Average Optimal Policies in a Controlled Queueing System with Dual Admission Control *†

Mark E. Lewis
Department of Industrial and Operations Engineering
University of Michigan, 1205 Beal Avenue, Ann Arbor, MI 48109-2117
melewis@engin.umich.edu
734-763-0519

submitted July 17, 2000, revised December 25, 2000, accepted March 1, 2001

Abstract

We consider a controlled $M/M/1$ queueing system where customers may be subject to two potential rejections. The first occurs upon arrival and is dependent on the number of customers in the queue and the service rate of the customer **currently** in service. The second which may or may not occur, occurs immediately prior to the customer receiving service. That is, after each service completion the customer in the front of the queue is assessed and the service rate of that customer is revealed. If the second decision-maker recommends rejection, the customer is denied service with a fixed probability. We show the existence of long-run average optimal monotone switching curve policies. Further, we show that the average reward is increasing in the probability that the second decision-maker's recommendation of rejection is honored. Applications include call centers with delayed classifications and manufacturing systems when the server is responsible for multiple tasks.

1 Introduction

Traditional multi-class admission control models assume that upon the arrival of a customer, the controller immediately knows the class of the customer. However, not only may the assessment of customers be time-consuming, but since the customer may wait in queue after admission, this assessment may no longer be valid by the time the customer begins service. Furthermore, as the queue length grows, the amount of information that must be stored to control the system becomes burdensome. In each scenario, it would seem more advantageous for the decision-maker to do the assessment and make the decision to serve the customer at a later time, say immediately prior to the customer beginning service. The obvious problem with this formulation is that the system is charged with holding customers regardless of whether it would

* *AMS Subject Classifications:* **primary-90C40:** Markov and semi-Markov decision processes, **secondary-60K25** queueing theory
† *IAOR Subject Classifications:* **primary-3160:** Markov processes, **secondary-3390:** queues: theory

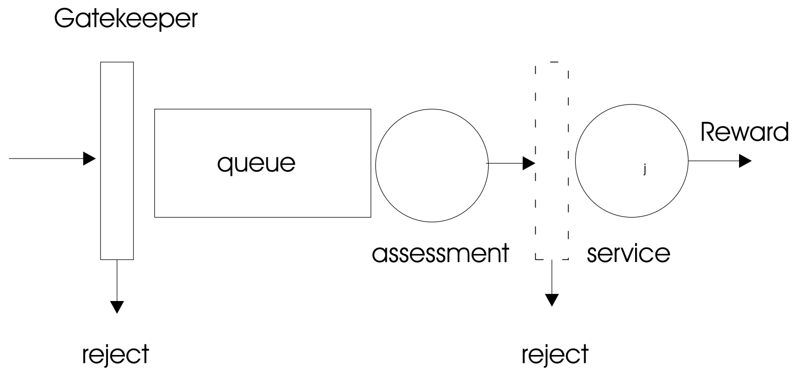


Figure 1: A dual admission controlled queueing system. The dashed box indicates that the server may recommend a rejection that occurs with probability q.

have been more cost effective to reject them upon arrival. A remedy to either problem; congestion control upon arrival or immediately before service, is to have the customers meet two decision-makers. The first provides congestion control when customers arrive to the system as is typical in admission control models. The second occurs after each service completion when the server assesses the customer in the front of the queue, the service requirements of the customer are revealed and the server may decide it is unwise to serve the customer. See Figure 1.

Suppose now that the server's recommendation to reject a customer may not always be honored. That is to say that with some probability the decision is ignored and the server must serve the customer. Given our previous discussion it seems intuitive that each decision-maker would allow admittance to the queue and to service until such time that the number of customers reaches some level or *control limit*, after which admittance would not be allowed. Since the service requirements of the customer in service have been revealed, the limits should depend on this information as well. Such a policy is called a *switching curve policy*. Furthermore, since the second decision-maker adds flexibility to the process, the higher the probability this decision is honored, the higher the long-run average reward should be.

As further motivation for this study, we assert that this model serves as a baseline for several models in practice. Consider for example a manufacturing setting where one machine must perform multiple tasks. The gatekeeper decides which jobs will be accepted and prepared for service as job requests arise. As service will not commence immediately, and assessing each part, and keeping this information may not be feasible, the gatekeeper does not know the service rate of the job when it arrives. The assessment of the job will be done immediately prior to service. The server then can decide if the job should be processed or rerouted somewhere else in the system.

Similarly, in telecommunications, call centers receive a call and often do not immediately classify the call's importance. The gatekeeper corresponds to traditional congestion control and relates to the call originally getting through. After sitting in queue for some time, perhaps while some information is gathered about the customer and the present state of the system, a classification is made. The server may then decide if the caller should be served or asked to call at a later time.

We add the further generality that the server's recommendation may be "overruled" in which case it would be forced to serve the customer. By allowing the probability that a rejection by the server is honored to vary from zero to one, we may evaluate the value of having a second decision-maker. The remainder of the paper is devoted to proving the following two results.

Theorem 1 *There exists a long-run average optimal policy that is a generalized monotone switching curve policy (see Definition 5).*

Theorem 2 *Let q be the probability that a rejection by the server is honored. The optimal gain, $g(q)$, is increasing in q.*

The paper is organized as follows: Section 2 is dedicated to a short literature review. We gather necessary definitions from average reward Markov decision process theory in Section 3. The model formulation can be found in Section 4. The proofs of the main results are in Section 5. We conclude with some examples and concluding remarks in Sections 6 and 7, respectively.

2 Literature Review

There is a vast literature on the control of queues and we will not give an exhaustive review here. Instead, we will highlight those papers of particular relevance to our current work. The idea of a *control limit policy* or a *trunk reservation policy* (see Definition 4) was originally reported by Yadin and Naor [18] for a server with vacations and proved to be optimal by Heyman [9]. This was later applied to a finite customer class, finite capacity, queueing system by Miller [14] and continued by Lippman and Ross [13] for a single-server with uncountable customer classes. Helm and Waldmann [7] showed in a very general setting if enough information is included in the state space the existence of control limit policies is still guaranteed. When the state space is more than one dimension, for example when there is more than one station, a control limit policy is called a *switching curve policy*. Hajek [6] discusses such a model with two stations where routing decisions must be made. A good survey of early papers on admission and service control in queueing systems was produced by Stidham [16].

More recently, optimal control limit admission policies for queues which operate in series or with delayed information were studied by Ghoneim and Stidham [5] and Altman and Stidham [4], respectively. Furthermore, Xu and Shanthikumar [17] discuss a queue with *expulsion control* akin to our second decision-maker. Lewis and Puterman [10, 11] showed that optimal policies may vary when a more sensitive optimality criterion, *bias optimality*, is used if the reward was received upon service completion as opposed to upon admittance into the system; the bias exhibits implicit discounting. None of the above mentioned papers are concerned with the interplay between **two** decision-makers making decisions about the same customers at varying points of the queueing process and with differing information.

For models with more than one decision-maker *stochastic game theory* has proved useful (cf. Altman [2]). In essence, most stochastic game models assume that two or more decision-makers are active at different times with different objective functions. A (Nash) equilibrium is sought where no player can benefit by deviating from the proposed policy. For a nice example of how this methodology has been applied to the control of queues see Altman [1]. There is also a survey by the same author that covers stochastic games and many related queueing problems in telecommunications systems (see [3]). In the present model, each decision-maker has the same objective. Hence, the second decision-maker adds the value of another opportunity to control congestion; there is no competition between the gatekeeper and the server.

3 Average Reward Markov Decision Processes

In this section we briefly discuss the components of a Markov decision process (MDP) and define the optimality criterion. Although the notation may be slightly different, all of the following ideas are outlined nicely in Puterman [15] or Hernandez-Lerma and Lasserre [8] . The problem we pose may be modeled as an infinite horizon, finite action, countable state space Markov decision process. However, we make an assumption on the holding cost so as to assure that an optimal policy need only consider a finite subset of this state space. Let the state space be denoted $\mathbb{X}$ and the action space $\mathbb{A}$. When in state $x \in \mathbb{X}$ the set of available actions will be written A_x. A deterministic decision rule d is a map from $\mathbb{X}$ to $\mathbb{A}$ such that when in state x the action $d(x)$ is used. A deterministic, Markovian policy π is a sequence of decision rules that describes what decisions will be made for every decision epoch. That is to say that π is of the form $\{d_1, d_2, \ldots\}$. We are interested in the class of stationary, deterministic policies which use the same decision rule for all decision epochs. A policy in this class is of the form $\{d, d, \ldots\}$ and is denoted d^∞. Let P_d be the one step transition matrix whose xy element is the probability of going to state y, given that the system is currently in state x and action $d(x)$ is employed. Similarly, let $r(x, a)$ be the expected reward when action a is chosen in state x. Suppose Π is the set of all non-anticipatory policies. We employ the convention that a function of a policy, π, or decision rule, d, and a state, x, may be written $f(\pi, x)$ or $f_\pi(x)$, similarly for d. We are now ready to define the gain of a policy π. Let X_k be the state of the system at stage k and d_k the decision rule at stage k under a particular policy π. The $k-$stage expected cost is

$$J_k(\pi, x) \equiv E_x^\pi \left[\sum_{n=0}^{k-1} r(X_n, d_n(X_n)) \right]. \tag{1}$$

Definition 3 *The* **long-run average reward** *or* **gain** *of a policy π given that the system started in state x, denoted $g(\pi, x)$ is given by*

$$g(\pi, x) \quad = \quad \liminf_{k \to \infty} J_k(\pi, x)/k. \tag{2}$$

Furthermore, let the **optimal expected average reward** *be denoted $g^*(x)$*

$$g^*(x) \quad = \quad \sup_{\pi \in \Pi} g(\pi, x). \tag{3}$$

A policy π^* is called **long-run average** *or* **gain optimal** *if $g(\pi, x) = g^*(x)$ for all $x \in \mathbb{X}$.*

We call an MDP *unichain* if all stationary policies generate a Markov chain with one recurrent class (and possibly some transient states). It is well-known that for the set of stationary, deterministic policies on a unichain, finite state and action space MDP, the limit in (2) exists and is independent of the initial state x (cf. Section 8.4 of Puterman [15]). Furthermore, restricting attention to this class of policies in fact is no restriction at all since the supremum over this smaller set is equivalent to that over all policies when the MDP is unichain, has bounded rewards, and finite state and action spaces (see Theorem 8.4.5 of Puterman).

The optimal gain of an MDP may be computed by solving the following system of linear equations (cf. Chapter 8 of Puterman [15]):

$$g = \max_{d \in D} P_d g \tag{4}$$

$$h = \max_{d \in D} \{r_d - g + P_d h\}. \tag{5}$$

for g and h. Specifically, the optimal gain satisfies (4) and there exists some vector h which together with the optimal gain satisfies (5). Moreover, the gain is the unique vector with this property. We refer to (4) and (5) as the *average optimality equations* (AOE). When the gain of all stationary policies is constant, i.e. when the MDP is unichain, (4) is superfluous and (5) can be used to find gain optimal policies. The process discussed is modeled as a Markov Decision Process with average reward optimality criterion in the next section.

4 Model Formulation

Suppose customers arrive to a single-server infinite capacity queueing system in accordance with a Poisson process of rate $\lambda < \infty$. Upon arrival customers meet a gatekeeper that must decide if the customer will be admitted to the system. The gatekeeper knows the service rate of the customer currently in service and the number of customers in the system. However, the class of the arriving customer (and those in the queue) is unknown to the gatekeeper. If accepted, customers immediately join the queue, if one exists, and await service. Assume when there are i customers in the system there is a holding cost $c(i)$ per unit time. We make the further assumptions that $c(i)$ is convex, increasing in i, $c(0) = 0$, and $c(1) > 0$. Immediately following a service completion, or when a customer arrives to an empty system, the server begins assessing the customer at the front of the queue. The time to complete the assessment is assumed exponential with rate α, $0 < \alpha < \infty$, and once the assessment is complete, the service rate of the customer is revealed.

After assessment, the customer is of Class j with probability p_j, $j = 1, \ldots, \ell$ such that $\sum_{j=1}^{\ell} p_j = 1$. The reward to be received (after service) if this customer is served, is then $R(j)$ and the service time is exponentially distributed with rate $\mu_j < \infty$. We assume both μ_j and $R(j)$ are strictly non-negative for $j \geq 1$. The service rates are assumed to be strictly increasing in j while the rewards are assumed to be non-decreasing. When the server is assessing a customer, a Class 0 customer is said to be in service. When the assessment is complete, the server may recommend rejection of the customer. However, this rejection will occur with probability $0 \leq q \leq 1$. If a customer's rejection is honored, a cost $c_r \geq 0$ is incurred and the customer leaves the system. On the other hand, if the server decides to serve the customer or the rejection is not honored the customer begins service at the prescribed rate. To avoid trivialities we assume that it is optimal to accept an arriving customer when the system is empty. In order to do so we assume that the expected reward of a customer arriving to an empty system is larger than the expected cost of that customer when it is the only customer to be accepted. That is,

$$\sum_{k=1}^{\ell} p_k R(k) > c(1) \left[\frac{1}{\alpha} + \sum_{k=1}^{\ell} \frac{p_k}{\mu_k} \right] + \min\{0, q \left[c_r - c(1) \sum_{k=1}^{\ell} \frac{p_k}{\mu_k} \right] \}. \tag{6}$$

Customers that are rejected by either the gatekeeper or the server (when the rejection is honored) are lost forever.

We consider a *uniformized*, discrete-time, embedded process. While it should be clear that we need only consider the state of the system when customers arrive and when an assessment has just been completed, for convenience we also embed at the service completions. Furthermore, uniformization introduces "dummy" transitions where the system state does not change, but time continues to progress. In doing so, policies under this discrete-time formulation have the same long-run average reward as the original system. For a complete discussion of this method, see Lippman [12]. Without loss of generality assume the uniformization constant $\Lambda \equiv \lambda + \mu_\ell + \alpha = 1$. Hence, in the uniformized chain when there are i customers in the system and a Class j customer is in service the arrival rate λ and the current service rate μ_j represent the *probability* that the next event is an arrival or departure, respectively. Similarly when there are i customers in the system and the server is assessing the first customer with μ_j replaced by α; the probability the next event is an assessment completion.

When the expected increase in the holding costs caused by accepting an arriving customer exceeds the expected reward offered by the customer it is optimal to reject that customer. When there are s customers in the system, a lower bound on the expected increase in cost to the system for the arriving customer is $c(s+1)$. To see this, suppose we start two processes on the same probability space. The first, process 1, starts with $s+1$ customers, while the second, process 2, starts with s customers. Assume further that the i^{th} customer for each process is of the same (unknown) class for $i \leq s$ and that the servers of both processes begin either serving a class j customer or assessing the customer in the front of the queue. Now assume that process 1 uses a fixed policy π while process 2 uses the same policy except that decisions are made as though it had the same number of customers as process 1. What all of this implies is that the processes follow the same sample paths except that process 1 continues to have one more customer in the system until such time that the queue of process 1 drains, after which the processes couple. The difference in cost between the two processes until the first transition is $c(s+1) - c(s)$. Similarly, the difference in costs until the second transition is at least $c(s) - c(s-1)$ (recall the convexity of c). Continue in this fashion for the first s transitions. Summing these differences yields $c(s+1)$ as claimed.

Since the holding cost is assumed convex, increasing and strictly positive when the system is non-empty, we have $c(i) \to \infty$ as $i \to \infty$. Let U be the smallest non-negative integer such that $c(U) > R(\ell)$. It must then be optimal to reject all arriving customers when there are U or more customers in the system. Note that when this policy is employed there may be i customers in the system for only a finite amount of time for any $i > U$ (transience). These states have no effect on the long-run average reward. Thus, we may truncate the state space at any level higher than U without affecting the long-run average reward of any reasonable policy. In order to start the inductive arguments that follow we assume that the number of customers remains below $U + 2$. This leads to the following Markov Decision Process formulation:

- The state space $\mathbb{X}$:

$$\mathbb{X} = [\{0, 1, \ldots, U+2\} \times \{0\}] \cup [\{1, 2, \ldots, U+2\} \times \{1, 2 \ldots, \ell\}]. \tag{7}$$

Note when $x = (i, j)$ there are i customers in the system and the current customer in service is of

Class j. Furthermore, a zero$^{\text{th}}$ class has been added to denote when there are i customers in the system and an assessment is in progress or when the system is empty.

- The action space is $\mathbb{A} \equiv \bigcup_{x \in \mathbb{X}} A_x$. We let 1 correspond to when the gatekeeper should accept an arriving customer while 2 corresponds to reject. Thus,

$$A_x = \begin{cases} \{1,2\} \times \{0\} & x \in \{\{1,2,\ldots,U-1\} \times \{1,2,\ldots,\ell\}\}, \\ \{1,2\} \times \mathcal{P}(\{1,2,\ldots\ell\}) & x \in \{\{0,1,\ldots,U-1\} \times \{0\}\}, \\ \{2\} \times \{0\} & x \in \{\{U,U+1,U+2\} \times \{1,\ldots,\ell\}\}, \\ \{2\} \times \mathcal{P}(\{1,2,\ldots\ell\}) & x \in \{\{U,U+1,U+2\} \times \{0\}\}, \end{cases} \tag{8}$$

where $\mathcal{P}(\{1,2,\ldots\ell\})$ denotes the power set of $\{1,2,\ldots\ell\}$, i.e., the set of all subsets of $\{1,2,\ldots\ell\}$. Thus, in state $x = (i,0)$ if the action $(1,B)$ has been chosen, the gatekeeper accepts arriving customers and the server recommends admittance to service only those customers whose class is in the set B. Similarly for $(2,B)$ and reject. Note that choosing to recommend service to Class 0 has been made available when a Class j customer ($j > 0$) is in service for convenience and consistency. Furthermore, if i^* is the smallest $i \leq U$ for which "reject" is optimal for all classes $j \geq 0$, all states (i,j) such that $i < i^*$ are recurrent while those with $i \geq i^*$ are transient. That is to say, there is one recurrent class; the MDP described above is unichain.

- The rewards $r(x,a)$ such that $x \in \mathbb{X}$ and $a \in \mathbb{A}_x$:

$$r((i,j),(b,B)) = \begin{cases} 0 & i = 0, \\ \mu_j R(j) - c(i) & 0 < i \leq U+2, j \neq 0, \\ -c(i) - qc_r \sum_{k \notin B} p_k & 0 < i \leq U+2, j = 0. \end{cases} \tag{9}$$

Next we state the AOE for the above model. We include in the notation the added dependence on q. For $1 \leq i \leq U+2$ and $j \geq 1$

$$\begin{aligned} h^q(i,j) &= \mu_j R(j) - c(i) - g(q) + \mu_j h^q(i-1,0) + [1-(\lambda+\mu_j)]h^q(i,j) \\ &\quad + \lambda \max\{h^q(i+1,j), h^q(i,j)\}, \end{aligned} \tag{10}$$

For $i \geq 1$ and $j = 0$

$$h^q(i,0) = -c(i) - g(q) + \alpha \left(\sum_{k=1}^{\ell} p_k [qV^q(i,k) + (1-q)h^q(i,k)] \right) + [1 - (\lambda + \alpha)]h^q(i,0)$$
$$+\lambda \max\{h^q(i+1,0), h^q(i,0)\}, \tag{11}$$

where

$$V^q(i,j) = \max\{h^q(i,j), h^q(i-1,0) - c_r\}. \tag{12}$$

Note that when $i \geq U$ we replace the "max" in (10) and (11) with $h^q(i,j)$ and $h^q(i,0)$, respectively, since rejecting customers is assumed to be optimal in these states. When the system is empty,

$$h^q(0,0) = -g(q) + (1-\lambda)h^q(0,0) + \lambda h^q(1,0). \tag{13}$$

In each state (i,j) for $j > 0$ the first (second) term in the maximum corresponds to accepting (rejecting) a customer when they arrive to the system. In the states of the form $(i,0)$ the first (second) term of $V^q(i,j)$ corresponds to serving (not serving) a customer after they have been assessed, while the terms in the second maximum in (11) are analogous to those when $j > 0$.

For a function $f(i,j)$ on the state space $\mathbb{X}$, let $\Delta_i f(i,j) \equiv f(i+1,j) - f(i,j)$. Thus, $f(i,j)$ is concave in i (for fixed j) if $\Delta_i^2 f(i,j) \leq 0$. Similarly, define $\Delta_j f(i,j) \equiv f(i,j+1) - f(i,j)$. From (10) and (11) the decision to accept or reject arriving customers is dependent on whether $\Delta_i h^q(i,\cdot)$ is positive or negative. Similarly, from (12) the decision to serve customers depends on $h^q(i,j) - (h^q(i-1,0) - c_r)$.

The final definitions of this section are the usual definition of a monotone switching curve (cf. Altman [2]) and a generalization of that definition.

Definition 4 *Suppose in each state $x \in \mathbb{X}$ there are but two actions, say action 1 and action 2. A stationary, deterministic policy is a* ***switching curve policy*** *if it may be described by a curve in $\mathbb{X}$ that separates $\mathbb{X}$ into two connected regions. In the first region the policy calls for action 1 to be used, while in the second region action 2 is used. Furthermore, a switching curve policy is called a* ***monotone switching curve policy*** *if the curve dividing $\mathbb{X}$ into two regions is monotone.*

Definition 5 *A stationary, deterministic policy d^∞ is called a* ***generalized monotone switching curve policy*** *with respect to the partial orderings $\prec_1$ on $\mathbb{X}$ and $\prec_2$ on $\mathbb{A}$, if whenever $y_1 \prec_1 y_2$, $d(y_1) \prec_2 d(y_2)$.*

Figure 2 shows an example of a monotone switching curve policy on a subset of the state space. It should be clear that on this subset the policy satisfies Definition 4. To see that it also satisfies Definition 5, let the partial ordering $\prec_{\mathbb{X}}$ be such that

$$(i,j) \prec_{\mathbb{X}} (m,n) \begin{cases} \text{if } j=0, i \leq m \text{ and } n=0, \\ \text{if } j \neq 0, i \leq m \text{ and } n \leq j. \end{cases} \tag{14}$$

We say that $(a, B_1) \prec_{\mathbb{A}} (b, B_2)$ if

1. $a \leq b$,
2. $j \in B_i$ implies $(j+1) \in B_i$ for $i = 1, 2$, and
3. $B_2 \subseteq B_1$.

Note that by (8) when $j \neq 0$, $B_1 = B_2 = \{0\}$; conditions 2 and 3 above hold. Furthermore, these states are not comparable to those with $j = 0$. Theorem 1 may now be restated in the following manner.

Alternative statement of Theorem 1: There exists a long-run average optimal policy d^∞ such that d is a generalized monotone switching curve policy with the partial orderings $\prec_{\mathbb{X}}$ on $\mathbb{X}$ and $\prec_{\mathbb{A}}$ on $\mathbb{A}$.

The monotone switching curve policy has the intuition that the gatekeeper is less willing to allow customers to enter the system when there are more customers in the system or if the server is working slower. Moreover, the lowest class recommended for service by the server increases as the number of customers increases. Given these observations in order to prove Theorem 1 we show that there exists $\{L_j\}_{j=0}^{\ell}$ and $\{S_j\}_{j=1}^{\ell}$ such that for $j \geq 1$

$$d(i,j) = \begin{cases} \{1, \{0\}\} & \text{for } i < L_j, \\ \{2, \{0\}\} & \text{for } i \geq L_j, \end{cases} \tag{15}$$

$$d(i,0) = \begin{cases} \{1, \{k \mid 1 \leq i < S_k\}\} & \text{for } i < L_0, \\ \{2, \{k \mid 1 \leq i < S_k\}\} & \text{for } i \geq L_0. \end{cases} \tag{16}$$

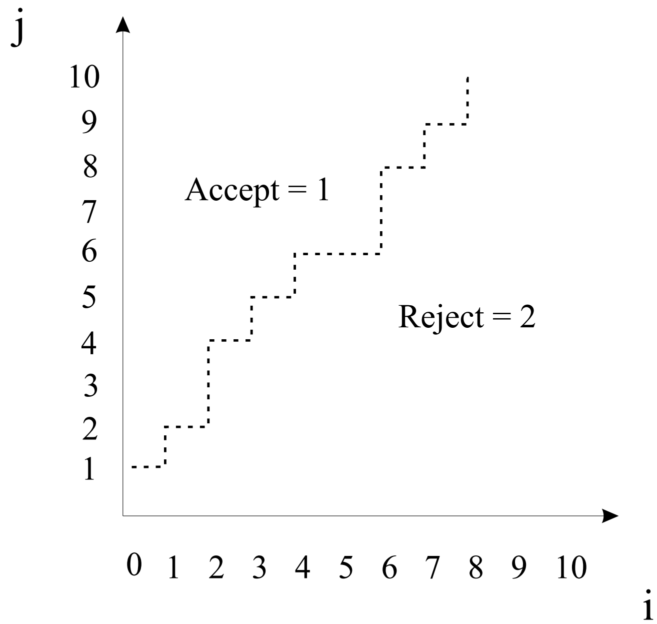


Figure 2: An example of a monotone switching curve.

Furthermore, $\{L_j\}_{j=0}^{\ell}$ and $\{S_j\}_{j=1}^{\ell}$ are ordered by customer class; $L_1 \leq L_2 \leq \cdots \leq L_\ell$ and $S_1 \leq S_2 \leq \cdots \leq S_\ell$. The elements L_j are called **arrival** control limits while the S_j are called **service** control limits for Class j. Notice as is displayed in Figure 2, one may make the alternative statement that for each fixed i, there exists control limits L_i' such that for $0 < j < L_i'$ it is optimal to reject customers in state (i, j).

Given Theorem 1, a few observations are in order. If d^∞ is a monotone switching curve policy with the prescribed partial orderings, customers are accepted by the gatekeeper when serving Class j as long as the number of customers in the system is less than L_j, while customers of Class k are recommended for service by the server when the number of customers is less than S_k; hence the term “control limits”. Furthermore, we make no assertion about the relationship of L_0 to the other control limits.

5 Results

In this section we prove the main results of the paper; the existence of a gain optimal policy that is a generalized monotone switching curve and that the optimal gain is increasing in q. The proofs of the following two results are delayed until the Appendix.

Proposition 6 *For fixed i and j such that $1 \leq i \leq U$ and $1 \leq j \leq \ell$ the following hold,*

1. $\Delta_i h^q(i,j) \leq \Delta_i h^q(i-1,0)$,

2. $\Delta_i^2 h^q(i,0) \leq 0$,

3. $\Delta_i^2 h^q(i,j) \leq 0$,

Proposition 7 *For each i and j such that $i \leq U+1$ and $1 \leq j \leq \ell$, $\Delta_j \Delta_i h^q(i,j) \geq 0$.*

This leads to the proof of Theorem 1.

Proof of Theorem 1: We first show the existence of arrival and service control limits. Fix j such that $j \in \{0, 1, \ldots, \ell\}$. Recall that it is optimal to reject an arriving customer when in state (i,j) if $\Delta_i h^q(i,j) \leq 0$. Let i^* be the lowest number of customers in the system for which this holds. The last two statements of Proposition 6 imply that $\Delta_i h^q(i,j) \leq 0$ for all $i \geq i^*$; i^* is an arrival control limit. Since j was an arbitrary customer class, we have the existence of optimal arrival control limits.

It is optimal not to serve a Class j customer after assessment when $h^q(i,j) \leq h^q(i-1,0) - c_r$. A similar argument to the previous for arrival control limits holds for the existence of service control limits using statement 1 of Proposition 6.

Suppose L_j is a gain optimal arrival control limit of Class j customers where $j \geq 1$. For all states i such that $i < L_j$ we have $\Delta_i h^q(i,j) \geq 0$. Proposition 7 guarantees that we also have $\Delta_i h^q(i,j+1) \geq 0$ for all $i < L_j$. Hence, $L_j \leq L_{j+1}$. The result follows. The order of the service control limits follows in the same manner. ∎

The final contribution of this paper examines the average reward as the probability that rejection by the server is honored increases. We first state a useful lemma which appears as Proposition 8.6.1b in Puterman [15].

Lemma 8 (Puterman [15]) *For a constant g_n and a vector h_n on the state space $\mathbb{X}$. Suppose $d_{n+1} \in arg\max_{d \in D}\{r_d + P_d h_n\}$. If*

$$\max_{d \in D}\{r_d - g_n 1 + (P_d - I)h_n\} > 0$$

for a state s which is recurrent under d_{n+1}, then $g_{n+1} > g_n$.

We are now ready to prove Theorem 2.

Proof of Theorem 2: Assume that it is not optimal to recommend serving some customer class k' when in state $(i, 0)$ where $i \leq L_\ell$ (so $(i, 0)$ is recurrent). Suppose $(g(q), h^q)$ satisfies the AOE when the probability that rejection by the server is honored is q. Let $q'' > q$. Consider the system where q'' is the probability that rejection by the server is honored. Let $g(q'')$ denote the optimal gain for this system. Finally, let $d' \in arg\max_{d \in D}\{r_d + P_d h^q\}$.

$$\begin{aligned} B(g(q), h^q) &\equiv -c(i) - g(q) + \alpha\left(\sum_{k=1}^{\ell} p_k[q''V^q(i,k) + (1-q'')h^q(i,k)]\right) \\ &\quad +[1-(\lambda+\alpha)]h^q(i,0) - h^q(i,0) + \lambda\max\{h^q(i+1,0), h^q(i,0)\} \qquad (17) \\ &= \alpha\left(\sum_{k=1}^{\ell} p_k[(q''-q)V^q(i'',k) - (q''-q)h^q(i,k)]\right) > 0, \qquad (18) \end{aligned}$$

where the equality follows since $(g(q), h^q)$ satisfies the AOE for the system with the probability of an honored rejection is q and the inequality follows since $V^q(i, k') > h^q(i, k')$ and $q'' > q$. Applying Lemma 8 we have $g(q'') \geq g_{d'} > g(q)$. If our assumption that it is not optimal to serve a Class k' customer in some state $(i, 0)$ does not hold, i.e., it is optimal to serve every Class k for all i, then the model reduces to the model without the dependence on q; $g(q)$ is independent of q. ■

6 Examples

In this section we present several examples to emphasize the results shown in the previous section. The first example shows that the optimal control limits may change as q changes.

Example 1

Suppose the parameter values are those found in Table 1. Recall when there are i customers in the system it is optimal to accept (reject) an arriving customer during assessment if $\Delta_i h^q(i, 0) \geq (\leq)\, 0$. We see in Figure 3(a) the optimal control limit, L_0, is 2 for q less than approximately .2. It then changes to 3 for

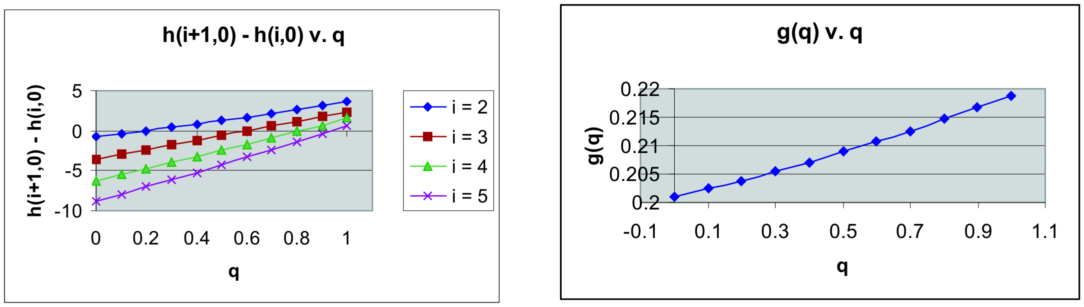
No. of classes	2	
Arrival rate	0.0385	
Service rates	Class 1	0.0385
	Class 2	0.3846
Assess rate	0.5769	
U	10	
R(j)	10	
c_r	1	
Probabilities	Class 1	0.3000
	Class 2	0.7000
c(i)	i*.25	

Table 1: Data for Example 1

$.2 \leq q \leq .6$, 4 for $.6 \leq q \leq .8$, and is 5 for $q \geq .8$. Of course $\Delta_i h^q(i+1,0) \leq \Delta_i h^q(i,0)$ for all q and all i as expected from Proposition 6. Furthermore, one might notice that $\Delta_i h^q(i,0)$ is increasing in q. Although we were unable to prove this holds in general, it held true in every example we considered. Intuitively, this implies that the control limit L_0 increases in q; the more likely the server's decision to reject is honored, the more willing the gatekeeper is to let customers join the queue.

As a decision-maker one would like to know the increase in the long-run average reward say from when $q = 0$ (server's rejection never honored) to $q = 1$ (server's rejection always honored). This in some sense gives a measure of the value of adding a second decision-maker. Figure 3(b) shows the optimal long-run average reward as a function of q. In this example the increase is approximately 8%. The fact that the system prefers Class 2 customers is evidenced by the fact that the optimal service control limits are $S_1 = 0$ (never serve Class 1) and $S_2 = 5$ (always serve Class 2) for all q. The optimal arrival control limit $L_1 = 0$ for all q and L_2 changes from 3 to 4 at $q \approx .35$, from 4 to 5 at $q \approx .75$ and from 5 to 6 at $q \approx .9$.

Example 2



(a) (b)

Figure 3: (a) Shows $\Delta_i h^q(i,0)$ versus q. (b) Shows $g(q)$ versus q. Each are computed using the data from Example 1.

In Example 2 we consider only two cases; when $q = 0$ and when $q = 1$. The optimal control levels for $q = 0$ are

$$S_1 = 1,\ S_2 = 8,\ S_3 = S_4 = 11, \tag{19}$$

$$L_0 = 4,\ L_1 = 3,\ L_2 = L_3 = L_4 = 4. \tag{20}$$

However, when $q = 1$,

$$S_1 = 4,\ S_2 = S_3 = S_4 = 11, \tag{21}$$

$$L_0 = 7,\ L_1 = 6,\ L_2 = L_3 = L_4 = 7. \tag{22}$$

This difference in the optimal policy implies that for some models the second decision-maker can make a crucial difference. In fact, in the present example there is a 27% increase in the optimal gain. This may not be that surprising since Class 4 customers have significantly higher service rates and higher probability that they will arrive while someone is in assessment or service.

7 Conclusions

We have shown that the optimal structure of admission control and service policies is that of monotone switching curve policy. The control limits are ordered by the service rate and the reward offered by each

No. of classes	4	
Arrival rate	0.1250	
Service rates	Class 1	0.0500
	Class 2	0.1500
	Class 3	0.3000
	Class 4	0.5000
Assess rate	0.3750	
U	10	
R	10	
c_r	0	
Probabilities	Class 1	0.1667
	Class 2	0.1667
	Class 3	0.1667
	Class 4	0.5000
c(i)	i*.1000	

Table 2: Data for Example 2

class of customer. Moreover, each control limit is inherently linked to the probability that the server's rejection is honored by the system. We concluded that one way to measure the value of the second decision-maker is to consider the change in the long-run average reward as q changes. In Example 2 the percent increase was large and the policy so significantly different for varying values of q that the savvy manager would need to give serious consideration to modifying the system with only one decision-maker.

There is a nice alternative proof of Theorem 2 that was provided by the referee. For each fixed policy in the system whose second decision-maker is honored with probability q, the proof uses a sample path argument to construct a policy in the system whose second decision-maker is honored with probability $q' \geq q$ that has the same average reward. Since the optimal average reward in the latter system maximizes the average reward over all policies in this system, the result follows.

It might also be noted that in fact the addition of a second decision-maker is equivalent to a single gatekeeper that performs admission control upon arrival and upon assessment of each customer. This follows since events may only occur one at a time. However, we find the explanation with two decision-makers more illuminating.

8 Acknowledgements

I would like to thank Dr. Maarten Oosten for suggesting this problem during one of our many conversations while we were both visiting the University of British Columbia. I would also like to thank Professors Robert D. Foley and Hayriye Ayhan of the Georgia Institute of Technology and Professor Rhonda Righter of The University of California at Santa Barbara for many interesting discussions during the early stages of this research.

Further thanks must also be given to the referee for many helpful comments and considerations that aided in expanding the scope and in the enhancing the readability of the paper.

9 Appendix

Proof of Proposition 6: Fix j. Consider the following inductive statement we refer to as statement $\mathcal{I}_i$

$$\Delta_i h^q(i,j) \leq \Delta_i h^q(i-1,0), \tag{23}$$

$$\Delta_i^2 h^q(i-2,0) \leq 0, \tag{24}$$

$$\Delta_i^2 h^q(i,j) \leq 0, \tag{25}$$

for $i = 2, \ldots, U$. We begin by proving $\mathcal{I}_U$. We do so by proving (23), (24), (25) separately.

Proof of (23)**:** From (10) we have,

$$\begin{aligned} \Delta_i h^q(U,j) &= -\Delta_i c(U) + \mu_j \Delta_i h^q(U-1,0) + [1-(\lambda+\mu_j)]\Delta_i h^q(U,j) \\ &\quad + \lambda \Delta_i h^q(U,j). \end{aligned} \tag{26}$$

Hence, $\mu_j[\Delta_i h^q(U,j) - \Delta_i h^q(U-1,0)] = -\Delta_i c(U) \leq 0$ and (23) holds for $i = U$. In fact, (23) trivially also holds for $i = U-1$ since,

$$\mu_j[\Delta_i h^q(U-1,j) - \Delta_i h^q(U-2,0)] = -\Delta_i c(U-1) - \lambda \max\{\Delta_i h^q(U-1,j), 0\} \leq 0. \tag{27}$$

Proof of (24)**:** From (11),

$$\alpha\Delta_i^2 h^q(U-2,0) = -\Delta_i c(U-1) + \alpha(\sum_{k=1}^{\ell} p_k[q(\max\{h^q(U-1,0) - c_r - h^q(U,k), 0\}$$
$$- \max\{h^q(U-2,0) - c_r - h^q(U-1,k), 0\})$$
$$+\Delta_i h^q(U-1,k) - \Delta_i h^q(U-2,0)]). \quad (28)$$

Consider

$$b_1(U,k) \equiv \max\{h^q(U-1,0) - c_r - h^q(U,k), 0\} - \max\{h^q(U-2,0) - c_r - h^q(U-1,k), 0\}. \quad (29)$$

Let $\mathcal{S}_i$ be the set of customer classes that are optimal to serve when there are i customers in the system. Similarly for N_i and "do not serve". For example, if the optimal decision rule is to serve Class j customers when there are i and $i+1$ customers in the system then we write $j \in \mathcal{S}_i\mathcal{S}_{i+1}$. Furthermore, recall that it is optimal not to serve Class j customers when there are i customers in the system if $h^q(i-1,0) - c_r - h^q(i,j) \geq 0$. By (27) this implies that $\mathcal{N}_{U-1} \subseteq \mathcal{N}_U$. There are three cases to consider.

$$b_1(U,k) = \begin{cases} 0 & \text{if } k \in \mathcal{S}_{U-1}\mathcal{S}_U, \\ -h^q(U-2,0) - c_r - h^q(U-1,k) & \text{if } k \in \mathcal{S}_{U-1}\mathcal{N}_U, \\ \Delta_i h^q(U-2,0) - \Delta_i h^q(U-1,k) & \text{if } k \in \mathcal{N}_{U-1}\mathcal{N}_U. \end{cases} \quad (30)$$

The non-negativity of $b_1(U,k)$ in the first two cases is trivial since $k \in \mathcal{S}_{U-1}$. Suppose the third case holds. Combining $b_1(U,k)$ with the rest of the summand in (28) we have

$$b_1(U-1,k) + \Delta_i h^q(U-1,k) - \Delta_i h^q(U-2,0)] = (1-q)[\Delta_i h^q(U-1,k) - \Delta_i h^q(U-2,0)] \leq 0, \quad (31)$$

where the inequality follows from (27). Since $c(i)$ is increasing in i, $-\Delta_i c(i) \leq 0$, and (24) holds.

Despite the fact that for the inductive step we need only show $\Delta_i^2 h^q(U-2,0) \leq 0$, to complete the proof of Proposition 6 we require that the concavity holds at $(U-1,0)$ and $(U,0)$ as well. It is not difficult to see that a similar argument to the one just made yields these results.

Proof of (25)**:** From (10)

$$\mu_j \Delta_i^2 h^q(U,j) \;=\; -\Delta_i^2 c(U) + \mu_j \Delta_i^2 h^q(U-1,0) \;\le 0, \tag{32}$$

where the inequality follows since $c(i)$ is convex in i and $\Delta_i^2 h^q(U-1,0) \le 0$ by the previous remarks. Thus, statement $\mathcal{I}_U$ holds.

Assume now that statement $\mathcal{I}_{i+1}$ holds. We must prove (23), (24) and (25) for i.

Proof of (23)**:** From (10),

$$\mu_j[\Delta_i h^q(i,j) - \Delta_i h^q(i-1,0)] \;=\; -\Delta_i c(i) + \lambda[\max\{\Delta_i h^q(i+1,j),0\} - \max\{\Delta_i h^q(i,j),0\}]. \tag{33}$$

Denote the set of customer classes which if they are in service it is optimal to accept an arriving when there are i customers in the system by $\mathcal{A}_i$. Similarly if it is optimal to reject a customer by $\mathcal{R}_i$. Note that $j \in \mathcal{A}_i$ if $\Delta_i h^q(i,j) \ge 0$. Consider the second term in (33)

$$b_2(i,j) \equiv \max\{\Delta_i h^q(i+1,j),0\} - \max\{\Delta_i h^q(i,j),0\}. \tag{34}$$

There are four possibilities

$$b_2(i,j) = \begin{cases} 0 & \text{if } j \in \mathcal{R}_i \mathcal{R}_{i+1}, \\ -\Delta_i h^q(i,j) & \text{if } j \in \mathcal{A}_i \mathcal{R}_{i+1}, \\ \Delta_i^2 h^q(i,j) & \text{if } j \in \mathcal{A}_i \mathcal{A}_{i+1}, \\ \Delta_i h^q(i+1,j) & \text{if } j \in \mathcal{R}_i \mathcal{A}_{i+1} \end{cases} \tag{35}$$

In each of the first two cases the non-positivity is guaranteed since “accept” being optimal in state (i,j) implies $\Delta_i h^q(i,j) \ge 0$. Since $c(i)$ is increasing in i (33) yields (23).

The last two cases require more care. Suppose $j \in \mathcal{A}_i \mathcal{A}_{i+1}$ so that $b_2(i,j) = \Delta_i^2 h^q(i,j)$. With a little algebra in (33) we obtain

$$\begin{aligned} (\lambda+\mu_j)[\Delta_i h^q(i,j) - \Delta_i h^q(i-1,0)] \;&=\; -\Delta_i c(i) + \lambda[\Delta_i h^q(i+1,j) - \Delta_i h^q(i-1,0)] \\ &=\; -\Delta_i c(i) + \lambda[\Delta_i h^q(i+1,j) - \Delta_i h^q(i,0)] \\ &\quad\; +\lambda \Delta_i^2 h^q(i-1,0). \end{aligned} \tag{36}$$

The last two terms are non-positive by the induction hypothesis. Now suppose $j \in \mathcal{R}_i\mathcal{A}_{i+1}$ so that $b_2(i,j) = \Delta_i h^q(i+1,j)$. Then (33) yields,

$$\begin{aligned}(\lambda+\mu_j)[\Delta_i h^q(i,j) - \Delta_i h^q(i-1,0)] &= -\Delta_i c(i) + \lambda[\Delta_i h^q(i+1,j) + \Delta_i h^q(i,j) - \Delta_i h^q(i-1,0)] \\ &= -\Delta_i c(i) + \lambda[\Delta_i h^q(i+1,j) - \Delta_i h^q(i,0)] + \Delta_i^2 h^q(i-1,0)] \\ & \quad \Delta_i h^q(i,j) \end{aligned} \tag{37}$$

The non-positivity of $\Delta_i h^q(i,j)$ holds since "reject" is optimal in state (i,j). The second and third terms are non-positive by the induction hypothesis. Hence, (23) holds.

Proof of (24)**:** This follows in the same manner as the proof of $\Delta_i^2 h^q(U-2) \leq 0$.

Proof of (25)**:** From (10)

$$\begin{aligned}\mu_j \Delta_i^2 h^q(i,j) &= -\Delta_i^2 c(i) + \mu_j \Delta_i^2 h^q(i-1,0) + \lambda[\max\{\Delta_i h(i+2,j),0\} - \max\{\Delta_i h(i+1,j),0\} \\ & \quad -[\max\{\Delta_i h(i+1,j),0\} - \max\{\Delta_i h(i,j),0\}]] \end{aligned} \tag{38}$$

By the induction hypothesis $\mathcal{A}_{i+2} \subseteq \mathcal{A}_{i+1}$, hence there are six cases to consider. With a little algebra,

$$(\lambda+\mu_j)\Delta_i^2 h^q(i,j) = \begin{cases} -\Delta_i^2 c(i) + \mu_j \Delta_i^2 h^q(i-1,0) + \lambda \Delta_i^2 h(i+1,j) & \text{if } j \in \mathcal{A}_i\mathcal{A}_{i+1}\mathcal{A}_{i+2}, \\ -\Delta_i^2 c(i) + \mu_j \Delta_i^2 h^q(i-1,0) + \lambda \Delta_i h(i+1,j) & \text{if } j \in \mathcal{A}_i\mathcal{R}_{i+1}\mathcal{R}_{i+2}, \\ -\Delta_i^2 c(i) + \mu_j \Delta_i^2 h^q(i-1,0) - \lambda \Delta_i h(i+1,j) & \text{if } j \in \mathcal{A}_i\mathcal{A}_{i+1}\mathcal{R}_{i+2}, \end{cases} \tag{39}$$

and

$$\mu_j \Delta_i^2 h^q(i,j) = \begin{cases} -\Delta_i^2 c(i) + \mu_j \Delta_i^2 h^q(i-1,0) + \lambda[\Delta_i^2 h(i+1,j) - \Delta_i h(i+1,j)] & \text{if } j \in \mathcal{R}_i\mathcal{A}_{i+1}\mathcal{A}_{i+2}, \\ -\Delta_i^2 c(i) + \mu_j \Delta_i^2 h^q(i-1,0) & \text{if } j \in \mathcal{R}_i\mathcal{R}_{i+1}\mathcal{R}_{i+2}, \\ -\Delta_i^2 c(i) + \mu_j \Delta_i^2 h^q(i-1,0) - 2\lambda \Delta_i h(i+1,j) & \text{if } j \in \mathcal{R}_i\mathcal{A}_{i+1}\mathcal{R}_{i+2}. \end{cases} \tag{40}$$

Each of the cases above is non-negative by the induction hypothesis or by noting which action is optimal in the respective states as before.

To complete the proof one need only show $\Delta_i^2 h(1,j) \leq 0$. However, this follows in precisely the same manner as the previous argument. Hence $\mathcal{I}_i$ holds and the result is proven. ∎

Proof of Proposition 7: By induction. Consider the case of $i = U + 1$. From (10)

$$\begin{aligned} \Delta_i h^q(U+1,j) &= -\Delta_i c(U+1) + \mu_j \Delta_i h^q(U,0) + [1-(\lambda+\mu_j)]\Delta_i h^q(U+1,j) \\ &\quad + \lambda \Delta_i h^q(i,j). \end{aligned}$$

Thus,

$$\mu_{j+1}\Delta_j \Delta_i h^q(U+1,j) = (\mu_{j+1}-\mu_j)[\Delta_i h^q(U,0) - \Delta_i h^q(U+1,j)] \ \geq \ 0$$

where the inequality follows from Proposition 6; the result holds for $U + 1$. Assume that it holds for $i + 1$ and consider

$$\begin{aligned} \Delta_i h^q(i,j) &= -\Delta_i c(i) + \mu_j \Delta_i h^q(i-1,0) + [1-(\lambda+\mu_j)]\Delta_i h^q(i,j) \\ &\quad + \lambda[\max\{\Delta_i h^q(i+1,j),0\} - \max\{\Delta_i h^q(i,j),0\}] + \lambda \Delta_i h^q(i,j). \end{aligned} \tag{41}$$

A little algebra yields,

$$\begin{aligned} \mu_{j+1}\Delta_j \Delta_i h^q(i,j) &= (\mu_{j+1}-\mu_j)[\Delta_i h^q(i-1,0) - \Delta_i h^q(i,j)] \\ &\quad + \lambda[\max\{\Delta_i h^q(i+1,j+1),0\} - \max\{\Delta_i h^q(i+1,j),0\} \\ &\quad - \max\{\Delta_i h^q(i,j+1),0\} + \max\{\Delta_i h^q(i,j),0\}]. \end{aligned} \tag{42}$$

Let

$$b_3(i,j) \equiv (\mu_{j+1}-\mu_j)[\Delta_i h^q(i-1,0) - \Delta_i h^q(i,j)], \tag{43}$$

and

$$\begin{aligned} b_4(i,j) &\equiv \max\{\Delta_i h^q(i+1,j+1),0\} - \max\{\Delta_i h^q(i+1,j),0\} \\ &\quad - \max\{\Delta_i h^q(i,j+1),0\} + \max\{\Delta_i h^q(i,j),0\}. \end{aligned} \tag{44}$$

Proposition 6 and the assumption that μ_j is increasing assures us that $b_3(i,j) \geq 0$. Hence, we must show $b_4(i,j) \geq 0$.

Note that since $h^q(i,j)$ is concave in i, $\mathcal{A}_{i+1} \subseteq \mathcal{A}_i$. Furthermore, by the induction hypothesis ($j \in$

$\mathcal{A}_{i+1}) \Rightarrow (j+1 \in \mathcal{A}_{i+1})$. Hence, there are 8 cases to consider. These cases lead to,

$$(\lambda + \mu_{j+1})\Delta_j \Delta_i h^q(i,j) = \begin{cases} b_3(i,j) + \Delta_j \Delta_i h^q(i+1,j) & \text{if } (j \in \mathcal{A}_i\mathcal{A}_{i+1}), (j+1 \in \mathcal{A}_i\mathcal{A}_{i+1}), \\ b_3(i,j) + \Delta_i h^q(i+1,j) & \text{if } (j \in \mathcal{A}_i\mathcal{R}_{i+1}), (j+1 \in \mathcal{A}_i\mathcal{A}_{i+1}), \\ b_3(i,j) & \text{if } (j \in \mathcal{A}_i\mathcal{R}_{i+1}), (j+1 \in \mathcal{A}_i\mathcal{R}_{i+1}), \end{cases} \tag{45}$$

and

$$b_4(i,j) = \begin{cases} \Delta_i h^q(i+1,j) - \Delta_i^2 h^q(i,j) & \text{if } (j \in \mathcal{A}_i\mathcal{A}_{i+1}), (j+1 \in \mathcal{R}_i\mathcal{A}_{i+1}) \\ \Delta_i h^q(i,j) & \text{if } (j \in \mathcal{A}_i\mathcal{R}_{i+1}), (j+1 \in \mathcal{R}_i\mathcal{R}_{i+1}) \\ -\Delta_i h^q(i,j) & \text{if } (j \in \mathcal{R}_i\mathcal{R}_{i+1}), (j+1 \in \mathcal{A}_i\mathcal{R}_{i+1}) \\ 0 & \text{if } (j \in \mathcal{R}_i\mathcal{R}_{i+1}), (j+1 \in \mathcal{A}_i\mathcal{R}_{i+1}) \end{cases} \tag{46}$$

Each case in the (45) and (46) is non-negative by applying the induction hypothesis, Proposition 6, or by appealing to the fact that "accept" or "reject" is optimal in the respective states. Since in each case (42) implies $\Delta_j \Delta_i h^q(i,j) \geq 0$, the result follows. ■

References

[1] Eitan Altman. Non zero-sum stochastic games in admission, service, and routing control in queueing systems. *Queueing Systems*, 23:256–279, 1996.

[2] Eitan Altman. A Markov game approach for optimal routing into a queueing network. In M. Bardi, T.E.S. Raghavan, and T. Parthasarathy, editors, *Annals of Dynamic Games*, volume 5 of *Stochastic and Differential Games, Theory and Numerical Methods*, pages 359–376. Birkhaüser, Boston, Basel, Berlin, 1999.

[3] Eitan Altman. Applications of markov decision processes in telecommunications: A survey. In Eugene Feinberg and Adam Shwartz, editors, *Markov Decision Processes: Models, Methods, Directions and Open Problems*. Kluwer Academic Publishers, tentatively 2001. submitted.

[4] Eitan Altman and Shaler Stidham, Jr. Optimality of monotone policies for two-action markovian decision processes, with applications to control of queues with delayed information. *Queueing Systems*, 1995.

[5] Hussein A. Ghoneim and Shaler Stidham, Jr. Control of arrivals to two queues in series. *European Journal of Operational Research*, 21:399–409, 1985.

[6] Bruce Hajek. Optimal control of two interacting service stations. *IEEE Transactions on Automatic Control*, AC-29(6):491–499, June 1984.

[7] Werner E. Helm and Karl-Heinz Waldmann. Optimal control of arrivals to multi-server queues in a random environment. *Journal of Applied Probability*, 21:602–615, 1984.

[8] Onésimo Hernández-Lerma and Jean B. Lasserre. *Discrete-Time Markov Control Processes: Basic Optimality Criteria*. Springer-Verlag Inc., New York, 1996.

[9] Daniel P. Heyman. Optimal operating policies for m/g/1 queueing systems. *Operations Research*, 16:362–382, 1968.

[10] Mark E. Lewis and Martin L. Puterman. A note on bias optimality in controlled queueing systems. *Journal of Applied Probability*, 37(1), 2000.

[11] Mark E. Lewis and Martin L. Puterman. A probabilistic analysis of bias optimality in unichain Markov decision processes. *IEEE Transactions on Automatic Control*, 46(1):1–5, January 2001.

[12] Steven A. Lippman. Applying a new device in the optimization of exponential queueing systems. *Operations Research*, 23(4):687–712, 1975.

[13] Steven A. Lippman and Sheldon M. Ross. The streetwalker's dilemma: A job shop model. *SIAM Journal of Applied Mathematics*, 20(3):336–342, 1971.

[14] Bruce L. Miller. A queueing reward system with several customer classes. *Management Science*, 16(3):234–245, 1969.

[15] Martin L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley and Sons, Inc., New York, 1994.

[16] Shaler Stidham, Jr. Optimal control of admission to a queueing system. *IEEE Transactions on Automatic Control*, AC-30(8):705–713, 1985.

[17] Susan H. Xu and J. George Shanthikumar. Optimal expulsion control – a dual approach to admission control of an ordered-entry system. *Operations Research*, 41(6):1137–1152, November-December 1993.

[18] M. Yadin and P. Naor. Queueing systems with a removable service station. *Operations Research Quarterly*, 14(4):393–405, 1963.