Operations Research Letters
Optimal admission control for tandem loss systems with two stations

Daniel F. Silva\textsuperscript{a,}\textsuperscript{*}, Bo Zhang\textsuperscript{b}, Hayriye Ayhan\textsuperscript{a}

\textsuperscript{a} H. Milton Stewart School of Systems and Industrial Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, United States
\textsuperscript{b} IBM T. J. Watson Research Center, 1101 Kitchawan Road, Route 134, Yorktown Heights, NY 10598, United States

\textbf{ARTICLE INFO}

\textbf{Article history:}
Received 10 December 2012
Received in revised form 5 February 2013
Accepted 1 April 2013
Available online 10 April 2013

\textbf{Keywords:}
Admission control
Loss systems
Tandem queues
Markov decision processes
Stochastic dynamic programming

\textbf{ABSTRACT}

We study a system of two queues in tandem with finite buffers, Poisson arrivals to the first station, and exponentially distributed service times at both stations. Losses are incurred either when a customer is rejected at the time of arrival to the first station or when the second station is full at the time of service completion at the first station. The objective is to determine the optimal admission control policy that minimizes the long-run average cost.

\textsuperscript{*} Corresponding author. Tel.: +1 4049914840.
E-mail address: dfsi3@gatech.edu (D.F. Silva).

1. Introduction

We consider a system with two stations in tandem and one server at each station. Arrivals to the system follow a Poisson process with rate $\lambda$, and service times at each station follow an exponential distribution with rates $\mu_i$, $i = 1, 2$. This note will mainly focus on the case that the first station has a buffer capacity of one, while the second station has an arbitrary finite buffer capacity. However, we also have results for systems with arbitrary finite buffers at both stations. Upon each arrival a gatekeeper has to decide (based on full knowledge of the state of the system) whether to admit or reject the incoming arrival. If an arrival is not admitted, a cost $c_1$ is incurred. If a customer completes service at the first station and at that time the second station is full, the customer is lost and a penalty cost of $c_2$ is incurred. Note that if the first station is full at the time of an arrival, then the incoming customer has to be rejected and the cost of $c_1$ is incurred. Our objective is to determine an admission policy at the first station that minimizes the long-run average cost.

Communication systems such as the Internet can be modeled as queuing networks with loss; see for example Bertsekas and Gallager [1] or Spicer and Ziedins [12] and the references therein for a description of these loss systems, and where they arise. To the best of our knowledge, the first paper to provide exact analytical results for a long-run average cost optimal policy for tandem queues with loss was Zhang and Ayhan [13] which considered a system with a finite buffer at the first station and a unitary buffer at the second station. They find that the optimal policy is of threshold type on the cost $c_2$ and provide a closed form expression for this threshold (which does not depend on the buffer size). Several related control problems in tandem queues can be found in the literature. The most relevant to this research are Ku and Jordan [5,7,6], Sheu and Ziedins [11], Ghoneim and Stidham [3], Hordijk and Koole [4], Chang and Chen [2] and Leskelä and Resing [8]. In the interest of space, we refer the reader to Zhang and Ayhan [13] for details on these references.

The remainder of this paper is organized as follows. In Section 2, we provide a Markov Decision Process formulation of the problem and provide some properties of the optimal policy for the general problem. We then focus on the more specific case of having a unitary buffer at the first station and completely characterize the optimal policy for this case. However, since the exact implementation of the optimal policy can be difficult for larger buffer sizes, in Section 3 we present numerical results that show that a heuristic policy has near optimal long-run average cost performance for larger systems.

2. Characterization of the optimal policy

In this section, we begin by modeling the system as a Markov Decision Process and proceed to prove the intuitive result that if $c_1 \geq c_2$, then the optimal policy always accepts incoming customers unless the first buffer is full at the time of arrival. Afterwards, we focus on the specific problem of unitary buffer at the
first station and show that there are only two policies that could be optimal depending on a threshold of $c_2$. We provide closed form expressions for the threshold when the buffer size at the second station is 10 or less. Finally, we investigate the effects of choosing the wrong policy.

2.1 Preliminaries

We use uniformization (see Lippman [9]) similarly to Zhang and Ayhan [13] to formulate the admission control problem as a discrete-time Markov decision process (MDP). Let $B_i$, $i = 1, 2$ be the buffer sizes at each station. Without loss of generality, we assume $\mu_1 + \mu_2 + \lambda = 1$. We then have a discrete-time Markov chain with state space \( S = \{q_1, q_2\} \in \mathbb{Z}^2 : 0 \leq q_1 \leq B_1, 0 \leq q_2 \leq B_2 \), where $q_i$ is the number of customers at station $i = 1, 2$.

Let 0 denote rejecting the next potential arrival and 1 denote accepting the next potential arrival. Then the sets of allowable actions are $A_0(q_2) = \{0\}$ if $q_1 = B_1$, and $A_0(q_2) = \{0, 1\}$, for any other $(q_1, q_2)$. Let $p(c_1(q_1, q_2), d) \in S$. Let $p(c_1(q_1, q_2), d)$ denote the transition probability when action $d$ is taken in state $(q_1, q_2)$. Let $\pi_1 = 1$ if $A$ is true, and 0 otherwise. We have

\[
p(p(s)(q_1, q_2), d) = \begin{cases} 
\mu_1, s = ((q_1 - 1)^+, q_2 + 1_{(q_1 > 0, q_2 < B_2)}) \\
\mu_2, s = (q_1, q_2 - 1^+) \\
\lambda, s = (q_1 + 1_{(d = 1)}, q_2),
\end{cases}
\]

with $p((0, 0))(0, 0), 0) = 1$. Let $r(s, d)$ denote the expected reward received when action $d$ is taken in state $s$, then

\[
r((q_1, q_2), d) = -c_1 \lambda I(\delta = 0) - c_2 \mu_1 I((q_1 > 0 \land q_2 < B_2)).
\]

This model is unique since regardless of the policy and the initial state, state $(0, 0)$ can be reached with probability 1, due to a long enough interarrival time. Therefore, $(0, 0)$ is recurrent, and all the states accessible from $(0, 0)$, together with $(0, 0)$, form a recurrent class; other states form a set of transient states. Hence, the existence of a deterministic stationary optimal policy is guaranteed by the finiteness of the state space and the action space (see Theorem 9.1.8 in Puterman [10]). Therefore, throughout the remainder of this paper, a policy $\pi$ is identified with a binary-valued function $\pi : S \rightarrow \{0, 1\}$.

Let $S_q = \{q_1, q_2\} \in S : 0 \leq q_1 \leq B_1 - 1\}$ be the set of states at which a choice of acceptance or rejection needs to be made. Each possible combination of actions to be taken at each state in $S_q$ comprises a stationary deterministic policy $\pi$.

First, we analyze the corresponding finite-horizon model under the expected total reward criterion. Let $v_{n,s}(s)$ be the expected total reward over $n$ periods under policy $\pi$ if the initial state is $s$ and $v_n(s)$ be the optimal $n$-period reward with the initial state $s$, or $v_n(s) = \inf_{s \in S} v_{n,s}(s)$.

Let $(a)^+ = \max(a, 0)$. Then the optimality equation $\forall (q_1, q_2) \in S$ is as follows:

\[
v_{n}(q_1, q_2) = \mu_1 [\lambda (q_1 - 1)^+, q_2 + 1_{(q_1 > 0 \land q_2 < B_2)}] + v_{n-1}(q_1 - 1)^+ + \lambda \max(-c_1 + v_{n-1}(q_1, q_2), v_{n-1}(q_1 + 1, q_2)).
\]

The optimal action at $(q_1, q_2) \in S_q$ with $n$ remaining periods is

\[
d_n(q_1, q_2) = \begin{cases} 
0, & \text{if } v_{n-1}(q_1 + 1, q_2) > -c_1 + v_{n-1}(q_1, q_2) \\
1, & \text{if } v_{n-1}(q_1 + 1, q_2) \leq -c_1 + v_{n-1}(q_1, q_2). \quad (1)
\end{cases}
\]

The following proposition is intuitive and will be useful in our analysis.

**Proposition 1.** $v_n(q_1 + 1, q_2) \geq v_n(q_1, q_2) - \max\{c_1, c_2\}, \forall q_1 < B_1, q_2 \leq B_2, \forall n \in \mathbb{Z}^+.$

**Proof.** Consider two systems with the same number of periods remaining in the decision horizon: system I under any policy $\pi$ with the initial state $(q_1, q_2)$, and system II with the initial state $(q_1 + 1, q_2)$ for any $q_1 < B_1$. Let system II operate under the same policy $\pi$ as if the initial state were $(q_1, q_2)$. Mark the last customer at station I in system II and always hold the marked customer as late as possible; specifically, serve the marked customer only if there is no one else at his station, and if there is an arrival to his station during his service, his service is preempted (in this case, his remaining service time, when he resumes service, equals in distribution to his original service time by the exponential assumption). Also, if a loss occurs at the marked customer’s station before he leaves the system, he is immediately unmarked and treated the same as all other customers afterwards.

With the two systems operating the aforementioned way, there are only two possibilities: (1) the marked customer leaves system II before any difference arises in the occurrence of losses in these two systems; after that, both systems evolve identically and thus the same total cost is incurred in both systems by the end of the decision horizon. (2) One more loss occurs in system II than in system I at one of the two stations (where the marked customer is located at that moment), incurring an extra cost $c_1$ or $c_2$, and after that both systems evolve identically. In other words, by initially holding one more customer at station I, system II either has the same total cost, or $c_1$ or $c_2$ more than system I. Because this is true for any policy $\pi$, the proposition then follows. □

2.2. Optimal policy

We will use a similar convention as proposed in Zhang and Ayhan [13] for naming a particular policy. Specifically, consider a policy where the gatekeeper admits an arrival whenever possible, we call this policy $\pi_c$, or greedy policy. Now consider the case where $c_1 \geq c_2$. We can use Proposition 1 to obtain the following.

**Theorem 1.** If $c_1 \geq c_2$, the greedy policy $\pi_c$ is optimal for the $n$-period problem under the expected total cost criterion, $\forall n \in \mathbb{Z}^+$.

**Proof.** It follows from $c_1 \geq c_2$ and Proposition 1 that $v_n(q_1 + 1, q_2) \geq v_n(q_1, q_2) - c_1$.

This, together with Eq. (1), then implies the optimality of the greedy policy. □

**Corollary 1.** If $c_1 \geq c_2$, the greedy policy $\pi_c$ is long-run average reward optimal.

For the rest of the paper we focus on the case where $B_1 = 1$ and $B_2 = B < \infty$. Consider a policy where the gatekeeper admits an arrival whenever the current state $(q_1, q_2)$ satisfies $q_1 = 0$ and $q_2 < B$. We call this policy $\pi_B$ or prudent policy throughout. Since the capacity of the first station is now 1, it is intuitive that a deterministic policy that rejects at $(0, q_2)$ for some $q_2 < B$ cannot be optimal. This is expected as any arrival admitted into the system at such states cannot be lost, so by rejecting at these states the system would incur a cost of $c_1$ when there is zero probability of incurring $c_2$. In the next proposition we formalize this idea.

**Proposition 2.** If a given policy $\pi$ it is always optimal to accept at $(0, q_2)$ for any $q_2 < B$. Or conversely a deterministic policy that rejects at $(0, q_2)$ for some $q_2 < B$ cannot be optimal.

**Proof.** We prove by induction. The result holds for $q_2 = 0$ since rejecting at $(0, 0)$ yields the long-run average cost $\lambda c_1$ per time unit; accepting at $(0, 0)$ while rejecting at all other states yields $\lambda c_1 \delta$, where $\delta$, the long-run fraction of time that the resulting system is not in state $(0, 0)$, must be strictly less than 1. Now suppose the desired result holds for $q_2 = 0, 1, \ldots, j - 1$, where $j < B$. Then we show it is also true for $q_2 = j$. 
Consider two systems: system I under any policy $\pi$, under which the rejection action is taken at state $(0, j)$, and system II also under $\pi$ except that the acceptance action is taken at state $(0, j)$. Both systems start with the same initial state. Because, for system I, all states $(q_1, q_2)$ with $q_1 + q_2 \geq j + 1$ are transient, we can assume that the policy $\pi$ prescribes the rejection action for all these states. Also, by the induction hypothesis, it suffices for us to assume that the policy $\pi$ prescribes the acceptance action for $(0, q_2)$, where $q_2 = 0, 1, \ldots, j - 1$. Note that both assumptions are effective for system II as well, because it also operates under $\pi$ (except at state $(0, j)$).

Consider any sample path. If state $(0, j)$ is never seen by an arrival on this sample path, both systems evolve identically. We now show that, if $(0, j)$ is ever seen by an arrival at some point in time, system II will have no more loss cost than system I from that point, say $T_1$, to the next time point when both systems reach the same state, say $T_2$.

First, we note that no loss ever occurs at station 2 in either system because $j < B$ and there are never more than $j + 1$ customers in either system. Second, at any point in time between $T_1$ and $T_2$ the states of the two systems at a customer arrival epoch must have one of the following forms.

- System I at $(0, j)$ and system II at $(1, j)$. The arrival is rejected in both systems in this case.
- System I at $(0, i)$ and system II at $(0, i + 1)$, where $i < j$. In this case, the arrival is accepted in both systems by the induction hypothesis and, also, for $i = j - 1$, using the assumption that the acceptance action is taken at state $(0, j)$ in system II.
- System I at $(0, j)$ and system II at $(0, j + 1)$. In this case, the arrival is rejected in both systems.
- System I at $(0, i)$ and system II at $(1, i)$, for some $i < j$. The arrival is accepted in system I by induction hypothesis but rejected in system II. Between $T_1$ and $T_2$ this can only occur once, after which both systems immediately reach the same state.

Therefore, in the interval $(T_1, T_2)$, at most $c_1$ more cost is incurred in system II than in system I. Also, because, at time $T_1$, system I has a rejection/loss cost $c_1$ while system II does not have any cost, we conclude that system II has no more loss cost than system I from $T_1$ to $T_2$. Because $\pi$ is arbitrary, it is optimal to accept at $(0, j)$.

So far, we have shown that for a given value of $B$ and a known set of parameters either the prudent policy must be optimal or the greedy policy must be optimal. Also, that if $c_1 > c_2$, then $\pi_G$ is always optimal. We leave the question of determining which of the two policies is optimal when $c_1 < c_2$. In order to answer the question we compute the long-run average cost under both policies and compare the cost values. Define $C_G(B)$ and $C_P(B)$ as the long-run average cost of operating a system with buffer $B$ under policies $\pi_G$ and $\pi_P$, respectively.

Let $(q_1, q_2) \in S$ be the stationary distribution of the Continuous Time Markov Chain (CTMC) model of this system operating under the prudent policy, that is $S = \{(q_1, q_2) \in \mathbb{Z}^2 : 0 \leq q_1 \leq 1, 0 \leq q_2 \leq B\}$ and $\hat{P}_{(0,1)}(q_1, q_2) \in S$ be the stationary distribution of the resulting CTMC when operating under the greedy policy $\pi_G$. Define

$$C_G(B) = c_G \left( \frac{1}{\mu_G} \left( \sum_{i=0}^{B-1} (P_{(1,0)} - \hat{P}_{(1,0)}) + P_{(0,B)} \right) - 1 \right).$$

**Proposition 3.** If for a fixed buffer size $B$, $c_G > c_P(B)$, then $C_G(B) < C_P(B)$ and $\pi_G$ is long-run average cost optimal for this system; otherwise $\pi_G$ is optimal.

**Proof.** Note that $P_{(1,0)} = 0$. Then, for fixed $B$:

$$C_G(B) = c_G \left( \sum_{i=0}^{B-1} (P_{(1,0)} - \hat{P}_{(1,0)}) + P_{(0,B)} \right).$$

Similarly, for the greedy policy:

$$C_P(B) = c_P \left( \sum_{i=0}^{B-1} (P_{(1,0)} - \hat{P}_{(1,0)}) + P_{(0,B)} \right).$$

The result then immediately follows from **Proposition 2** and the comparison of $C_G(B)$ with $C_P(B)$. □

Next we provide closed form expressions for $c_G(B)$ when $B \leq 10$ (since the computation of $c_G(B)$ for $B > 10$ is difficult and tedious).

Following the procedure described above, we obtain expressions of the form

$$c_P(B) = c_1 \left( 1 + \frac{\alpha(B)}{\beta(B)} \right).$$

We have a closed form expression for $\alpha(B)$, when $B \leq 10$, which is

$$\alpha(B) = \mu_2^{B+1} \sum_{i=1}^{[B/2]} \left( \lambda^{i-1} + 1 \right) \mu_1^{i-1} \left( \mu_1 + \mu_2 \right)^{B-2i+1} \left( \mu_1 + \mu_2 \right)^{B-i}.$$  

for $B$ we get an expression for each $B$

$$\beta(1) = \lambda \mu_1 + \mu_2 + \mu_3,$$

$$\beta(2) = \mu_1 \mu_2 (\mu_1 + \mu_2) + \lambda \mu_2 (\mu_1 + \mu_2)^2 + \lambda \mu_3 (\mu_1 + \mu_2)^2 + \lambda \mu_4 (\mu_1 + \mu_2)^2 + \mu_5 (\mu_1 + \mu_2)^2,$$

$$\beta(3) = \mu_1 \mu_2 (\mu_1 + \mu_2)^2 + \lambda \mu_3 (\mu_1 + \mu_2)^3 + \lambda \mu_4 (\mu_1 + \mu_2)^3 + \lambda \mu_5 (\mu_1 + \mu_2)^3 + \mu_6 (\mu_1 + \mu_2)^3,$$

$$\beta(4) = \mu_1 \mu_2 \mu_3 (\mu_1 + \mu_2)^3 + \lambda \mu_4 (\mu_1 + \mu_2)^4 + \lambda \mu_5 (\mu_1 + \mu_2)^4 + \lambda \mu_6 (\mu_1 + \mu_2)^4 + \mu_7 (\mu_1 + \mu_2)^4,$$

$$\beta(5) = \mu_1 \mu_2 \mu_3 (\mu_1 + \mu_2)^2 + \lambda \mu_4 (\mu_1 + \mu_2)^2 + \mu_5 (\mu_1 + \mu_2)^2,$$

$$\beta(6) = \mu_1 \mu_2 \mu_3 \mu_4 (\mu_1 + \mu_2)^3 + \lambda \mu_5 (\mu_1 + \mu_2)^3 + \lambda \mu_6 (\mu_1 + \mu_2)^3 + \mu_7 (\mu_1 + \mu_2)^3,$$

$$\beta(7) = \mu_1 \mu_2 \mu_3 \mu_4 \mu_5 (\mu_1 + \mu_2)^4 + \lambda \mu_6 (\mu_1 + \mu_2)^4 + \mu_7 (\mu_1 + \mu_2)^4,$$

$$\beta(8) = \mu_1 \mu_2 \mu_3 \mu_4 \mu_5 \mu_6 (\mu_1 + \mu_2)^5 + \lambda \mu_7 (\mu_1 + \mu_2)^5.$$

$\lambda = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 + \lambda_6 + \lambda_7 + \lambda_8.$
\[ \beta(7) = \mu_1 \mu_2 (\mu_1 + \mu_2)^4 + \lambda \mu_6 (\mu_1 + \mu_2)^3 \times (\mu_1^4 + 3 \mu_1^2 \mu_2 + 12 \mu_1 \mu_2^2 + 26 \mu_1^2 \mu_2^2 + 29 \mu_1 \mu_2^4 + 6 \mu_2^6) \times (\mu_1^3 + 5 \mu_1^2 \mu_2 + 15 \mu_1 \mu_2^2 + 35 \mu_1^2 \mu_2^2 + 68 \mu_1^3 \mu_2^4 + 93 \mu_1^2 \mu_2^6 + 65 \mu_1 \mu_2^8 + 15 \mu_2^{10}) + \lambda_4 \mu_1 \mu_2^2 (\mu_1 + \mu_2)^2 + 10 \mu_1 \mu_2^2 + 25 \mu_1 \mu_2^4 + 5 \mu_1 \mu_2^6 + 6 \mu_1^2 \mu_2^8 + 21 \mu_1^3 \mu_2^{10} + 26 \mu_1^4 \mu_2^{12} + 15 \mu_1^5 \mu_2^{14} + \lambda_5 \mu_1 \mu_2^2 (\mu_1 + \mu_2)^2 + 3 \mu_1 \mu_2^2 + 4 \mu_1^2 \mu_2^4 + 5 \mu_1 \mu_2^6 + 6 \mu_1^2 \mu_2^8 + 7 \mu_1^3 \mu_2^{10} + 6 \mu_1^4 \mu_2^{12} + \lambda_7 \mu_1 \mu_2^2 (\mu_1 + \mu_2)^2 + 3 \mu_1 \mu_2^2 + \mu_1^2 \mu_2^4 + \mu_1 \mu_2^6 + \mu_2^8) \]

\[ \beta(8) = \mu_1 \mu_2 (\mu_1 + \mu_2)^3 + \lambda_2 \mu_1 \mu_2 (\mu_1 + \mu_2)^3 \times (\mu_1 + 3 \mu_1 \mu_2 + 9 \mu_1^2 \mu_2 + \lambda_3 \mu_1 (\mu_1 + \mu_2)^3) \times (\mu_1 + 4 \mu_1 \mu_2 + 13 \mu_1^2 \mu_2 + 30 \mu_1^3 \mu_2 + 37 \mu_1^2 \mu_2^2 + 7 \mu_2^4) + \lambda_5 \mu_1 \mu_2 (\mu_1 + \mu_2)^3 + 21 \mu_1^3 \mu_2^2 + 221 \mu_1^4 \mu_2^4 + 116 \mu_1^5 \mu_2^6 + 21 \mu_1^6 \mu_2^8 + \lambda_5 \mu_1 \mu_2 (\mu_1 + \mu_2)^3 + 4 \mu_1 \mu_2^2 + 5 \mu_1^2 \mu_2^4 + 5 \mu_1 \mu_2^6 + 6 \mu_1^2 \mu_2^8 + 7 \mu_1^3 \mu_2^{10} + 6 \mu_1^4 \mu_2^{12} + \lambda_7 \mu_1 \mu_2 (\mu_1 + \mu_2)^3 + 3 \mu_1 \mu_2^2 + \mu_1^2 \mu_2^4 + \mu_1 \mu_2^6 + \mu_2^8) \]

\[ \beta(9) = \mu_1 \mu_2 (\mu_1 + \mu_2)^2 + \lambda_3 \mu_1 \mu_2 (\mu_1 + \mu_2)^2 \times (\mu_1^4 + 3 \mu_1^2 \mu_2 + \lambda_2 \mu_1 \mu_2 (\mu_1 + \mu_2)^2 + 10 \mu_1 \mu_2^2 + \lambda_5 \mu_1 \mu_2 (\mu_1 + \mu_2)^2 + 4 \mu_1^4 \mu_2^4 + 14 \mu_1^3 \mu_2^4 + 34 \mu_1^2 \mu_2^8 + 6 \mu_1^3 \mu_2^4 + 21 \mu_1^4 \mu_2^8 + 28 \mu_1^5 \mu_2^{12} + 34 \mu_1^4 \mu_2^{12} + 21 \mu_1^5 \mu_2^{12} + \lambda_7 \mu_1 \mu_2 (\mu_1 + \mu_2)^2 + 5 \mu_1 \mu_2^2 + 7 \mu_1^2 \mu_2^4 + 8 \mu_1 \mu_2^6 + 7 \mu_1^2 \mu_2^8 + \lambda_8 \mu_1 \mu_2 (\mu_1 + \mu_2)^2 + \mu_1^2 \mu_2^4 + \mu_1 \mu_2^6 + \mu_2^8) \]

**Proposition 4.** Under \( \sigma_{\tau_c} \), the stationary probability of being in state (0, B). \( p(0, B) \) is strictly decreasing in \( B \) for \( 0 \leq B \leq 10 \).

Now consider the greedy policy.

**Proposition 5.** Under \( \sigma_{\tau_c} \), the stationary probability of being in state (0, B). \( \hat{p}(0, B) \) is strictly decreasing in \( B \) for \( 0 \leq B \leq 10 \).

Finally, consider state (1, B) under the greedy policy, where cost \( c_2 \) may be incurred.

**Proposition 6.** Under \( \sigma_{c_2} \), the stationary probability of being in state (1, B). \( \hat{p}(1, B) \) is strictly decreasing in \( B \) for \( 0 \leq B \leq 10 \).
The above propositions lead us to propose the following conjecture.

**Conjecture 1.** Under $\pi_P$ as $B \to \infty$, $p_{(0,B)}$ converges monotonically to some value $p^*_{(0,B)} \geq 0$. Under $\pi_G$ as $B \to \infty$ the probabilities $\hat{p}_{(0,B)}$ and $\hat{p}_{(1,B)}$ converge monotonically to values $\hat{p}^*_{(0,B)} \geq 0$ and $\hat{p}^*_{(1,B)} \geq 0$, respectively.

As there is no closed form expression for the stationary probabilities for general values of $B$, we provide numerical evidence to support our conjecture. We conduct numerical experiments considering a system where $B$ attains all integer values from 1 to 50, and also multiples of 50 up to 500. We generate 1000 sets of parameters independently and solve the problem numerically for each set with each value of the buffer size. In order to generate $\lambda$, $\mu_1$, and $\mu_2$, we sample from a continuous uniform distribution between 1 and 100, while $c_1$, $c_2$ are sampled from a continuous uniform distribution between 1 and 1000. If $c_1 \geq c_2$, we discard those cost values and generate a new set until $c_1 < c_2$, since we know (from Theorem 1) that if $c_1 \geq c_2$ then $\pi_G$ is optimal. In order to analyze the results we divided the datasets into six cases as explained in Table 1.

**Table 1.** Cases studied.

<table>
<thead>
<tr>
<th>Case</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$\lambda &lt; \mu_1 &lt; \mu_2$</td>
</tr>
<tr>
<td>B</td>
<td>$\lambda &lt; \mu_2 &lt; \mu_1$</td>
</tr>
<tr>
<td>C</td>
<td>$\mu_1 &lt; \lambda &lt; \mu_2$</td>
</tr>
<tr>
<td>D</td>
<td>$\mu_1 &lt; \mu_2 &lt; \lambda$</td>
</tr>
<tr>
<td>E</td>
<td>$\mu_2 &lt; \lambda &lt; \mu_1$</td>
</tr>
<tr>
<td>F</td>
<td>$\mu_2 &lt; \mu_1 &lt; \lambda$</td>
</tr>
</tbody>
</table>

**Fig. 1.** Average $p_{(0,B)}$ for increasing values of $B$ for cases A–E.

**Fig. 2.** Average $\hat{p}_{(0,B)}$ for increasing values of $B$ for cases A–E.

**Fig. 1** illustrates how $p_{(0,B)}$ changes with respect to $B$ for each of the cases above. **Fig. 2** shows $\hat{p}_{(0,B)}$ for different values of $B$. The results for $\hat{p}_{(1,B)}$ follow a similar pattern to those shown in **Fig. 2** and were omitted in the interest of space. All probability values in these figures were calculated as averages for a fixed buffer size over the experiments described above. Our observations show that the probability of being in the states of interest decreases dramatically as a function of the buffer size until it practically vanishes by $B = 10$ for cases A and C and by $B = 50$ for cases B and D but never vanishes for cases E and F (i.e., when $\mu_2$ is smaller than $\mu_1$ and $\lambda$). For cases E and F, as $B$ increases, all probabilities of interest monotonically decrease to a strictly positive value.
These results suggest that choosing a non-optimal policy (amongst the prudent and greedy policies) may not have a significant impact on the cost when parameters satisfy cases A–D, but might have an impact when they satisfy cases E and F. In particular, we can conjecture that \( c^*_{\leq 10} \) can be a good approximation for \( c^*(B) \) for \( B \geq 10 \) in cases A–D; however, the same may not hold in cases E and F. Thus, it is important to develop heuristic policies that yield good cost performance for larger values of B under these scenarios as the risk remains that we will incur unnecessary costs by choosing the wrong policy.

3. Heuristic and numerical experiments

As we have shown, the optimal policy is characterized by a threshold \( c^*(B) \). We provided closed form expressions for \( c^*(B) \) for \( B \leq 10 \). However, it is difficult to obtain a closed form expression for \( c^*(B) \) for \( B > 10 \). In the previous section we saw that for large values of B, the probabilities of being in a full state vanish quickly as B increases for all systems where \( \mu_2 > \lambda \) and/or \( \mu_2 > \mu_1 \) but they are not negligible otherwise (namely cases E and F). Our numerical results below show that for a wide range of parameters \( c^*(10) \) can be used whenever \( B > 10 \) and still obtain near optimal results.

For the datasets described in Section 2.3, we calculated both the optimal long-run average cost and the long-run average cost under a heuristic policy, which uses the optimal threshold if \( B \leq 10 \) and uses \( c^*(10) \) otherwise. The average percentage of additional cost incurred over the optimal cost for the heuristic policy was less than 0.01% for all six cases A–F. The worst case performance of the heuristic policy was also close to the optimal policy, with worst case additional cost incurred over the optimal cost of 0.08% for case E and 0.02% for case F, and less than 0.001% for cases A–D. Thus, the heuristic policy has near optimal performance in all cases.

Acknowledgment

This work was supported in part by the National Science Foundation under grants CMMI-0856600 and CMMI-0969747.

References