Baltzer Journals June 28, 1997

On the Optimal Control of Arrivals to a Single Queue with Arbitrary Feedback Delay

Joy Kuri[1] and Anurag Kumar[2]

[1] *Departement de Génie Electrique et de Génie Informatique, École Polytechnique, Case Postale 6079, Succ A, H3C 3A7, Montréal, CANADA*
E-mail: kuri@comm.polymtl.ca

[2] *Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore-560012, INDIA*
E-mail: anurag@ece.iisc.ernet.in

We consider a problem of admission control to a single queue in discrete time. The controller has access to k–step old queue lengths only, where k can be arbitrary. The problem is motivated, in particular, by recent advances in high–speed networking where information delays have become prominent.

We formulate the problem in the framework of Completely Observable Controlled Markov Chains, in terms of a multi–dimensional state variable. Exploiting the structure of the problem, we show that under appropriate conditions, the multi–dimensional Dynamic Programming Equation (DPE) can be reduced to a unidimensional one. We then provide simple computable upper and lower bounds to the optimal value function corresponding to the reduced unidimensional DPE. These upper and lower bounds, along with a certain relationship among the parameters of the problem, enable us to deduce partially the structural features of the optimal policy. Our approach enables us to recover simply, in part, the recent results of Altman & Stidham, who have shown that a multiple–threshold–type policy is optimal for this problem. Further, under the same relationship among the parameters of the problem, we provide easily computable upper bounds to the multiple thresholds and show the existence of simple relationships among these upper bounds. These relationships allow us to gain very useful insights into the nature of the optimal policy. In particular, the insights obtained are of great importance for the problem of actually computing an optimal policy because they reduce the search space enormously.

Keywords: Markov Decision Theory, Delayed Feedback Information, Threshold Policies

1 Introduction

The problem of optimal control of arrivals to a queue under delayed feedback information has recently become important. It is motivated, in particular, by developments in high–speed networking. In high–speed networks, link transmission times have decreased sharply relative to the constant signal propagation times. This has affected the control of such systems, since by the time the state information reaches the controller, the state may have changed, due to arrivals/departures. Owing to this, the state of a system as seen by the controller at any time is *not* the present state of the system. Hence, it becomes important to consider problems of optimal control under the constraint of delayed feedback information.

A standard framework often applied to problems of optimal control is that of Markov Decision Theory (with appropriate Markovian assumptions on the system's random variables). Problems with delayed feedback information can be naturally modelled in the framework of Markov Decision Theory as Partially Observable Controlled Markov Chains (PO–CMC). In this approach, the "state" of the PO–CMC is defined as a conditional probability measure on the space of the underlying unobservable system state (the so–called "information state"). Often, however, a direct formulation in terms of a Completely Observable Controlled Markov Chain (CO–CMC) is possible by appropriately defining an enlarged state so that the enlarged state encompasses all the relevant information. This approach has been followed in Altman & Nain ([1]), Kuri & Kumar ([10], [11]), Artiges ([5]), Kuri ([12]) and recently in Altman & Stidham ([4]). Related work on control with delayed information can be found in Koole ([8]), Altman & Koole ([3]) and Altman & Koole ([2]).

In Altman & Nain ([1]), Kuri & Kumar ([10], [11]), Artiges ([5]), discrete–time optimal control problems in discrete–time queueing models were studied. The feedback delay was restricted to one time–step and it was shown that the optimal control policies are of the threshold or switching types. In all of the above papers, the technique of Value Iteration (see, for example, Kumar & Varaiya [9], Hajek [6]) was used to establish that the optimal value function had properties which implied that the optimal policy was of threshold/switchover type.

In Kuri ([12]), Altman & Stidham ([4]) and the present paper, arbitrary delays are considered. In a general context, Altman & Stidham ([4]) use a stochastic dominance approach, coupled with the inductive dynamic programming technique to establish that for two–action Markovian Decision Processes, threshold type policies are optimal.

We focus on a specific discrete–time queueing control problem. There is a single queue to which customers arrive randomly at the beginning of each slot. Departures also occur randomly at the end of each slot. A controller decides, at every slot, whether an arrival can be admitted to the queue or not. The constraint is that the controller has access only to k–slot old queue lengths, where k can be arbitrary. Admission to the queue yields a reward, but this is counter–balanced by the cost of holding customers in the queue. The objective is to minimize the net discounted cost over an infinite horizon.

In terms of this problem, the results of Altman & Stidham ([4]) say the following: for a particular pattern of arrivals to the queue during the k–slot long feedback delay period, the optimal policy is characterised by a threshold on the queue length. This existence result specifies the nature of the optimal policy; however, several important questions remain: in particular, how are the thresholds corresponding to different arrival patterns related?

In the present paper (as in Kuri ([12])), the problem is formulated as a CO–CMC in terms of a multidimensional state variable, which gathers all the relevant information. We then show that *under appropriate conditions*, the multidimensional Dynamic Programming Equation (DPE) can be reduced to a unidimensional DPE. This is done by exploiting the observation that there are many different states in the multidimensional formulation for which the optimal costs are actually equal. This fact can be further exploited in a straightforward way to see that, for suitable values of the parameters of the problem, it is optimal to not accept an arrival whenever the observed queue length is larger than a certain value.

The contribution of this paper is four–fold. Firstly, we utilise the observation that the optimal costs corresponding to different multidimensional states are equal to reduce the dimensionality of the state space to unity under appropriate conditions. This is a considerable simplification of the problem. Secondly, we provide simple *computable* upper and lower bounds to the optimal value function corresponding to the "reduced" unidimensional DPE. Thirdly, when a certain relationship between the parameters of the problem holds, our approach easily leads to a quick partial characterization of the optimal policy. Thus, in a very simple way, we recover, in part, the nature of the optimal policy. Fourthly, for the same relationship among the parameters, our approach enables us to provide easily computable upper bounds to the thresholds. In addition, it is shown that there are simple relationships between the upper bounds to the thresholds corresponding to different arrival patterns. These relationships allow us to gain very useful insights into the nature of the optimal policy. In particular, the insights obtained are of great importance for the problem of actually computing an optimal policy because they reduce the search space enormously.

The outline of the paper is as follows. In the next section, we consider the underlying queueing model. This is followed in Section 3 by the formulation of the problem in terms of a CO–CMC with a multi–dimensional state space. Section 4 shows how the state space can be collapsed to a single–dimensional one under appropriate conditions. In Section 5, we provide *computable* upper and lower bounds to the single–dimensional optimal value function. This is followed in Section 6 by computable upper bounds on the thresholds for the optimal policy.

2 The Underlying Queueing Model

We consider a control system in discrete time. The duration between two successive time epochs is defined as a "slot." There is a single queue to which arrivals may occur at the beginning of each slot (just after the slot boundary that determines the beginning of the slot). Precisely at the boundary of each slot, a controller decides, based on the latest available information, whether to admit or not an arrival that may occur immediately afterwards. A departure from the queue may occur at the end of each slot (just prior to the boundary that determines the end of the slot). In terms of the above description, slot n commences with the time epoch at which the controller makes a decision regarding admission of an arrival that may occur immediately afterwards and lasts till the next instant at which the controller is required to decide again.

$a(n) \in \{0,1\}$ denotes the occurrence or non–occurence of an arrival at time $n+$, where $n+$ means immediately after the slot boundary (as in Figure 1). We assume that $a(n)$ is distributed as a Bernoulli random variable with parameter λ:

$$a(n) = \begin{cases} 1 & w.p. \quad \lambda \\ 0 & w.p. \quad 1-\lambda \end{cases}$$

($w.p.$ stands for "with probability"). Similarly, $d(n+1) \in \{0,1\}$ denotes a departure from the queue at the end of slot n, at time $(n+1)-$, where $(n+1)-$ means immediately before the slot boundary. $d(n)$ is assumed to be Bernoulli with parameter μ:

$$d(n) = \begin{cases} 1 & w.p. \quad \mu \\ 0 & w.p. \quad 1-\mu \end{cases}$$

Let $q(n)$ denote the queue length at n. Let $u(n)$ denote the control action taken at the beginning of the n^{th} slot. We note that the control action $u(n)$ in slot n is taken precisely at the boundary where slot n begins. $u(n) = 1$ corresponds to admitting the arrival which may occur at $n+$, and $u(n) = 0$ corresponds to refusing entry. The embedding convention described above is shown in Figure 1.

We then have

$$q(n+1) = (q(n) + a(n)I_{\{u(n)=1\}} - d(n+1))^+$$

where $I_{\{u(n)=1\}}$ is 1 if $u(n) = 1$, and 0 otherwise.

For a control problem without feedback delay, the controller knows $q(n)$ before deciding on $u(n)$. Therefore, for this problem with no feedback delay, $q(n)$ qualifies as the "state". Now consider a control problem with a delay of 1 time slot. At time $(n+1)$, the controller knows $q(n)$, and also whether a customer was admitted at the beginning of slot n. In other words, the controller knows the product $a(n) \times I_{\{u(n)=1\}}$. Let us denote $i(n) = a(n)I_{\{u(n)=1\}}$, where $i(n)$ will be referred

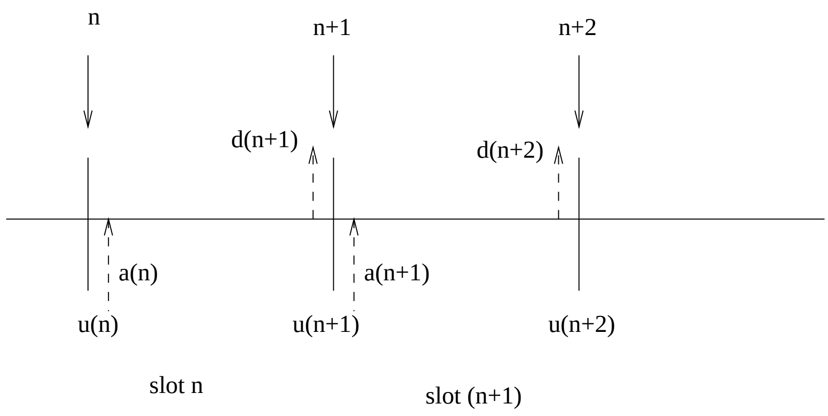

Figure 1: Diagram showing the embedding convention.

to as the "indicator" variable for "an admission in" slot n. Note that $i(n)$ is 1 only when the last control action $u(n)$ was 1 and an arrival occurred at time $n+$ $(a(n) = 1)$. At time $(n+1)$, therefore, $(q(n), i(n))$ qualifies as a state variable for the 1–step delay problem, since, by virtue of the Markovian assumptions on the random variables, it encompasses all known and relevant information. Note that our assumptions of Bernoulli distributions ensure that for a given choice of the sequence $\{u(n)\}$, the process $\{q(n)\}$ evolves as a Markov Chain.

We now introduce some notation in the context of the 1–step delay problem. This will be useful in considering delays of any number of steps. Given a current state $(q(n), i(n)) = (l, i)$, let P_i denote a matrix whose (l, m) element is given as follows: for $l, m = 0, 1 \ldots$ and $i \in \{0, 1\}$,

$$P_i(l, m) = Prob(q(n+1) = m | q(n) = l, i(n) = i)$$

It is clear that a row of P_i specifies the conditional probability distribution of the random variable $q(n+1)$. Obviously,

$$P_0(l, m) = \begin{cases} \mu & \text{if } m = (l-1) \\ 1-\mu & \text{if } m = l \\ 0 & \text{otherwise} \end{cases}$$

when $l \geq 1$, and $P_0(0, 0) = 1$. Similarly,

$$P_1(l, m) = \begin{cases} \mu & \text{if } m = l \\ 1-\mu & \text{if } m = (l+1) \\ 0 & \text{otherwise} \end{cases}$$

From the structure of the matrices P_0 and P_1, the properties in the following Lemma are obvious. Let $P_i(l,\cdot)$ denote the l^{th} row of P_i. Let $X \leq_{st} Y$ denote that random variable X is stochastically less than random variable Y. Then

Lemma 1
For the matrices P_0 and P_1, the following hold:

1. $P_i(l,\cdot) \leq_{st} P_i(l+1,\cdot)$, $l \geq 0$
2. $P_0(l,\cdot) \leq_{st} P_1(l,\cdot)$, $l \geq 0$
3. $P_0(l+1,\cdot) = P_1(l,\cdot)$, $l \geq 0$

We exploit the fact that the rows of P_i are stochastically increasing by using the following well–known result (see Stoyan [13]):

Lemma 2
If X and Y are random variables, then $X \geq_{st} Y \Leftrightarrow$ for all non–decreasing functions f, $E[f(X)] \geq E[f(Y)]$.

3 The Controlled Markov Chain

We turn now to the formulation of the discrete–time Controlled Markov Chain that describes the control problem for a feedback delay of k slots, where $k \geq 1$ (for a formal treatment, see [12]). Just as in the case of the 1–slot delay problem, it is easy to see that the "state" at time n, $s(n)$, can be defined as:

$$s(n) = (q(n-k), i(n-k), i(n-k+1), \ldots, i(n-1))$$

This multi–dimensional state variable captures all the relevant information.

The action at time n, $u(n)$ lies in the set $\{0,1\}$, as before.

We define the one–step cost function of the CMC based on the following considerations. Each queued customer is assumed to cost the system an amount $b < 1$ per slot (a "holding cost"). If n customers are queued, the holding cost per slot is nb. We define the vector $\underline{b}$ to indicate the holding costs corresponding to the number of customers queued:

$$\underline{b} = (0, b, 2b, 3b, \ldots)^t$$

where $(\cdots)^t$ denotes transpose. Each accepted arrival is assumed to yield a one–time reward of 1 upon admission. Let the state $s(n)$ of the CMC be given as $s(n) = (x, i_k, i_{k-1}, \ldots, i_1)$, where i_j denotes the indicator j steps back from the *current time*, i.e., $i(n-j)$. Let $u(n) = u$ denote the action taken. Then, we define the one–step cost $c(s(n), u(n))$ as the "net expected holding cost over step n":

$$c(s(n), u(n)) = bE(q(n)|s(n)) - \lambda(1-b)u$$

Let $P_{(i_k i_{k-1} \cdots i_2 i_1)}$ denote the matrix product

$$P_{i_k} \times P_{i_{k-1}} \times \cdots P_{i_2} \times P_{i_1}.$$

For $x = 0, 1, 2 \cdots$, row x of $P_{(i_k i_{k-1} \cdots i_2 i_1)}$ gives the distribution of the current queue length given a k–step old observed queue length of x and an indicator sequence of $(i_k i_{k-1} \cdots i_2 i_1)$. Following usual practice, we will denote a product of the matrix $P_{(i_k i_{k-1} \cdots i_2 i_1)}$ and the vector $\underline{b}$ as the vector $P_{(i_k i_{k-1} \cdots i_2 i_1)}\underline{b}$. Further, $P_{(i_k i_{k-1} \cdots i_2 i_1)}\underline{b}(x)$ will denote the x^{th} element of the vector $P_{(i_k i_{k-1} \cdots i_2 i_1)}\underline{b}$. Note that $P_{(i_k i_{k-1} \cdots i_2 i_1)}\underline{b}$ gives a vector of expected holding costs over the current step. In terms of this notation, we can rewrite the one–step cost as

$$c(s(n), u(n)) = (P_{(i_k i_{k-1} \cdots i_2 i_1)}\underline{b})(x) - \lambda(1-b)u$$

$(P_{(i_k i_{k-1} \cdots i_2 i_1)}\underline{b})(x)$ denotes the holding cost over the current step when the initial state is $(x, (i_k i_{k-1} \cdots i_2 i_1))$. The term $\lambda(1-b)u$ denotes the expected reward for action u, and the net expected cost is simply the difference of the two terms.

We shall use the standard discounted cost criterion as our performance metric. Corresponding to a policy π, an initial state $s(0)$, and a discount factor $\beta \in (0, 1)$, the cost criterion is given by

$$E^{\pi}_{s(0)}[\sum_{n=0}^{\infty} \beta^n c(s(n), u(n))]$$

where, as usual, $E^{\pi}_{s(0)}$ is the expectation defined with respect to the probability measure induced by policy π and initial state $s(0)$ (see, for example, [7]).

The dynamic programming equation for the k–step delay problem can now be written down. For $s = (x, (i_k i_{k-1} \cdots i_2 i_1))$, let $\underline{i} = (i_k i_{k-1} \cdots i_2 i_1)$; then we write $s = (x, \underline{i})$. Let $V^{\pi}_{\beta,(\underline{i})}(x)$ denote the expected cost when the initial state is $(x, \underline{i})$ and the policy followed is π. For each $\underline{i}$, we view $V^{\pi}_{\beta,(\underline{i})}(x)$ as a vector indexed by $x \in \{0, 1, 2, \cdots\}$. As usual, we define

$$V^{*}_{\beta,(\underline{i})}(x) = \min_{\pi} V^{\pi}_{\beta,(\underline{i})}(x)$$

and the Dynamic Programming equation is

$$V^{*}_{\beta,(i_k i_{k-1} \cdots i_2 i_1)}(x)$$

$$
\begin{aligned}
= \quad & \min\{(P_{(i_k i_{k-1}\cdots i_2 i_1)}\underline{b})(x) + \beta(P_{i_k}V^*_{\beta,(i_{k-1}i_{k-2}\cdots i_1 0)})(x), \\
& (P_{(i_k i_{k-1}\cdots i_2 i_1)}\underline{b})(x) - \lambda(1-b) + \beta\overline{\lambda}(P_{i_k}V^*_{\beta,(i_{k-1}i_{k-2}\cdots i_1 0)})(x) \\
& +\beta\lambda(P_{i_k}V^*_{\beta,(i_{k-1}i_{k-2}\cdots i_1 1)})(x)\} \\
= \quad & (P_{(i_k i_{k-1}\cdots i_2 i_1)}\underline{b})(x) + \beta(P_{i_k}V^*_{\beta,(i_{k-1}i_{k-2}\cdots i_1 0)})(x) \\
& + \beta\lambda\min\{0, (P_{i_k}V^*_{\beta,(i_{k-1}i_{k-2}\cdots i_1 1)})(x) \\
& - (P_{i_k}V^*_{\beta,(i_{k-1}i_{k-2}\cdots i_1 0)})(x) - \frac{1-b}{\beta}\} \qquad (1)
\end{aligned}
$$

The first expression in $\min\{\cdots,\cdots\}$ corresponds to a control action 0 (disallow admission) and the second expression corresponds to a control action 1 (allow admission). When the control action is 0, the next state can be characterized as follows: the arrival indicator part of the next state is $(i_{k-1}i_{k-2}\cdots i_1 0)$ (since admission is disallowed), while the observed queue length distribution is given by row x of the matrix P_{i_k}. Therefore, the "cost–to–go" from the next step becomes $(P_{i_k}V^*_{\beta,(i_{k-1}i_{k-2}\cdots i_1 0)})(x)$. Similarly, the terms appearing in the expression for control action 1 can be explained.

From Equation (1), we observe that there is a "coupling" between different indicator sequences. Writing down Equation (1) for each possible indicator sequence, we would get a system of 2^k coupled equations. Essentially, the optimal value function is a real–valued map on the $(k+1)$–dimensional space $\{0,1\}^k \times \mathcal{N}$. This is only to be expected since our state itself is $(k+1)$–dimensional.

It can be shown without difficulty that

(P1) $\forall x \geq 0$, and any indicator sequence $\underline{i}$,

$$V^*_{\beta,(\underline{i})}(x+1) \geq V^*_{\beta,(\underline{i})}(x)$$

Further, let $\underline{j}$ be any sequence and let $\underline{i}$ be "derived" from $\underline{j}$ by flipping one or more 0's occurring in $\underline{j}$ to 1's.

(P2) $\forall x \geq 0$, and $\underline{j}$ and $\underline{i}$ as above,

$$V^*_{\beta,(\underline{i})}(x) \geq V^*_{\beta,(\underline{j})}(x)$$

Lemma 3
$V^*_{\beta,(\underline{i})}(x)$ possesses P1 and P2.

Proof: The proof uses Lemma 1 and relies on the familiar technique of showing that when a function possessing properties P1 and P2 is transformed by the Dynamic Programming Operator, the resulting transformed function also has P1 and P2. (The space of such functions is non–empty since the "zero function", which is identically 0 everywhere, has these properties trivially). We omit the details ([12]) since it is a matter of writing down the expressions and checking. □

4 Reduction to a Single Dimensional State Variable

We show below that *for appropriately large values of the observed queue length*, $V^*_{\beta,(i_k i_{k-1}\cdots i_2 i_1)}(.)$ can be written in terms of $V^*_{\beta,\underline{0}}(.)$; here, $\underline{0}$ means that in $(i_k i_{k-1}\cdots i_2 i_1)$,

$$i_k = i_{k-1} = \cdots = i_2 = i_1 = 0.$$

The implication of this is that, for sufficiently large observed queue lengths, an equation of the form of (1) can be "reduced" to one involving only $V^*_{\beta,\underline{0}}(.)$. As we shall see later, it is thus enough to consider only 1 equation instead of the 2^k coupled equations which correspond to the 2^k distinct sequences of $(i_k i_{k-1}\cdots i_2 i_1)$ possible.

The key to the "reduction" referred to above is the observation that for appropriately large values of the observed queue length (k steps old), the current queue length distribution generated by this observed queue length and a given string of indicators, is exactly the same as that generated by a suitably modified observed queue length value (k steps old) and a *particular* indicator string, viz., one with the first $(k-1)$ bits being zeroes. In other words, if the last known queue length is sufficiently large, then the arrivals occurring during the delay period may be taken to have occurred at the beginning itself. This is explained below.

In general, for an observed queue length x, the current queue length would be given by

$$((\cdots((x + i_k - D_{k-1})^+ + i_{k-1} - D_{k-2})^+ + \cdots + i_2 - D_1)^+ + i_1 - D_0)^+$$

where D_j, $0 \le j \le k-1$, is the virtual service random variable j steps back from the current time (it is Bernoulli, taking the value 1 with probability μ). This expression can be simplified when x is sufficiently large for it allows us to ignore the + in $(\cdots)^+$. We explore this idea now.

Let $\underline{i}$ be the given indicator string and let $\underline{i}^{'}$ represent the first or leading $(k-1)$ entries of $\underline{i}$. Let $z(\underline{i}^{'})$ denote the number of zeroes in $\underline{i}^{'}$, so that $0 \le z(\underline{i}^{'}) \le (k-1)$. Consider an observed queue length of $(l + z(\underline{i}^{'}))$, $l \ge 0$. Then the current queue length is given by:

$$((\cdots((l + z(\underline{i}^{'}) + i_k - D_{k-1}) + i_{k-1} - D_{k-2}) + \cdots + i_2 - D_1) + i_1 - D_0)^+$$

Note that since we are starting with $(l + z(\underline{i}^{'}))$ customers (the observed queue length), the expression

$$(\cdots((l + z(\underline{i}^{'}) + i_k - D_{k-1}) + i_{k-1} - D_{k-2}) + \cdots + i_2 - D_1)$$

is non–negative. But

$$z(\underline{i}^{'}) + i_k + i_{k-1} + \cdots + i_2 = (k-1)$$

and using this we get

$$\begin{aligned}&(\cdots((l+z(\underline{i}^{'})+i_k-D_{k-1})+i_{k-1}-D_{k-2})+\cdots+i_2-D_1)\\ &= \quad (\cdots((l+k-1-D_{k-1})-D_{k-2})-\cdots-D_1)\end{aligned}$$

Thus, the current queue length can also be written as

$$((\cdots((l+k-1-D_{k-1})-D_{k-2})-\cdots-D_1)+i_1-D_0)^+$$

and this is nothing but the current queue length generated by an observed queue length of $(l+k-1)$, and the indicator string $(0^{(k-1)}i_1)$.

Now recall that in our problem the one–step cost is determined by the distribution of current queue length. Therefore, if we have two distinct initial states such that the distribution of queue length at time 0 due to each is the same, the costs corresponding to the two states must be the same. Hence we have, for $l \geq 0$,

$$V^*_{\beta,(\underline{i}'1)}(l+z(\underline{i}^{'})) = V^*_{\beta,\underline{0}}(l+k) \tag{2}$$

and

$$V^*_{\beta,(\underline{i}'0)}(l+z(\underline{i}^{'})) = V^*_{\beta,\underline{0}}(l+k-1) \tag{3}$$

Equations (2) and (3) indicate how large the observed queue length (appearing in brackets) should be (corresponding to the given indicator sequence) for the simplification to go through. They also show that the optimal costs for many different states are equal. When $x = l+z(\underline{i}^{'})$, for some $l \geq 0$, we can ignore the $+$ in $(\cdots)^+$ at all but the last place, and this allows us to "pull" all the 1–valued indicators in $\{i_k i_{k-1}\ldots i_2\}$ to the observed queue length position. Clearly, this could not be done if we had fewer than $z(\underline{i}^{'})$ customers to start with.

Theorem 4
For $x \geq k$, the DPE can be written as

$$\begin{aligned}V^*_{\beta,\underline{0}}(x) \quad &= \quad (P_{\underline{0}}\underline{b})(x)+\beta(P_{(0)}V^*_{\beta,\underline{0}})(x)\\ &\quad +\beta\lambda\min\{0,\mu(V^*_{\beta,\underline{0}}(x)-V^*_{\beta,\underline{0}}(x-1))+\\ &\quad \overline{\mu}(V^*_{\beta,\underline{0}}(x+1)-V^*_{\beta,\underline{0}}(x))-\frac{1-b}{\beta}\}\end{aligned}$$

Proof : Suppose the sequence of indicators is denoted by $\underline{i} = (i_k i_{k-1}\cdots i_2 i_1)$. Let the observed queue length be $(l+z(\underline{i}))$, $l \geq 0$; that is, the observed queue length is greater than or equal to the number of zeroes in the indicator sequence. The DPE is given by equation (1):

$$V^*_{\beta,(i_k i_{k-1}\cdots i_2 i_1)}(l+z(\underline{i})) =$$

$$
\begin{aligned}
&(P_{(i_k i_{k-1}\cdots i_2 i_1)}\underline{b})(l+z(\underline{i})) + \beta(P_{i_k}V^*_{\beta,(i_{k-1}i_{k-2}\cdots i_1 0)})(l+z(\underline{i})) \\
&+ \beta\lambda \min\{0, (P_{i_k}V^*_{\beta,(i_{k-1}i_{k-2}\cdots i_1 1)})(l+z(\underline{i})) \\
&- (P_{i_k}V^*_{\beta,(i_{k-1}i_{k-2}\cdots i_1 0)})(l+z(\underline{i})) - \frac{1-b}{\beta}\}
\end{aligned}
\tag{4}
$$

Firstly, by virtue of Equations (2) and (3), we have

$$
V^*_{\beta,(\underline{i})}(l+z(\underline{i})) = V^*_{\beta,\underline{0}}(l+k) \tag{5}
$$

Secondly, the distribution of current queue length given an observed queue length of $(l+z(\underline{i}))$ and indicator string $\underline{i}$ is the same as that given an observed queue length of $(l+k)$ and indicator string $\underline{0}$; hence, $(P_{\underline{i}}\underline{b})((l+z(\underline{i})) = (P_{\underline{0}}\underline{b})(l+k)$.

Next we consider the term $(P_{i_k}V^*_{\beta,(i_{k-1}i_{k-2}\cdots i_1 0)})(l+z(\underline{i}))$.

Case $i_k = 1$

By the definition of P_1, we have

$$
\begin{aligned}
&(P_1 V^*_{\beta,(i_{k-1}i_{k-2}\cdots i_1 0)})(l+z(\underline{i})) \\
&= \mu V^*_{\beta,(i_{k-1}i_{k-2}\cdots i_1 0)}(l+z(\underline{i})) + \overline{\mu} V^*_{\beta,(i_{k-1}i_{k-2}\cdots i_1 0)}(l+z(\underline{i})+1) \\
&= \mu V^*_{\beta,\underline{0}}(l+k-1) + \overline{\mu} V^*_{\beta,\underline{0}}(l+k) \qquad \text{(using Equation (3))} \\
&= (P_1 V^*_{\beta,\underline{0}})(l+k-1)
\end{aligned}
$$

and similarly,

$$
\begin{aligned}
(P_1 V^*_{\beta,(i_{k-1}i_{k-2}\cdots i_1 1)})(l+z(\underline{i})) = \\
(P_1 V^*_{\beta,\underline{0}})(l+k)
\end{aligned}
$$

Case $i_k = 0$

In this case we have

$$
\begin{aligned}
&(P_0 V^*_{\beta,(i_{k-1}i_{k-2}\cdots i_1 0)})(l+z(\underline{i})) \\
&= \mu V^*_{\beta,(i_{k-1}i_{k-2}\cdots i_1 0)}(l+z(\underline{i})-1) + \overline{\mu} V^*_{\beta,(i_{k-1}i_{k-2}\cdots i_1 0)}(l+z(\underline{i})) \\
&= \mu V^*_{\beta,\underline{0}}(l+k-1) + \overline{\mu} V^*_{\beta,\underline{0}}(l+k) \qquad \text{(using Equation (3))} \\
&= (P_1 V^*_{\beta,\underline{0}})(l+k-1)
\end{aligned}
$$

and correspondingly,

$$
(P_0 V^*_{\beta,(i_{k-1}i_{k-2}\cdots i_1 1)})(l+z(\underline{i})) = (P_1 V^*_{\beta,\underline{0}})(l+k)
$$

Hence, substituting term for term, the equation (4) becomes

$$\begin{aligned}
&V^{*}_{\beta,\underline{0}}(l+k) = \\
&\quad (P_{\underline{0}}\underline{b})(l+k) + \beta(P_1 V^{*}_{\beta,\underline{0}})(l+k-1) \\
&\quad + \beta\lambda \min\{0, (P_1 V^{*}_{\beta,\underline{0}})(l+k) - (P_1 V^{*}_{\beta,\underline{0}})(l+k-1) \\
&\quad - \frac{1-b}{\beta}\}
\end{aligned} \tag{6}$$

Now expanding the terms in equation (6), we have, for $x \geq k$,

$$\begin{aligned}
V^{*}_{\beta,\underline{0}}(x) \quad = \quad & (P_{\underline{0}}\underline{b})(x) + \beta(P_{(0)} V^{*}_{\beta,\underline{0}})(x) \\
& +\beta\lambda \min\{0, \mu(V^{*}_{\beta,\underline{0}}(x) - V^{*}_{\beta,\underline{0}}(x-1)) + \\
& \overline{\mu}(V^{*}_{\beta,\underline{0}}(x+1) - V^{*}_{\beta,\underline{0}}(x)) - \frac{1-b}{\beta}\}
\end{aligned} \tag{7}$$

This ends the proof. $\square$

Initially, we had 2^k coupled equations. The optimal value function could be viewed as a map from the $(k+1)$–dimensional space $\{0,1\}^k \times \mathcal{N}$ to the real line. The reduction procedure shows that when $x \geq k$, it is possible to fix k elements of the $(k+1)$, and thereby obtain what is essentially a function of *one* variable, viz., x. We emphasize that this simplification is possible only when $x \geq k$. We now gain some insight into the structure of the optimal policy by considering the expression occurring in the $\min\{0, \cdots\}$ term in Equation (7).

5 Bounds for the Optimal Value Function

We shall now study the function $V^{*}_{\beta,\underline{0}}(x)$. Bounds for $V^{*}_{\beta,\underline{0}}(x)$, $x \geq 0$, can be obtained as shown below.

Consider a policy $\tilde{\pi}$ that *never accepts*. Let $V^{\tilde{\pi}}_{\beta,\underline{0}}(x)$ represent the cost corresponding to the state $(x, \underline{0})$ when the policy followed is $\tilde{\pi}$. Clearly, $V^{\tilde{\pi}}_{\beta,\underline{0}}(0) = 0$. We have, for $x \geq 1$,

$$V^{\tilde{\pi}}_{\beta,\underline{0}}(x) = (P_{\underline{0}}\underline{b})(x) + \beta(\overline{\mu} V^{\tilde{\pi}}_{\beta,\underline{0}}(x) + \mu V^{\tilde{\pi}}_{\beta,\underline{0}}(x-1)) \tag{8}$$

Using this recursion along with $V^{\tilde{\pi}}_{\beta,\underline{0}}(0) = 0$, we can *compute* the cost function $V^{\tilde{\pi}}_{\beta,\underline{0}}(x)$ corresponding to the policy $\tilde{\pi}$.

Lemma 5
$V^{\tilde{\pi}}_{\beta,\underline{0}}(x)$ *is convex non–decreasing.*

Proof: It is clear that $V^{\tilde{\pi}}_{\beta,\underline{0}}(x)$ is non–decreasing. To see that $V^{\tilde{\pi}}_{\beta,\underline{0}}(x)$ is convex, we proceed as follows. From Equation (8), we have for $x \geq 1$,

$$V^{\tilde{\pi}}_{\beta,\underline{0}}(x) = \frac{(P_{\underline{0}}\underline{b})(x)}{1-\beta\overline{\mu}} + \frac{\beta\mu}{1-\beta\overline{\mu}} V^{\tilde{\pi}}_{\beta,\underline{0}}(x-1) \tag{9}$$

Now using Equation (9),

$$(V^{\tilde{\pi}}_{\beta,\underline{0}}(x+2) - V^{\tilde{\pi}}_{\beta,\underline{0}}(x+1)) \geq (V^{\tilde{\pi}}_{\beta,\underline{0}}(x+1) - V^{\tilde{\pi}}_{\beta,\underline{0}}(x))$$

iff

$$\begin{aligned} &\frac{(P_{\underline{0}}\underline{b})(x+2) - (P_{\underline{0}}\underline{b})(x+1)}{1-\beta\overline{\mu}} \\ &+\frac{\beta\mu}{1-\beta\overline{\mu}}(V^{\tilde{\pi}}_{\beta,\underline{0}}(x+1) - V^{\tilde{\pi}}_{\beta,\underline{0}}(x)) \\ &\qquad\qquad\qquad \geq \quad (V^{\tilde{\pi}}_{\beta,\underline{0}}(x+1) - V^{\tilde{\pi}}_{\beta,\underline{0}}(x)) \end{aligned}$$

Rearranging terms, the above is equivalent to

$$(V^{\tilde{\pi}}_{\beta,\underline{0}}(x+1) - V^{\tilde{\pi}}_{\beta,\underline{0}}(x)) \leq \frac{(P_{\underline{0}}\underline{b})(x+2) - (P_{\underline{0}}\underline{b})(x+1)}{1-\beta} \tag{10}$$

Thus, in order to prove that $V^{\tilde{\pi}}_{\beta,\underline{0}}(x)$ is convex, we need to check that the condition in Inequality (10) is satisfied for all x. In order to this, we shall be using the following fact, which follows from Lemma 2:
If $\underline{b}(.)$ is convex non–decreasing, then $(P_0\underline{b})(.)$ is also convex non–decreasing. Repeated application of the above fact shows that $(P_{\underline{0}}\underline{b})(.)$ is convex non–decreasing.

Now consider $V^{\tilde{\pi}}_{\beta,\underline{0}}(1) - V^{\tilde{\pi}}_{\beta,\underline{0}}(0)$. We have

$$V^{\tilde{\pi}}_{\beta,\underline{0}}(1) - V^{\tilde{\pi}}_{\beta,\underline{0}}(0) = \frac{(P_{\underline{0}}\underline{b})(1)}{1-\beta\overline{\mu}}$$

Recalling that $(P_{\underline{0}}\underline{b})(0) = 0$, we see that by the convexity of $\underline{b}$,

$$(P_{\underline{0}}\underline{b})(1) - (P_{\underline{0}}\underline{b})(0) \leq (P_{\underline{0}}\underline{b})(2) - (P_{\underline{0}}\underline{b})(1)$$

Therefore,

$$\frac{(P_{\underline{0}}\underline{b})(1) - (P_{\underline{0}}\underline{b})(0)}{1-\beta\overline{\mu}} \leq \frac{(P_{\underline{0}}\underline{b})(2) - (P_{\underline{0}}\underline{b})(1)}{1-\beta}$$

and the statement in Inequality (10) is true for $x = 0$.

Now assume that Inequality (10) is true for $x = (y-1) \geq 0$. Then,

$$\begin{aligned}
V^{\tilde{\pi}}_{\beta,\underline{0}}(y+1) - V^{\tilde{\pi}}_{\beta,\underline{0}}(y) &= \frac{(P_{\underline{0}}\underline{b})(y+1) - (P_{\underline{0}}\underline{b})(y)}{1-\beta\overline{\mu}} \\
&\quad + \frac{\beta\mu}{1-\beta\overline{\mu}}(V^{\tilde{\pi}}_{\beta,\underline{0}}(y) - V^{\tilde{\pi}}_{\beta,\underline{0}}(y-1)) \\
&\leq \frac{(P_{\underline{0}}\underline{b})(y+1) - (P_{\underline{0}}\underline{b})(y)}{1-\beta\overline{\mu}} \\
&\quad + \frac{\beta\mu}{1-\beta\overline{\mu}}\frac{(P_{\underline{0}}\underline{b})(y+1) - (P_{\underline{0}}\underline{b})(y)}{1-\beta} \\
&\quad \text{(induction hypothesis)} \\
&= \frac{(P_{\underline{0}}\underline{b})(y+1) - (P_{\underline{0}}\underline{b})(y)}{1-\beta} \\
&\leq \frac{(P_{\underline{0}}\underline{b})(y+2) - (P_{\underline{0}}\underline{b})(y+1)}{1-\beta}
\end{aligned}$$

where we have used the convexity of $\underline{b}(.)$ to obtain the last line. Hence the induction step is complete. □

The following lemma shows that the optimal value function $V^{*}_{\beta,\underline{0}}(x)$ can be bounded above and below in terms of the cost function $V^{\tilde{\pi}}_{\beta,\underline{0}}(x)$. Note that this provides computable upper and lower bounds to the optimal value function for all $x \geq 0$.

Lemma 6
We have, $\forall x \geq 0$,

$$V^{\tilde{\pi}}_{\beta,\underline{0}}(x) - \frac{\lambda(1-b)}{1-\beta} \leq V^{*}_{\beta,\underline{0}}(x) \leq V^{\tilde{\pi}}_{\beta,\underline{0}}(x)$$

Proof: $V^{*}_{\beta,\underline{0}}(x) \leq V^{\tilde{\pi}}_{\beta,\underline{0}}(x)$ follows because $V^{*}_{\beta,\underline{0}}(x)$ is the optimal cost and $\tilde{\pi}$ is a particular policy. To prove $V^{\tilde{\pi}}_{\beta,\underline{0}}(x) - \frac{\lambda(1-b)}{1-\beta} \leq V^{*}_{\beta,\underline{0}}(x)$, we argue as follows. Consider the policy $\tilde{\pi}$ that never accepts. If $\tilde{\pi}$ is followed, the resulting queue length over any step in the horizon will be stochastically the smallest. Hence, for any policy, the expected cost over any step will be lower bounded by

(Expected queue length at step if $\tilde{\pi}$ is followed - $\lambda(1-b)$)

When this is discounted by β and added over all steps, we get $V^{\tilde{\pi}}_{\beta,\underline{0}}(x) - \frac{\lambda(1-b)}{1-\beta}$. Hence we have $V^{\tilde{\pi}}_{\beta,\underline{0}}(x) - \frac{\lambda(1-b)}{1-\beta} \leq V^{*}_{\beta,\underline{0}}(x)$. □

The computable upper and lower bounds to $V^*_{\beta,\underline{0}}(x)$ for all $x \geq 0$ in Lemma 6, enable us to obtain a simple lower bound to $(V^*_{\beta,\underline{0}}(x+1) - V^*_{\beta,\underline{0}}(x))$: $\forall x \geq 0$,

$$\begin{aligned}(V^*_{\beta,\underline{0}}(x+1) - V^*_{\beta,\underline{0}}(x)) &\geq V^{\tilde{\pi}}_{\beta,\underline{0}}(x+1) - \frac{\lambda(1-b)}{1-\beta} - V^{\tilde{\pi}}_{\beta,\underline{0}}(x) \\ &\stackrel{\text{def}}{=} LB(x)\end{aligned}$$

It is immediate from the convexity of $V^{\tilde{\pi}}_{\beta,\underline{0}}(x)$ that $LB(x)$ is non–decreasing.

6 Bounds on the Acceptance Threshold

We show now that the bounds on $V^*_{\beta,\underline{0}}(x)$ (which hold for all $x \geq 0$) can be combined with the reduced DPE given by Equation (7) (which holds for $x \geq k$) to give upper bounds on the thresholds of the optimal policy, when a specific relationship between the parameters is satisfied.

The fact that $LB(x)$ increases with x has the following implication. In equation (7), consider the term $\min\{0, \cdots\}$ on the right hand side. We know that for $x \geq k$,

$$\begin{aligned}&\mu(V^*_{\beta,\underline{0}}(x) - V^*_{\beta,\underline{0}}(x-1)) + \\ &\overline{\mu}(V^*_{\beta,\underline{0}}(x+1) - V^*_{\beta,\underline{0}}(x)) \\ &\quad\geq \mu LB(x-1) + \overline{\mu} LB(x)\end{aligned}$$

Now $(\mu LB(x-1) + \overline{\mu} LB(x) - \frac{1-b}{\beta})$ increases with x. Therefore, if this quantity becomes positive for some $\tilde{x}$, it remains positive for all $x > \tilde{x}$. Thus, if the parameters of the problem, viz., b, β, λ, μ, k are such that $(\mu LB(x-1) + \overline{\mu} LB(x) - \frac{1-b}{\beta}) > 0$ for $x \geq \tilde{x}$, then for all $x \geq \tilde{x}$, we must have

$$\begin{aligned}&\mu(V^*_{\beta,\underline{0}}(x) - V^*_{\beta,\underline{0}}(x-1)) + \\ &\overline{\mu}(V^*_{\beta,\underline{0}}(x+1) - V^*_{\beta,\underline{0}}(x)) - \\ &\quad > \frac{1-b}{\beta}\end{aligned}$$

Hence, for $x \geq \max(k, \tilde{x})$, both the above and equation (7) will hold. Therefore, for all observed queue lengths $x \geq \max(k, \tilde{x})$, the optimal action is 0. We have thus shown the following:

Lemma 7
Suppose b, β, λ, μ and k are such that $(\mu LB(x-1) + \overline{\mu} LB(x) - \frac{1-b}{\beta})$ is positive from some x onwards. Let $\tilde{x}$ be the smallest value of x for which $(\mu LB(x-1) +$

Table 1
Correspondence between observed queue lengths

Obs. queue for seq. $\underline{i}$	Obs. queue for seq. 0^k
$z(\underline{i})$	k
$z(\underline{i})+1$	$(k+1)$
$z(\underline{i})+2$	$(k+2)$
$\cdots$	$\cdots$

$\overline{\mu}LB(x) - \frac{1-b}{\beta}) > 0$. Consider the all–zero indicator sequence. Then, for all observed queue lengths $x \geq \max(k, \tilde{x})$, the optimal action is 0.

In Altman & Stidham ([4]), it has been shown that the optimal policy for the k–step problem has a threshold structure, with a threshold for each indicator sequence $\underline{i}$. Our approach provides an upper bound to the threshold for appropriate values of the parameters. In particular, the upper bound to the threshold corresponding to the all–zero indicator sequence is precisely $\max(k, \tilde{x})$, as shown above.

We have seen that for the indicator sequence of all 0's, the upper bound to the threshold is given by $\max(k, \tilde{x})$. Now recall the fact that the multi–dimensional DPE for different indicator sequences can be reduced to the same unidimensional DPE (Equation (7)) for appropriately large observed queue lengths. It is then simple to obtain the upper bound to the threshold for any indicator sequence $\underline{i}$ as follows.

For a given set of parameters, we can first evaluate $\max(k, \tilde{x})$. This will be equal to some entry on the right–hand column of Table 1. The corresponding entry in the left–hand column gives the upper bound for sequence $\underline{i}$. This is given by $z(\underline{i}) + \max(0, \tilde{x} - k)$. Thus, we have shown the following:

Lemma 8
The upper bound to the threshold corresponding to the indicator sequence $\underline{i}$ is given by $z(\underline{i}) + \max(0, \tilde{x} - k)$.

It is immediate from the above that the upper bounds corresponding to *all* indicator sequences that have the same $z(\underline{i})$ are equal. The relationships between the upper bounds for different indicator sequences is also easily seen and summarised below.

Lemma 9
If $\underline{i}$ and $\underline{j}$ are two sequences such that the numbers of zeroes in them differ by $d \geq 0$, then the upper bounds to the thresholds corresponding to them also differ by d.

Our approach provides easily computable upper bounds to the thresholds corresponding to different indicator sequences. From a practical point of view, this will be very useful in the search for an optimal policy as it serves to limit the search space. In fact, since $z(\underline{i}) \in \{0, 1, \cdots, k\}$, it is obvious that there are just $(k+1)$ upper bounds for the entire set of 2^k possible indicator sequences, with the relationships between them being given by Lemma 9. Without the results in Lemma 8 and Lemma 9, it would be extremely difficult to actually compute the optimal policy. This becomes particularly significant as k becomes large.

Finally, we consider the question: under what conditions can we expect the expression $(\mu LB(x-1) + \overline{\mu} LB(x) - \frac{1-b}{\beta})$ to become eventually positive? Recalling that

$$LB(x) = V^{\tilde{\pi}}_{\beta,\underline{0}}(x+1) - \frac{\lambda(1-b)}{1-\beta} - V^{\tilde{\pi}}_{\beta,\underline{0}}(x)$$

we find that the limiting value of $LB(x)$, as x increases to ∞, is governed by $\lim_{x\to\infty}(V^{\tilde{\pi}}_{\beta,\underline{0}}(x+1) - V^{\tilde{\pi}}_{\beta,\underline{0}}(x))$. Now it is clear that

$$\lim_{x\to\infty}(V^{\tilde{\pi}}_{\beta,\underline{0}}(x+1) - V^{\tilde{\pi}}_{\beta,\underline{0}}(x)) = \frac{b}{1-\beta}$$

since the left hand side is the cost of holding 1 extra customer forever. Therefore, we have

$$\lim_{x\to\infty}(\mu LB(x-1) + \overline{\mu} LB(x) - \frac{1-b}{\beta}) = \frac{b}{1-\beta} - \frac{\lambda(1-b)}{1-\beta} - \frac{1-b}{\beta}$$

This enables us to conclude that the limiting value of $(\mu LB(x-1) + \overline{\mu} LB(x) - \frac{1-b}{\beta})$ is positive iff the following condition holds

$$\beta > \frac{1-b}{1-\lambda(1-b)} \tag{11}$$

Inequality (11) gives the simple condition among the parameters of the problem under which computable bounds can be obtained as in Lemmas 8 and 9. Note that the condition given in Inequality (11) does not involve the parameter μ, and, more interestingly, the feedback delay parameter k.

Acknowledgements

The authors are grateful to the referee and the Associate Editor for providing insightful comments.

References

[1] E. Altman and P. Nain, Closed-Loop control with delayed information, Performance Evaluation Review, 20 (1992) 193–204.

[2] E. Altman and G. Koole, Stochastic scheduling games with Markov decision arrival processes, Journal of Computers and Mathematics with Appl., (1993) 141–148.

[3] E. Altman and G. Koole, Control of a random walk with noisy delayed information, Systems and Control Letters, 24 (1995) 207–213.

[4] E. Altman and S. Stidham, Optimality of monotonic policies for two–action Markovian decision processes, with applications to control of queues with delayed information, report UNC/OR/TR–94–2, Univ. of North Carolina, Chapel Hill, 1994.

[5] D. Artiges, Optimal routing into two heterogeneous service stations with delayed information, IEEE Trans. on Autom. Control, 40 (1995) 1234–1236.

[6] B. Hajek, Optimal control of two interacting service stations, IEEE Trans. on Autom. Control, 29 (1984) 491–499.

[7] O. Hernández–Lerma, *Adaptive Markov Control Processes*, Springer–Verlag, 1989.

[8] G. Koole, Optimal repairman assignment in two symmetric maintenance models, European Journal of Operations Research, 82 (1995) 295–301.

[9] P. R. Kumar and P. P. Varaiya, *Stochastic Systems: Estimation, Identification and Adaptive Control*, Prentice–Hall, 1986.

[10] J. Kuri and Anurag Kumar, Optimal control of arrivals to queues with delayed queue length information, in *Proc. of 31st IEEE Conference on Decision and Control*, 1992, 991–992.

[11] J. Kuri and Anurag Kumar, Optimal control of arrivals to queues with delayed queue length information, IEEE Trans. on Autom. Control, 40 (1995) 1444–1450.

[12] Joy Kuri, *Optimal Control Problems in Communication Networks with Information Delays and Quality of Service Constraints*, Ph.D. Thesis, Indian Institute of Science, 1995.

[13] D. Stoyan, *Comparison Methods for Queues and Other Stochastic Models*, John Wiley & Sons, 1983.