{\tau_T} - T]}{1 - E[e^{-\alpha_0(x_{\tau_T} - T)}]} \\
&\leq \frac{1}{R_1 - 1} \cdot \frac{1}{\alpha_0 \left(1 - \frac{\alpha_0}{2} \cdot \frac{\mathbf{E}_\mu[(x_{\tau_T} - T)^2]}{\mathbf{E}_\mu[x_{\tau_T} - T]}\right)} \\
&\leq \frac{1}{\alpha_0(R_1 - 1)(1 - \frac{\alpha_0}{2})} = \frac{\beta}{\alpha_1(R_1 - 1)}, \quad (30)
\end{aligned}$$

where the second inequality follows from the fact in (26). Now for any  $D \geq \bar{D}$  we can write

$$\begin{aligned}
J^{\pi^{T^*}}(D, \epsilon) &= \mathbf{P}\left(\min_{t \geq 0} x_t \leq T^* \mid x_0 = D\right) J^{\pi^{T^*}}(T^*, \epsilon) \\
&= e^{-\alpha_0(D-T^*)} J^{\pi^{T^*}}(T^*, \epsilon)
\end{aligned}$$

where the inequality holds by Theorem 1. Combining this with (30) gives the result in (28).

If  $D \leq \bar{D}$ , the risky policy uses the costly server until the threshold  $T^*$  is crossed at  $\tau_{T^*}$  or the interruption event ( $\tau_e$ ), whichever happens first. Afterwards, no extra cost is incurred if an interruption has occurred. Otherwise, by (30) an extra cost of at most  $\frac{\beta}{\alpha_1(R_1 - 1)}$  is incurred, i.e.,

$$J^{\pi^{T^*}}(D, \epsilon) \leq \mathbf{E}[\min\{\tau_e, \tau_{T^*}\}] + \mathbf{P}(\tau_{T^*} < \tau_e) \frac{\beta}{\alpha_1(R_1 - 1)}.$$

By Doob's optional stopping theorem applied to the martingale  $Z_t = x_t - (R_1 - 1)t$ , we obtain

$$D = \mathbf{P}(\tau_{T^*} < \tau_e) \mathbf{E}[x_{\tau_{T^*}} | \tau_{T^*} < \tau_e] - (R_1 - 1) \mathbf{E}[\min\{\tau_e, \tau_{T^*}\}],$$

which implies

$$\mathbf{E}[\min\{\tau_e, \tau_{T^*}\}] \leq \frac{\mathbf{P}(\tau_{T^*} < \tau_e)(T^* + 1) - D}{R_1 - 1}$$

By combining the preceding relations we conclude that

$$J^{\pi^{T^*}}(D, \epsilon) \leq \frac{\mathbf{P}(\tau_{T^*} < \tau_e)}{R_1 - 1} \left( T^* + 1 + \frac{\beta}{\alpha_1} \right) - \frac{D}{R_1 - 1},$$

which immediately implies (29) by employing Lemma 1. ■

In the following, we compare the expected cost of the presented policies using numerical methods, and illustrate that the bounds derived in Theorems 3, 4 and 6 on the expected cost function are close to the exact value.

#### D. Performance Comparison

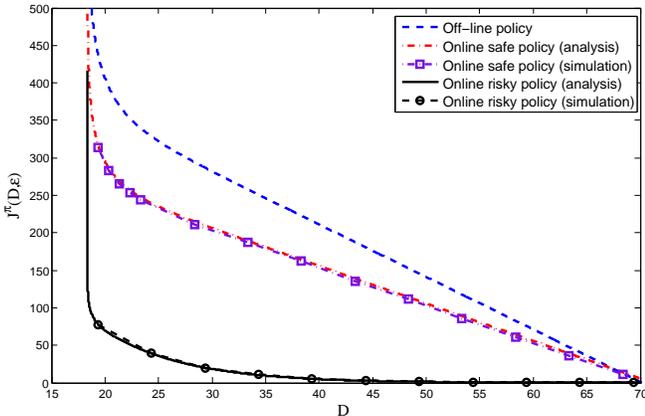


Fig. 2. Expected cost (units of time) of the presented control policies as a function of the initial buffer size for interruption probability  $\epsilon = 10^{-3}$ . The analytical bounds are given by Theorems 3, 4 and 6.

Figure 2 compares the expected cost functions of the off-line, online safe and online risky policies as a function of the initial buffer size  $D$ , when the interruption probability is fixed to  $\epsilon = 10^{-3}$ , the arrival rate from the free server is  $R_0 = 1.05$ , and the arrival rate from the costly server is  $R_c = R_1 - R_0 = 0.15$ . We plot the bounds on the expected cost given by Theorems 3, 4 and 6 as well as the expected cost function numerically computed by the Monte-Carlo method.

Observe that the expected cost of the risky policy is significantly smaller than both online safe and off-line policies. For example, the risky policy allows us to decrease the initial buffer size from 70 to 20 with an average of  $70 \times 0.15 \approx 10$  extra packets from the costly server. The expected cost in terms of the number packets received from the costly server is 43 and 61 for the online safe and off-line policy, respectively.

Moreover, note that it is merely the existence of the costly server as a backup that allows us to improve the user's quality of experience without actually using too many packets from the costly server. For example, observe that the risky policy

satisfies QoE metrics of  $D = 35$  and  $\epsilon = 10^{-3}$ , by only using on average about one extra packet from the costly server. However, without the costly server, in order to decrease the initial buffer size from 70 to 35, the interruption probability has to increase from  $10^{-3}$  to about 0.03 (see Theorem 1).

#### V. DYNAMIC PROGRAMMING APPROACH

In this section, we present a characterization of the optimal association policy in terms of the Hamilton-Jacobi-Bellman (HJB) equation. Note that because of the probabilistic constraint over the space of sample paths of the buffer size, the optimal policy is not necessarily Markov with respect to the buffer size as the state of the system. We take a similar approach as in [4] where by expanding the state space, a Bellman equation is provided as the optimality condition of an MDP with probabilistic constraint. In particular, consider the pair  $(x, p)$  as the state variable, where  $x$  denotes the buffer size and  $p$  represents the desired level of interruption probability. The evolution of  $x$  is governed by the following stochastic differential equation

$$dx = -dt + dN^u, \quad x_0 = D, \quad (31)$$

where  $N^u$  is a Poisson counter with rate  $R_u = R_0 + u \cdot R_c$ . For any  $(D, \epsilon) \in \mathcal{R}$  and any optimal policy  $\pi$ , the constraint  $p^\pi(D) \leq \epsilon$  is active. Hence, we consider the sample paths of  $p$  such that  $p_0 = \epsilon$  and  $\mathbf{E}[p_t] = \epsilon$  for all  $t$ , where the expectation is with respect to the Poisson jumps. Let  $\hat{p} = p + dp$  if a Poisson jump occurs in an infinitesimal interval of length  $dt$ . Also, let  $dp_0$  be the change in state  $p$  if no jump occurs. Therefore,

$$0 = \mathbf{E}[dp] = R_u dt (\hat{p} - p) + (1 - R_u dt) dp_0.$$

By solving the above equation for  $dp_0$ , we obtain the evolution of  $p$  as

$$dp = (p - \hat{p})(R_u dt - dN^u), \quad p_0 = \epsilon. \quad (32)$$

Similarly to the arguments of Theorem 2 of [4], by principle of optimality we can write the following dynamic programming equation

$$V(x, p) = \min_{u \in \{0,1\}, \hat{p} \in [0,1]} \{ u dt + \mathbf{E}[V(x + dx, p + dp)] \}. \quad (33)$$

If  $V$  is continuously differentiable, by Itô's Lemma for jump processes, we have

$$dV(x, p) = \frac{\partial V}{\partial x}(-dt) + \frac{\partial V}{\partial p} \cdot (p - \hat{p}) R_u dt + (V(x + 1, \hat{p}) - V(x, p)) dN^u,$$

which implies the following HJB equation

$$\frac{\partial V(x, p)}{\partial x} = \min_{u \in \{0,1\}, \hat{p} \in [0,1]} \left\{ u + \frac{\partial V}{\partial p} \cdot (p - \hat{p}) R_u + R_u (V(x + 1, \hat{p}) - V(x, p)) \right\} \quad (34)$$

The optimal policy  $\pi$  is obtained by characterizing the optimal solution of the partial differential equation in (34) together with the boundary condition  $V(x, 1) = 0$ . Since such equations are in general difficult to solve analytically, we use

the *guess and check* approach, where we propose a candidate for the value function and verify that it nearly satisfies the HJB equation almost everywhere. Moreover, we show that the trajectories of  $(x_t, p_t)$  steered by the optimal actions  $(u^*, \hat{p}^*)$  lie in a one-dimensional invariant manifold, leading to the risky policy defined in Definition 4.

For any  $(x, p) \in \mathcal{R}$  define

$$T(x, p) = \begin{cases} \frac{1}{\alpha_1 - \alpha_0} [\log(\frac{\theta}{p}) - \alpha_0 x], & \text{if } x \geq \frac{1}{\alpha_1} \log(\frac{\theta}{p}), \\ \frac{1}{\alpha_1} \log\left(\frac{p + \theta(1 - e^{-\alpha_1 x}) - 1}{p - e^{-\alpha_1 x}}\right), & \text{otherwise,} \end{cases} \quad (35)$$

where  $\theta = \frac{\alpha_1}{\alpha_0}$ . The candidate solution for HJB equation (34) is given by

$$\bar{V}(x, p) = \frac{1}{\alpha_0(1 - \frac{\alpha_0}{2})(R_1 - 1)} e^{-\alpha_0(x - T(x, p))}, \quad (36)$$

when  $x \geq \frac{1}{\alpha_1} \log(\frac{\theta}{p})$ , and

$$\bar{V}(x, p) = \frac{p + \theta(1 - e^{-\alpha_1 x}) - 1}{(R_1 - 1)(\theta - 1)} (T(x, p) + \frac{\beta}{\alpha_1}) - \frac{x}{R_1 - 1}, \quad (37)$$

when  $x < \frac{1}{\alpha_1} \log(\frac{\theta}{p})$ . Note that the candidate solution is derived from the structure if the expected cost of the risky policy (cf. Theorem 6). We may verify that  $\bar{V}$  satisfies the HJB equation (34) for all  $(x, p)$  such that  $x \geq \frac{1}{\alpha_1} \log(\frac{\theta}{p})$  or  $x \geq \frac{1}{\alpha_1} \log(\frac{\theta}{p}) - 1$ , but for other  $(x, p)$  the HJB equation is only approximately satisfied. This is due to bounding the overshoots, when computing the expected cost of the risky policy. The verification of HJB equation for our candidate solution is tedious but straightforward. We do not include it owing to space limitation.

**Theorem 7.** *Let  $\pi^*$  be the optimal association policy obtained from minimizing the right hand side of the HJB equation in (34) for the value function given by (36) and (37). Then  $\pi^*$  is a risky policy defined in Definition 4 with a threshold level  $T(D, \epsilon)$ , where  $D$  is the initial buffer size and  $\epsilon$  is the desired interruption probability.*

*Proof:* We sketch the proof owing to space limitation. First, we can show that the optimal action  $u^*(x, p)$  takes the following form

$$u^*(x, p) = \begin{cases} 0, & \text{if } x \geq \frac{1}{\alpha_1} \log(\frac{\theta}{p}); \\ 1, & \text{otherwise.} \end{cases} \quad (38)$$

Moreover, we may verify that for the initial condition  $(x_0, p_0) = (D, \epsilon)$ , the trajectory of  $(x_t, p_t)$  steered by the optimal actions  $(u^*, \hat{p}^*)$  is limited to a one-dimensional invariant manifold  $\mathcal{M}(D, \epsilon)$ , where

$$\mathcal{M}(D, \epsilon) = \left\{ (x, p) : p = \theta e^{-\alpha_0 x - (\alpha_1 - \alpha_0)T} \cdot \mathbb{I}_{\{x \geq T(D, \epsilon)\}} + \frac{(\theta - 1)e^{-\alpha_1 T(D, \epsilon)} - e^{-\alpha_1 x}(1 - \theta e^{-\alpha_1 T(D, \epsilon)})}{1 - e^{-\alpha_1 T(D, \epsilon)}} \mathbb{I}_{\{x < T(D, \epsilon)\}} \right\},$$

where  $T(D, \epsilon)$  is given by (35). Therefore, by plugging the above relation back into (38), we can show that the optimal action  $u^* = 0$  if and only if  $x \geq T(D, \epsilon)$ , i.e., the optimal policy given by the HJB equation is of the form of the risky policy in Definition 4 with threshold  $T = T(D, \epsilon)$ . ■

## VI. CONCLUSIONS AND FUTURE WORK

In this paper we studied the problem of selecting the access-networks in a heterogeneous wireless environment for media streaming applications. Our objective was to investigate the trade-offs between the network usage cost and the user's QoE requirements parameterized by initial waiting time and allowable probability of interruption in media playback. We analytically characterized and compared the expected cost of both off-line and online policies, finally showing that a threshold-based online risky policy achieves the lowest cost. Moreover, we derived an HJB equation for the problem of finding the optimal deterministic policy formulated as an MDP with a probabilistic constraint, and verified that the online risky policy nearly satisfies the HJB equation. Numerical analysis also confirmed our analytical results showing that merely the availability of a costly server used as a backup significantly improves QoE of media streaming without incurring a significant usage cost.

In the future, we would like to study more accurate models of channel variations such as the two-state Markov model due to Gillbert and Elliot. In this work we focused on deterministic network association policies. Another extension of this work would consist of studying randomized control policies. Finally, we would like to study more of the peer-to-peer aspect of the system to understand the decision making at the system level.

## REFERENCES

- [1] C. Labovitz, D. McPherson, and S. Iekel-Johnson. 2009 Internet Observatory report. In *NANOG-47*, October 2009.
- [2] R. Laroia. Future of Wireless? The Proximate Internet. In *Proc. of the Second International Conference on Communication Systems and Networks (COMSNETS)*, Bangalore, India, January 2010.
- [3] Knocking. <http://knockinglive.com>, 2010.
- [4] R. Chen. Constrained stochastic control with probabilistic criteria and search optimization. In *Proc. 43rd IEEE Conference on Decision and Control*, December 2004.
- [5] Y. P. Zhou, D. M. Chiu, and J. C. S. Lui. A simple model for analyzing P2P streaming protocols. In *Proc. IEEE ICNP 2007*.
- [6] 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.
- [7] 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.
- [8] S. Acedanski, S. Deb, M. Médard, and R. Koetter. How good is random linear coding based distributed networked storage. In *NetCod*, 2005.
- [9] C. Gkantsidis, J. Miller, and P. Rodriguez. Comprehensive view of a live network coding p2p system. In *Proc. ACM SIGCOMM*, 2006.
- [10] M. Wang and B. Li. R<sup>2</sup>: 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.
- [11] H. Chi and Q. Zhang. Deadline-aware network coding for video on demand service over p2p networks. In *PacketVideo*, 2006.
- [12] D. Kumar, E. Altman, and J-M. Kelif. Globally Optimal User-Network Association in an 802.11 WLAN and 3G UMTS Hybrid Cell. In *Proc. of the 20th International Teletraffic Congress (ITC-20)*, Ottawa, Canada, June 2007.
- [13] C. Feng and B. Li. On large-scale peer-to-peer streaming systems with network coding. In *Proceedings of the 16th ACM international conference on Multimedia*, Vancouver, Canada, October 2008.
- [14] 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].
- [15] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, 1997.