JB 82-05

Paper Entitled

"Adaptive Control of Two Competing Queues"

Proceedings of 1983 IEEE INFOCOM
pp. 427-435
1983

ADAPTIVE CONTROL OF TWO COMPETING QUEUES

John S. Baras[1] and Arthur J. Dorsey[2]

[1] Electrical Engineering Department
University of Maryland
College Park, Maryland 20742

[2] IBM-Federal Systems Building
21 Firstfield Road
Gaithersburg, Maryland 20748

ABSTRACT

The adaptive control of two queues competing for the services of a single server is treated. The arrival rates and service rates are considered constant but unknown. Convergence results for certainty-equivalence type, adaptive control schemes are established. Several possible extensions of the results are discussed briefly.

1. INTRODUCTION

Dynamic control of queueing systems is a subject of great interest presently, due to potential applications in performance evaluation and design of computer and communication networks and systems.

Extensive bibliographies and reviews of queueing control models and strategies can be found in Crabil et al [1], Sobel [2], Stidham and Prabhu [3]. Typically, classical queueing theory methods treat static or steady-state models and strategies. Recently examples of dynamic control of queueing systems based on point process models and their relevant theory have been given in studies by Hajek et al [4], Bremaud [5], Roseberg, Varaiya and Walrand [6], and the authors [7] when the queue sizes are not observable.

Quite often in practical applications, the parameter values required in typical stochastic queueing models are not known. For example the arrival rates or the service rates or both may be unknown and should be estimated from current and/or historical data. Thus in practice adaptive control of queues is required, in the sense of designing a control strategy which will be responsive to input and service characteristics as they may change in time. To our knowledge very few adaptive control studies of queueing systems have been undertaken. The present paper analyzes a simple adaptive control problem for two competing queues. The results obtained lead to useful practical strategies and can be interpreted also as parameter sensitivity analysis of optimal control strategies.

The organization of the present paper is as follows. In section 2 we formulate the problem and establish nomenclature. In section 3 our main results on certainty equivalence adaptive control are described. In section 4 a very brief description of a more complete method is given.

2. PROBLEM FORMULATION AND NOTATION

We consider two queues served by the same server in discrete time; thus in computer/communication applications jargon we have "synchronous systems". The time is divided into equal length time slots (which are prespecified), during which arrivals and service completions can occur. We let $t = 0,1,2,\ldots$ be the index of the time slots. The situation is depicted in figure 1 below. We let $x_1(t)$, $x_2(t)$

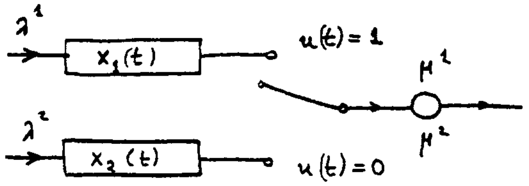

Figure 1. The server time allocation problem

Customers arrive into queues 1 and 2 according to two independent Bernoulli streams with constant rates λ^1, λ^2 respectively. Thus if we let $n_a^i(\cdot)$, $i = 1,2$ denote the two arrival processes,

$$\lambda^i = \Pr\{n_a^i(t) = 1\}, \quad i = 1,2. \tag{2.1}$$

The two queues compete for the services of a server. When the server serves queue i, $i = 1,2$, service completions follow a Bernoulli stream with constant rate μ^i, $i = 1,2$. Thus if we let $n_d^i(\cdot)$, $i = 1,2$, denote the two departure processes, their rates are time varying and depend directly on the queue size. Let $x_i(t)$ be the number of customers in queue i during time slot t, the customer in service (if any) included. The control to be selected concerns the server time allocation to queue 1 or to queue 2. Namely when $u(t) = 1$ and the server completes a services, the next customer to be served comes from queue 1, while if $u(t) = 0$ the next customer comes from queue 2. With this nomenclature the departure rates

$$\mu^i(t,k,v) = \Pr\{n_d^i(t)=1 \,|\, \begin{matrix}\text{past histories of } x_1,x_2 \\ n_a^1,n_a^2 \text{ and u, up to time t}\end{matrix}\}$$

$$\Pr\{n_d^i(t)=1 \,|\, x_i(t)=k,\ u(t)=v\},\ i=1,2, \tag{2.2}$$

take the simple form

$$\mu^1(t,k,v) = \begin{cases} \mu^1 v, & \text{if } k \neq 0 \\ 0, & \text{if } k = 0 \end{cases} \tag{2.3a}$$

$$\mu^2(t,k,v) = \begin{cases} \mu^2(1-v), & \text{if } k \neq 0 \\ 0 & ,\text{ if } k = 0 \end{cases} \tag{2.3b}$$

We assume that during each time slot at most one arrival and one service can occur when each queue operates alone. We further assume that both queues can grow without bound. This is done so that analytical treatment of the problem becomes possible. When the queues are bounded, e.g. due to finite buffer size in computer/communication systems, the methods used here lead to numerical treatment; analytical solutions are no longer achievable. In the latter case if N_i, $i = 1,2$, are the maximum queue sizes for each queue, we have additional constraints.

$$\lambda^i(t,k,v)=\Pr\{n_a^i(t)=1\} \begin{cases} \lambda^i & ,\text{ if } k\neq 0, \text{ all } t,\ v, \\ \\ 0 & ,\text{ if } k=N_i, \text{ all } t,\ v. \end{cases} \tag{2.6}$$

For further results in the case of bounded queues we refer the reader to [8].

The controller has available (for the purposes of selecting a strategy) both the departure and arrival data (slot by slot). Therefore the controller knows at all times the queue sizes. In [7] we studied the partially observed case where the controller had available only arrival data. The difference here is that the controller does not know the values of the parameters λ^i, μ^i, $i = 1, 2$. They have to be estimated on the basis of the observed data $n_a^i(s)$, $n_d^i(s)$, $s<t$, $i = 1, 2$. We assume that λ^i, μ^i, $i = 1,2$, are constant but unknown. We shall consider two cases: Apriori information on these parameters is of the form

$$\begin{aligned} \mu^i &\in M_i \quad , \quad i = 1, 2 \\ \lambda^i &\in \Lambda_i \quad , \quad i = 1, 2 \end{aligned} \tag{2.5}$$

where (a) M_i, Λ_i, are compact intervals or (b) M_i, Λ_i are finite sets.

Service is assumed not to be preemptive and server idling is not allowed. The decision slots, i.e. the slots when control values can change, are either service completion slots or arrival slots when the other queue is empty.

The controller performance is evaluated in this paper using two criteria. The first is the long term average cost

$$J_a = \liminf_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} c(x(k),u(k)) \tag{2.6}$$

where $c(x(k), u(k))$ is the instantaneous cost. The second is the infinite time average discounted cost

$$J_d = E\{\sum_{k=0}^{\infty} \beta^k c(x(k), u(k))\} \tag{2.7}$$

where $\beta\varepsilon(0,1]$ is the discount factor. In the course of the analysis we will use on occasion finite time average cost

$$J_f = E\ \{\sum_{k=0}^{n} c(x(k), u(k))\}. \tag{2.8}$$

In the present paper the instantaneous cost $c(x(k), u(k))$ will be taken as the instantaneous delay

$$c(x(k), u(k)) = c_1x_1(k) + c_2x_2(k) \tag{2.9}$$

where c_1, c_2 are positive constants modelling the relative weight the controller attaches on delays in queue 1 versus those occuring in queue 2.

At each decision slot the controller must assign the value 1 or 0 to the control variable u(t) based on the following information:

$$n_a^i(s),\ s = 0,1,2,\dots,\ t-1,\ i=1,2$$

$$n_d^i(s),\ s = 0,1,2,\dots,\ t-1,\ i=1,2$$

$$u(s),\ s=0,1,2,\dots,\ t-1.$$

To ease notation we shall use the notation

$$\begin{aligned} y^t &= \{n_a^i(s),\ n_d^i(s),\ s=0,1,\dots,t,i=1,2\} \\ u^t &= \{u(s),\ s = 1,\ 2,\ \dots\ ,\ t\}. \end{aligned} \tag{2.10}$$

Thus admissible control strategies γ are sequences of functions

$$\gamma = (g_1,\ g_2,\ \dots\ )\ , \tag{2.11}$$

where

$$u(t) = g_t\ (y^{t-1},\ u^{t-1}) \tag{2.12}$$

and each g_k takes values in $\{0,\ 1\}$. Note that at all times the controller knows the queue sizes, since

$$x_i(t) = n_i^a(t) - n_i^d(t) + x_i(0), \quad i=1,2. \tag{2.13}$$

Our objective is to derive optimal strategies which are adaptive. We consider two methods for analyzing adaptive control problems. The first method is usually known as "certainty-equivalence" type, adaptive control. In this method we first solve the stochastic optimal control problem with known parameters λ^i, μ^i. For the costs considered here the optimal strategy is stationary, i.e. $\gamma = \{g, g, g, \ldots\}$, where g depends on the parameter values; that is $g(x) = g(\theta,x)$, where $\theta = (\lambda^1, \lambda^2, \mu^1, \mu^2)$. At each time t = 1, 2, ... the controller selects by some method, e.g. maximum likelihood, an estimate of the unknown parameters which we denote by

$$\hat{\theta}(t) = (\hat{\lambda}^1(t), \hat{\lambda}^2(t), \hat{\mu}^1(t), \hat{\mu}^2(t)). \tag{2.14}$$

He then uses feedback control

$$u(t) = g(\hat{\theta}(t), x) \tag{2.15}$$

as a good candidate for adaptive control. In this method the analysis is focused on the following questions (problems):

(i) Does $\lim_{t\to\infty} \hat{\theta}(t)$ exist? In what sense? Characterize the limit.

(ii) If θ_0 is the true parameter value does $g(\hat{\theta}(t),x) \to g(\theta_0,x)$ as $t\to\infty$? Estimate the rate of convergence.

(iii) If $\hat{\theta}(t)\to\theta^*$ in some sense, how close is $J(\theta^*)$ (i.e. the cost corresponding to θ^*, which is achieved by the strategy (2.19)), to $J^*(\theta_0)$ (i.e. the optimal cost for the true parameter value)?

The second method treats the adaptive control problem as a stochastic control problem with partial observations. In this method λ^i, μ^i, i = 1, 2, are treated as additional state processes which are unobserved however. Since they are constant they have trivial transition probability matrices. We then apply stochastic control methodology for partially observed Markov chains [7] - [11] and obtain the optimal control strategy. The resulting strategy is of course adaptive by construction.

The consideration of infinite time and long term average costs is crucial for the first method only. The second method can be applied in principle for any cost, provided the problem is well posed. However, there is tremendous difference between the two methods, both in off line and on line computations. The second method, although straightforward, will be unfeasible numerically for the general case. Furthermore the first method suggests simple, practical, implementable control strategies. Finally, convergence rates and estimates cannot be computed in general for nonstationary strategies. An important part of the analysis is the comparison of the "optimal" costs achieved by the two methods in order to assess the first (more intuitive) method versus the global optimal performance provided by the second. Work on this latter subject is in progress and will be reported elsewhere.

3. "CERTAINTY-EQUIVALENCE" ADAPTIVE CONTROL

In this section we analyze fully the adaptive control problem posed in the previous section by the so called "certainty-equivalence" method. For the analysis that follows it is important to display explicitly the probabilistic model for the queues and in particular its dependence on the unknown parameters.

The queueing model is identical to the one developed in [7]; we shall therefore be brief. Due to our assumptions the two queues are independent, modulo the coupling provided by the control strategy. The transition probabilities for each queue modelled as a Markov chain with countable state space $\mathcal{I}$, the nonnegative integers, are given by (see [7] for details):

$$\begin{aligned}
P^i_{j,j}(v) &= \lambda^i\mu^i(j,v) + (1-\lambda^i)(1-\mu^i(j,v)) \\
P^i_{j,j+1}(v) &= \lambda^i(1-\mu^i(j,v)) \\
P^i_{j,j-1}(v) &= (1-\lambda^i)\mu^i(j,v) \\
P^i_{j,k} &= 0 \quad , \text{ elsewhere} \\
& i = 1, 2
\end{aligned} \tag{3.1}$$

where we have suppressed the time argument, since it does not enter explicitly. In view of (2.4) (2.5), letting

$$\begin{aligned}
b^1 &= \lambda^1(1-\mu^1) \ , \ b^2 = (1-\mu^2)\lambda^2 \\
d^1 &= \mu^1(1-\lambda^1) \ , \ d^2 = \mu^2(1-\lambda^2)
\end{aligned} \tag{3.2}$$

(3.1) reduce to

$$P^1(1) = \begin{bmatrix} 1-\lambda_1 & \lambda_1 & 0 & 0 \\ d_1 & 1-b_1-d_1 & b_1 & 0 \\ 0 & d_1 & 1-b_1-d_1 & b_1 \\ & \ddots & \ddots & \ddots \end{bmatrix} \qquad (3.3a)$$

$$P^1(0) = \begin{bmatrix} 1-\lambda_1 & \lambda_1 & 0 & 0 \\ 0 & 1-\lambda_1 & \lambda_1 & 0 \\ 0 & & 1-\lambda_1 & \lambda_1 \\ & & \ddots & \ddots \end{bmatrix} \qquad (3.3b)$$

$$P^2(1) = \begin{bmatrix} 1-\lambda_2 & \lambda_2 & 0 & 0 \\ & 1-\lambda_2 & \lambda_2 & 0 & 0 \\ 0 & & 1-\lambda_2 & \lambda_2 \\ & & \ddots & \ddots \end{bmatrix} \qquad (3.4a)$$

$$P^2(0) = \begin{bmatrix} 1-\lambda_2 & \lambda_2 & 0 & 0 \\ d_2 & 1-b_2-d_2 & b_2 & 0 & 0 \\ 0 & d_2 & 1-b_2-d_2 & b_2 \\ & \ddots & \ddots & \ddots \end{bmatrix} \qquad (3.4b)$$

The transition probability matrix for the Markov chain representing both queues is given by

$$P(v) = P^1(v) \otimes P^2(v) \text{ , for all } v \qquad (3.5)$$

where $\otimes$ indicates matrix tensor product. It is straightforward to establish that for any value of the control variable v (i.e. 0 or 1), P(v) will not be a block diagonal matrix and therefore any state will communicate with any other. That is P(v) is irreducible [12, p. 232] for each value of v. There are no absorbing states for any value of v. We assume that the values of λ^i, μ^i, are such that the Markov chain corresponding to each queue when operated alone consists of a single ergodic class. This condition can be expressed as a stability condition on the matrices $P^1(1)$ in (3.3a) and $P^2(0)$ in (3.4b) above. Thus our assumption becomes

$$\lambda^i/\mu^i<1, \; i =1,2 \; . \qquad (3.6)$$

This assumption is quite weak and will be satisfied by many practical systems. We shall consider only strategies such that

"P(v) in (3.5) corresponds to a single ergodic class" (3.7)

Following the steps described in section 2 we consider first the stochastic control problem when the parameters λ^i, μ^i, i = 1,2 are known. This problem can be thought of as a priority assignment problem in a multiple class queue [1] and has been investigated by several people in the past. To our knowledge all previous work has been directed at the continuous time problem. Thus Cox and Smith [13, p. 77] considered priority assignment in a single server queue with k classes, with arrivals modelled by independent Poisson streams with rates $\lambda^1, \ldots, \lambda^k$. Service times of different classes were modelled by independent random variables with probability distribution function $B_j(\cdot)$ for the j^{th} class. The waiting cost per unit time for class i was denoted by c_i. The cost considered in [13] was average waiting time per unit time (i.e. J_a in equation (2.10)). The admissible strategies were open loop; that is the controller could not use current information, such as queue size, at each decision epoch. The optimal open loop strategy derived in [13] is the so called "μc rule". Specifically if μ^i, i=1,...,k, denotes the inverse of the first moment of $B_i(\cdot)$ (i.e. $\mu^i = \frac{1}{\nu_i}$, where ν_i = average service time for the i^{th} class), optimal priority assignment ranks classes according to the value of the product $\mu^i c_i$, i = 1, 2, ... k; the classes with higher "μc values" given higher priority. Rykov and Lembert [14] and Kakalik [15] proved that the same open loop "μc rule" was optimal even among all feedback rules, which allowed the controller knowledge of queue size at each decision epoch only. That is in [14] dynamic priority assignment was analyzed for average waiting cost per unit time (see Eq. (2.10)), where u(t) could be a function of $x_1(t-1)$, $x_2(t-1)$, ... $x_K(t-1)$ which were assumed to be known to the controller. In [13] - [15] queues without bounds were considered. The rest of the model was identical to that treated in [13]. What is interesting in this result is that the optimal stationary policy does not depend on the arrival rates.

Harrison [16], [17], investigated the same model with the objective of maximizing the expected net present value of service rewards received minus holding costs incurred over an infinite planning horizon with a positive discount factor. He showed that there exists a very special type of priority assignment, called a modified static (i.e., open loop) policy which is optimal among all feedback policies knowning queue sizes. This particular rule assigns an open loop priority ranking, which can be computed explicitly given the system data,

i.e., λ^i, $B_i(\cdot)$, c_i, $i = 1, \ldots K$, via a finite step algorithm [17]. A particularly important feature of this policy is that in the case of two classes it reduces to the "μc rule"; that is it is independent of the arrival rates. However, in a queue with more than three classes, the determination of lower priorities does depend on the arrival rates of higher ranked classes [17]. In [16] - [17] unbounded queues were considered.

Finally, Mova and Ponomarenko [18] analyzed an (M/M/c) : (PRI/N/∞) system. Due to the finite bound on the queue length they demonstrated that the simple "μc rule" is not optimal via a numerical example. The equations characterizing the optimal policy, show that it is a true feedback strategy in the sense that it depends on the current queue size and furthermore on the arrival rates of all classes. Linear programming can be used to compute the optimal policy which is stationary.

In summary, previous work demonstrates that for the two competing queues problem in continuous time (i.e. asynchronous systems) the "μc rule" strategy is optimal for a variety of criteria when the parameters λ^i, μ^i, $i = 1, 2$ are known. This is remarkable and quite useful in practice since several objective functions usually serve as performance measures of a queueing system.

Our first objective in this section is to establish a similar result for the discrete time problem formulated in section 2. We consider first the infinite time discounted average aggregate waiting cost, i.e. Eq. (2.7). To obtain the optimal policy we have to overcome a slight technical problem: the instantaneous cost c(x(k), u(k)) in Eq. (2.12) is not uniformly bounded, because the state space is $\mathcal{I} \times \mathcal{I}$. We shall use results of Lippman [19] which allow polynomial (in the state) growth of the instantaneous cost. In the current model (see section 2) only a finite number of states are accessible from each state in one transition (see in particular (3.1) - (3.5)). Specifically from the state (i_1, i_2) only the states (i_1, i_2), $(i_1 + 1, i_2)$, $(i_1 - 1, i_2)$, $(i_1 - 1, i_2)$, $(i_1, i_2 + 1)$, $(i_1, i_2 - 1)$, $(i_1 + 1, i_2 + 1)$, $(i_1 + 1, i_2 - 1)$ and $(i_1 - 1, i_2 + 1)$ are accessible. Thus the modification (1′) [19, p. 720] of assumption (1) [19, p. 71] holds here. Assumption 1 [19, p. 718] is satisfied here with m = 1 and Assumption 2 [19, p. 719] is satisfied trivially here since only one arrival and one departure can occur in each queue during each time slot. A slight modification of the arguments in [19, Theorem 1] for discrete time, establishes Denardo's [20] N-stage contraction assumption in an appropriate metric space with weighted sup metric. These arguments provide the ideas behind the proof of the following results since the action set here is finite (i.e., the set {0, 1}).

<u>Theorem 1</u>: For $\beta\varepsilon(0,1)$ there is an optimal stationary policy γ_d^* for the infinite horizon discounted average aggregate delay problem. The optimal policy and optimal cost V^* are determined from the stationary Bellman functional equation

$$V_d^*\ (i_1, i_2) = c_1 i_1 + c_2 i_2 + \beta \min_{v\varepsilon\{0,1\}} \sum_{j_1,j_2\varepsilon\mathcal{I}} P_{i_1,i_2;j_1,j_2}(v)\ V_d^*\ (j_1,j_2) \qquad \text{for all } i_1, i_2 \varepsilon\ \mathcal{I}\ , \tag{3.8}$$

where P(v) is the transition probability matrix described by (3.1) - (3.5). Moreover (3.8) has a unique solution.

Let $g_d^*\ (\cdot,\cdot)$ be the function : $\mathcal{I} \times \mathcal{I} \rightarrow \{0,1\}$ that defines the optimal policy. That is

$$\gamma_{d,\beta}^* = (g_{d,\beta}^*, g_{d,\beta}^*, g_{d,\beta}^*, \ldots) \ . \tag{3.9}$$

It is determined implicitly from (3.8) via

$$g_d^*\ (i_1,i_2) = \arg\{\min_{v\varepsilon\{0,1\}} \sum_{j_1,j_2\varepsilon\mathcal{I}} P_{i_1,i_2;j_1,j_2}\ V_2^*\ (j_1,j_2)\}. \tag{3.10}$$

We have the following result:

<u>Theorem 2</u>: The optimal rule g_d^* is the "μc rule" and is independent of β. That is, if $\mu^2 c_2 > \mu^1 c_1$

$$g_d^*(i_1,i_2) = \begin{cases} 1 \quad , & \text{when } i_2 = 0 \\ \\ 0 \quad , & \text{when } i_2 \neq 0 \end{cases} \qquad (3.11)$$

while if $\mu^1 c_1 > \mu^2 c_2$

$$g_d^*(i_1,i_2) = \begin{cases} 1 \quad , & \text{when } i_1 \neq 0 \\ \\ 0 \quad , & \text{when } i_1 = 0 \end{cases} \qquad (3.12)$$

The proof of theorem 2 follows from direct manipulation of (3.8) and is omitted.

Remarks: 1. The optimal strategy is open loop. A similar result holds for multiple classes (more than 2), although the proof is more involved.

2. The optimal strategy does not depend on the arrival rates λ^1,λ^2. For more than two classes this is no longer valid.

3. When queues are finite the "μc rule" of theorem 2 is not optimal. These problems can only be treated numerically. We refer to [8] for details.

4. It is clear that both g_d^* and V^* depend implicitly on the unknown parameters λ^i, μ^i, $i=1,2$. More precisely g_d^* depends only on μ^i, $i=1,2$. We shall emphasize this dependence by writing $V^*(\theta,i_1,i_2)$ and $g^*(\theta,i_1,i_2)$, where $\theta = \lambda^1,\lambda^2,\mu^1,\mu^2)$, when discussing adaptive control. Similarly we shall write $P(\theta,v)$ for the transition probability matrix (3.5).

We consider next the average aggregate delay per unit time problem (see (2.6)). Again to overcome the unbounded cost difficulty we use modifications of arguments from Lippman [19] and Kushner [21]. From the discussion in the beginning of this section (see in particular (3.6), (3.7)) for each policy the Markov chain of the problem has states comprising a single ergodic class. Our analysis follows closer Kushner's development [21] with appropriate modifications. The result can be obtained by taking the limit $\beta \to 1$ on a normalized discounted cost following arguments similar to Lippman [19]. Let $\gamma = (g,g,\ldots)$ be a general stationary Markovian policy. Admissible stationary Markovian policies are only those for which $P(g)$ corresponds to a single ergodic class. Here $P(g)$ is the transition probability matrix (3.5) under policy γ. Let $\pi(g)$ be the corresponding row vector of anymptotic probabilities

$$\pi(g) = \pi(g)P(g) \qquad (3.13)$$

Let C denote the vector with components $C(i_1,i_2 = c_1 i_1 + c_2 i_2$. Let $\tilde{I}(i_1,i_2) = 1$ and define

$$\Delta(v) = \sum_{n=0}^{\infty} (P^{(n)} - \tilde{I}\pi(v))C \qquad (3.14)$$

and

$$\Gamma(v) = \sum_{j_1,j_2 \in \mathcal{I}} \pi_{j_1,j_2}(v)\,(c_1 j_1 + c_2 j_2)\ . \qquad (3.15)$$

By definition $\Gamma(v)$ is the limit of $\sum_{j_1,j_2 \in \mathcal{I}} P^{(n)}_{i_1,i_2;j_1,j_2} (c_1 j_1 + c_2 j_2)$ as $n \to \infty$. We can, with this notation, state our next result.

Theorem 3: There is an optimal stationary policy γ_a^* for the average aggregate delay per unit time problem. The optimal policy and optimal cost V_a^* are determined from the functional equation

$$\Delta(i_1,i_2) + V_a^*(i_1,i_2) = c_1 i_1 + c_2 i_2 +$$

$$+ \min_{v \in \{0,1\}} \sum_{j_1,j_2 \in \mathcal{I}} P_{i_1,i_2;j_1,j_2} \Delta(j_1,j_2), \qquad (3.16)$$

for all $i_1,\ i_2 \in \mathcal{I}$

The proof of theorem 3 involves modifications of arguments from [19] [21]. Again optimality is in the sense of nonanticipative policies [21, p. 158-159]. Let $g^*(\cdot,\cdot)$ be the function defining the optimal strategy γ_a^* of theorem 3. We have then the following result.

Theorem 4: The optimal rule g_a^* is the "μc rule" described by (3.11), (3.12).

We note that an alternate expression for V_a^* is

$$V_a^* = \sum_{i_1,i_2 \in \mathcal{I}} \pi(g_a^*,i_1,i_2)\,(c_1 i_1 + c_2 i_2) \qquad (3.17)$$

where $\pi(g_a^*)$ is the asymptotic probability vector corresponding to the "μc rule".

Remarks: 1. Optimal strategy is open loop.

2. Optimal strategy does not depend on $\lambda^1,\ \lambda^2$.

3. Results are valid for more classes.

4. Finite queues can only be treated numerically (see [8]).

5. Again we explicitly denote dependence on parameters by writing

$V^*_a(\theta,i_1,i_2)$, $g^*_a(\theta,i_1,i_2)$ and $P(\theta,v)$.

To complete the certainty equivalence adaptive control scheme we need only to estimate μ^1, μ^2 since the optimal strategy with known parameters does not depend on λ^1, λ^2. We use maximum likelihood estimates of μ^1, μ^2 based on the observations $n^i_d(s)$, $s<t$, $i = 1,2$. Namely let $\mu = (\mu^1, \mu^2)$ and

$$\ell(t,\mu) = \frac{Pr(x(t),\ x(t+1);\ u(t),\ \mu)}{Pr(x(t),\ x(t+1);\ u(t),\ \mu_o} \tag{3.18}$$

where μ_0 is the true value. Then if

$$L(t,\mu) = \prod_{s=0}^{t-1} \ell(s,\mu) \tag{3.19}$$

the max likelihood estimate is given by

$$\hat\mu(t) = \arg\max L(t,\mu). \tag{3.20}$$

We obtain then the following result in the case M_i (see section 2) i=1,2 are finite sets.

<u>Theorem 5</u>: $\hat\mu^1(t)$, $\hat\mu^2(t)$ converge to μ^{1^*}, μ^{2^*} with probability 1 as t→∞. Furthermore if $g^*(\mu,i_1,i_2)$ denotes the optimal feedback strategy with parameter μ

$$P_{i_1,i_2;j_1,j_2}(g^*(\mu^*,i_1,i_2),\mu^*) = P_{i_1,i_2;j_1,j_2}(g^*(\mu^*,i_1,i_2),\mu_0). \tag{3.21}$$

The proof of theorem 5 uses modified and simpler arguments than [22] due to the special structure of our model. In particular note that the "μc rule" (3.11) (3.12) is open loop. Thus (3.21) in view of (3.1)-(3.5) implies

$$\left.\begin{array}{l}\hat\mu^1(t) \to \mu^1_0 \\ \\ \hat\mu^2(t) \to \mu^2_0\end{array}\right\} \text{ as } t\to\infty \text{ w.p.l.} \tag{3.22}$$

For example applying (3.21) with $i_1 = 1$, $i_2 = 1$, $j_1 = 0$, $j_2 = 2$ we get

$$\mu^{1^*}(1-\lambda^1)\lambda^2 = \mu^1(1-\lambda^1)\lambda^2$$

for any λ^1, λ^2 which implies $\mu^{1^*} = \mu^1_0$.

<u>Corollary 1</u>: The control values $u(t) = g^*(\hat\mu(t), i_1,i_2)$ converge a.s. as t→∞ to the values $u(t) = g^*(\mu_0,i_1,i_2)$.

In view of the special structure of the problem the likelihood functions (3.18) (3.19) and the maximum likelihood estimators $\hat\mu^1(t)$, $\hat\mu^2(t)$ obtain a particularly nice form. Indeed [12]

$$\hat\mu^1(t) = \frac{\sum_{s=1}^{t} n^1_d(s)}{\sum_{s=1}^{t} u(s)} \tag{3.23}$$

$$\hat\mu^2(t) = \frac{\sum_{s=1}^{t} n^2_d(s)}{\sum_{s=1}^{t} (1-u(s))} \tag{3.24}$$

In case the denominators are zero we keep the previous estimate. Recursive expressions for (3.23) are easy to obtain

$$\hat\mu^1(t+1) = \frac{\sum_{s=1}^{t} n^1_d(s) + n^1_d(t)}{\sum_{s=1}^{t} u(s)+u(t+1)}$$

$$= \frac{\hat\mu^1(t) + n^1_d(t)/U(t)}{1 + \frac{u(t+1)}{U(t)}} \tag{3.25}$$

where

$$U(t+1) = U(t) + u(t+1) \tag{3.26}$$

Similarly

$$\hat\mu^2(t+1) = \frac{\hat\mu^2(t) + n^2_d(t)/V(t)}{1 + \frac{1-u(t+1)}{V(t)}} \tag{3.27a}$$

where

$$V(t+1) = V(t) + 1-u(t+1)\ . \tag{3.27b}$$

Finally the adaptive strategy is

$$u(t) = \begin{cases} 1 \text{ if } x_2(t) = 0, \text{ or if } x_1(t) \neq 0 \\ \qquad \text{and } \hat{\mu}^1(t)c_1 > \hat{\mu}^2(t)c_2 \\ 0 \text{ if } x_1(t) = 0, \text{ or if } x_2(t) \neq 0 \\ \qquad \text{and } \hat{\mu}^1(t)c_1 < \hat{\mu}^2(t)c_2 \end{cases} \qquad (3.28)$$

It is seen that the adaptive strategy changes from time to time. However corollary 1 establishes convergence to the correct policy.

There remains to obtain estimates on the costs obtained by the adaptive strategy. Work on this part of the problem is underway.

4. ADAPTIVE CONTROL AS STOCHASTIC CONTROL WITH PARTIAL OBSERVATIONS

The main utility of the second method is to obtain performance estimates for practical schemes, such as the one studied in section 3. As explained in section 2, we treat λ^1, λ^2, μ^1, μ^2 here as additional states with transition probabilities

$$P_{\lambda_i^1,\lambda_i^1} = 1, \; P_{\lambda_i^1,\lambda_j^1} = 0 \text{ for } i \neq j \qquad (4.1)$$

and similarly for the others. Here $\Lambda_1 = \{\lambda_1^1, \lambda_2^1, \ldots \lambda_K^1\}$. The augumented state is

$$x = \{\underbrace{i_1, i_2,}_{\text{queues}} \quad \underbrace{i_{\lambda_1}, i_{\lambda_2},}_{\text{arrival rate}} \quad \underbrace{i_{\mu_1}, i_{\mu_2}}_{\text{departure rate}}\} . \qquad (4.2)$$

The observed variables are

$$y(t) = \{n_a^1(t),\; n_d^1(t),\; n_a^2(t),\; n_d^2(t)\} . \qquad (4.3)$$

We are thus in the framework of [7] with

$$S_{ij}(t,v,\psi) = \Pr\{x(t+1) = j, y(t) = \psi | x(t) = i, u(t) = v\} , \; i,j, \in \mathbb{I} \times \mathbb{I} \times M_1 \times M_2 \times \Lambda_1 \times \Lambda_2 .$$

One can formulate and solve the infinite horizon discounted cost problem (eq. (2.7)) and average cost per unit time problem (eq. (2.6)) for this semi Markov model. The results allow treatment of finite queues numerically. Work on comparison between the two methods is continuing.

REFERENCES

[1] T.B. Crabill, D. Gross and M.J. Magazine, "A Classified Bibliography of Research on Optimal Design and Control of Queues", Operations Research, Vol. 25, No. 2, March-April 1977, pp. 219-232.

[2] M.J. Sobel, "Optimal Operation of Queues", in Mathematical Methods in Queueing Theory (A.B. Clarke, Edt.), Lecture Notes in Economics and Math. Systems, 98, 1974, pp. 231-261.

[3] S. Stidham, Jr. and N.V. Prabhu, "Optimal Control of Queueing Systems", Ibid, pp. 263-294.

[4] B. Hajek and T. Van Loon, "Decentralized Control of a Multi-Access Broadcast Channel", IEEE Trans. on Aut. Control, Vol. AC-27, No. 3, June 1982, pp. 559-569.

[5] P. Bremaud, "Optimal Thinning of a Point Process", SIAM J. Control and Optim., Vol. 17, No. 2, March 1979, pp. 222-230.

[6] Z. Rosberg, P. Varaiya and J. Walrand, "Optimal Control of Service in Tandem Queues", IEEE Trans. on Autom. Control, Vol. AC-27, No. 3, June 1982, pp. 600-610.

[7] J.S. Baras and A.J. Dorsey, "Stochastic Control of two Partially Observed Competing Queues", IEEE Trans. on Autom. Control, Vol. AC-26, No. 5, Oct. 1981, pp. 1106-1117.

[8] A.J. Dorsey, "Adaptive Control of Simple Queueing Systems", Ph.D. Thesis, Electrical Engineering Department, University of Maryland, College Park, 1982.

[9] P. Varaiya, "Notes on Stochastic Control", Unpublished Class Notes, University of California, Berkeley, Department of Electrical Engineering and Computer Sciences.

[10] R.D. Smallwood and E.J. Sondik, "Optimal Control of Partially Observable Markov Processes over a Finite Horizon", Oper. Res., Vol. 21, No. 5, pp. 1071-1088, 1973.

[11] E.J. Sondik, "The Optimal Control of Partially Observable Markov Processes over the Infinite Horizon: Discounted Costs", Oper. Res. Vol. 26, No. 2, pp. 282-304, 1978.

[12] D.P. Heyman and M.J. Sobel, Stochastic Models in Operations Research, Vol. 1, McGraw-Hill, 1982.

[13] D.R. Cox and W.L. Smith, Queues, London: Methuen, 1961.

[14] V.V. Rykov and E. Ye. Lembert, "Optimal Dynamic Priorities in Single Queueing Systems", Eng. Cybern. Vol 5, No. 1, 1967, pp. 21-30.

[15] J.S. Kakalik, "Optimal Dynamic Operating Policies for a Service Facility", Technical Report 47, Operations Research Center, MIT Cambridge, MA 1969.

[16] J.M. Harrison, "A Priority Queue with Discounted Linear Costs", Operations Research, Vol. 23, No. 2, March-April 1975, pp. 260-269.

[17] ____________, "Dynamic Scheduling of a

Multiclass Queue: Discount Optimality" Operations Research, Vol. 23, No. 2, March-April 1975, pp. 260-269.

[18] V.V. Mova and L.A. Ponomarenko, "On the Optimal Assignment of Priorities Depending on the State of a Servicing System with a Finite Number of Waiting Places", Eng. Cybern., Vol. 12, No. 5, 1974, pp. 66-72.

[19] S.A. Lippman, "Semi-Markov Decision Processes with Unbounded Rewards", Managm. Science, Vol. 19, No. 7, March 1973, pp. 717-731.

[20] E.V. Denardo, "Contraction Mappings in the Theory Underlying Dynamic Programming", SIAM Rev., Vol. 9, 1967, pp. 165-177.

[21] H. Kushner, Introduction to Stochastic Control, Holt, Rinehart and Winston, 1971.

[22] V. Borkar and P. Varaiya, "Adaptive Control of Markov Chains, I: Finite Parameter Set", IEEE Trans. on Autom. Control, pp. 953-957.