A QUEUEING REWARD SYSTEM WITH SEVERAL CUSTOMER CLASSES

Bruce L. Miller

January 1968

P- 3772

ABSTRACT

This paper considers an n-server queueing system with m customer classes distinguished by the reward associated with serving customers of that class. Our objective is to accept or reject customers so as to maximize the expected value of the rewards received over an infinite planning horizon. By making the assumptions of Poisson arrivals and a common exponential service time this problem can be formulated as an infinite horizon continuous time Markov decision problem. In Section 3 we customize the general algorithm for solving continuous time Markov decision problems to our queueing model. In Section 4 we obtain qualitative results about the form of the optimal policy. Section 6 reports the results of simulation tests which compare heuristic policies for our model when the service times associated with each customer class have arbitrary distributions. The "winning" policy is based on a rather intricate theorem whose proof comprises Section 5.

A QUEUEING REWARD SYSTEM WITH SEVERAL CUSTOMER CLASSES

Bruce L. Miller*

The RAND Corporation, Santa Monica, California

1. PROBLEM FORMULATION

We consider an n-server queueing system with m customer classes. A customer class k is distinguished by the reward r_k associated with serving a customer of that class. It is convenient to order the customer classes so that $r_1 > r_2 > \ldots > r_m$ and we assume $r_m > 0$. This assumption makes the analysis "cleaner" and is not restricting as customers with negative rewards would be rejected (see next paragraph) anyway. Customers of class k arrive according to a Poisson rate λ_k and their service time is exponentially distributed with a mean of $1/\mu$, independent of k.

Decisions are made at the instant a customer arrives. We have the choice (if some servers are free) of serving the arriving customer and obtaining the associated reward or of rejecting the customer in order to keep available servers free. We assume there is no backlogging of customers or preemption. Our objective (given in detail later) is to maximize the average number of rewards per unit time the system receives over an infinite planning horizon.

---

*Any views expressed in this paper are those of the author. They should not be interpreted as reflecting the views of The RAND Corporation of the official opinion or policy of any of its governmental or private research sponsors. Papers are reproduced by The RAND Corporation as a courtesy to members of its staff.

Much of the work presented here comes from part of the author's Ph.D. dissertation in the Operations Research Program at Stanford and was supported by the National Science Foundation under grant GP-3739.

We use a scenario to illustrate the basic relationships of the model. In this example there are 2 servers and 2 customer classes defined by $r_1 = 6$, $r_2 = 2$, $\lambda_1 = 1/\text{hour}$, $\lambda_2 = 2/\text{hour}$ and $\mu = 1/\text{hour}$. The initial state at time 0 is that both servers are free.

TIME	EVENT	DECISION	RESULTANT STATE	CUMULATED REWARD
0:21	Class 2 Arrival	Serve	1 Free Server, 1 Busy	2
0:35	Class 1 Arrival	Serve	2 Busy Servers	8
1:17	Service Finished	Does Not Apply	1 Free Server, 1 Busy	8
1:27	Class 2 Arrival	Don't Serve	1 Free Server, 1 Busy	8
1:37	Service Finished	Does Not Apply	2 Free Servers	8
1:39	Class 2 Arrival	Serve	1 Free Server, 1 Busy	10
1:45	Class 2 Arrival	Don't Serve	1 Free Server, 1 Busy	10
1:56	Class 1 Arrival	Serve	2 Busy Servers	16
2:17	Class 1 Arrival	Can't Serve	2 Busy Servers	16

Formulation as a Continuous Time Markov Decision Process

We begin the formulation as a continuous time Markov decision problem by letting the states be 0, 1, ..., n, where being in state i means i servers are free. We let A_i be the set of actions associated with state i where an action is a set of integers indicating which customer classes are to be served. For $i \geq 1$, A_i is the set of all possible subsets of {1, 2, ..., m} while A_0 has one action, the empty set. For example, one action might be the set {2,5,8} and in cases where this action applies customers of classes 2, 5, or 8 would be served and all other types of customers would be rejected.

With each action $a \in A_i$ we associate a reward rate $r(a) = \Sigma_{k \in a} \lambda_k r_k$ which is independent of the state. The reward rate is an expected reward rate. The chosen action a also determines a vector with components $q(j|i,a)$, $0 \leq j \leq n$, having the property that $q(j|i,a) \geq 0$, $j \neq i$, and $\Sigma_j\, q(j|i,a) = 0$. The interpretation of $q(j|i,a)$ is the transition rate into state j from state i using action a. We have

$$q(i+1|i,a) = (n-i)\mu \qquad \text{(servers becoming free)}$$

$$q(i-1|i,a) = \Sigma_{k\epsilon a}\lambda_k \qquad \text{(customers being served)}$$

$$q(i|i,a) = -(n-i)\mu - \Sigma_{k\epsilon a}\lambda_k \qquad \text{(transition rate out of } i\text{)}$$

and $\quad q(j|i,a) = 0 \qquad$ otherwise.

Let $F = X_{i=0}^{n} A_i$. For each decision vector $f \in F$, we associated a reward vector $r(f)$ whose i component is $r(f_i)$ and an infinitesimal generator matrix $Q(f)$ whose (i,j) element is $q(j|i,f_i)$. In this paper we will consider only stationary policies which are necessarily elements of F. This class of policies seems sufficient large since in [6, Theorem 7] it is shown that a stationary policy is optimal over the class of piecewise constant policies in the averaging case. Using the policy f means that when the system is in state i (i servers are free) we will serve an arriving customer of class k if and only if $k\epsilon f_i$.

We let $p_{ij}(t,f)$ be the probability our system is in state j at time t given the system was in state i at time 0 and we are using the policy f, and $P(t,f)$ be the corresponding matrix. It is well known that the probability transition matrix function $P(\cdot,f)$ is given by the solution of the differential equations

$$\frac{d}{dt} P(t,f) = P(t,f)\, Q(f) \tag{1}$$

with the initial condition $P(0,f) = I$.

In the infinite horizon averaging case we seek the $f \in F$ such that the vector of expected average rewards

$$x(f) = \lim T^{-1} \int_0^T P(t,f)\, r(f) \tag{2}$$

is maximized in all coordinates and this is called the optimal policy. This limit exists since $P(t,f)$ converges to a matrix $P^*(f)$ [1, p. 181, Theorem 1].

Related Models

A problem which we do not consider here, but which occurs at some point in the decision making process is how many servers to procure. This problem has been considered for related models by Tainter [7] and Whisler [8]. Both of their models have only one customer class.

2. AN APPLICATION

The author became interested in this class of problems when he worked for an oil company. The oil company rents a fixed number of tank cars to be used along with a pipeline to fill orders for liquid petroleum gas which come from different warehouse locations all over the country. The pipeline represents an unlimited supply method. When a tank car satisfies a request it leaves the supply depot, delivers the gas to the warehouse, and returns to the depot. When a request comes there is a known cost of fulfilling the order by tank car or by pipeline (typically a higher cost) and the decision maker must decide which to use.

This problem can be modeled by letting the tank cars be servers and the different warehouse locations be the customer classes. Shipping by pipeline is equivalent to not serving the customer. The service time is the turnaround time needed to ship the gas to the customer. The reward associated with a warehouse location is the pipeline tariff minus the tank car cost to that warehouse.

Our model represents reality reasonably well with respect to the assumptions of no backlogging and no preemption as well as the reward structure. The assumption of a common exponential service time is unrealistic, and this problem must be handled by an approximation such as those given in Sec. 6.

3. A SOLUTION ALGORITHM

The solution of finite state continuous time Markov decision problems is considered by Howard [2, Chapter 10] and further examined recently by the author [6].

We begin the algorithm by choosing an arbitrary $f \in F$ and calculating vectors $x(f)$ and $y(f)$ from the equations

$$x(f) = P^*(f)\, r(f) \qquad \text{(see eq. (2))}, \text{ and} \tag{3}$$

$$r(f) + Q(f)\, y(f) = x(f) \qquad P^*(f)\, y(f) = 0. \tag{4}$$

The vector $x(f)$ can be interpreted as the vector of steady state rewards (which we are trying to maximize), and $y(f)$ can be interpreted from the equation

$$y(f) = \int_0^\infty (P(t,f) - P^*(f))\, r(f)\, dt. \tag{5}$$

Then for each $a \in A_i$ consider the following inequalities:

$$Q_i(a)\, x(f) \geq 0\ , \tag{6}$$

$$r(a) + Q_i(a)\, y(f) \geq x_i(f) \tag{7}$$

where $Q_i(a)$ is a row vector. Let $G(f,i)$ be the set of actions a such that (6) holds strictly or that (6) holds with equality and (7)

holds strictly. If $G(f,i)$ is empty for all i then f is optimal, i.e., $x(f) \geq x(g)$ for all $g \in F$. If not we obtain a new policy g by setting $g_i = f_i$ if $G(f,i)$ is empty and $g_i \in G(f,i)$ otherwise, the particular choice being arbitrary. The policy g is then substituted for f, and we continue by calculating $x(f)$ and $y(f)$, etc. This algorithm is finite since it has been shown that no policy recurs.

Specialization to the Queueing Model

For our queueing reward model, we are able to restrict our attention to a policy set $F' \subset F$ (Theorem 4.6) where $F' = X_{i=0}^{n} A_i'$. The sets A_i' are defined by $A_0' = A_0 = \phi$, and A_i', $i \geq 1$, differs from A_i in that it does not include the action don't serve anyone, the empty set. If $f \in F'$ then all states communicate which implies the elements of $x(f)$ are identical.

If we let $\nabla y_i(f) = y_i(f) - y_{i-1}(f)$, then the first equation of (4) can be written

$$
\begin{aligned}
& \vdots && \\
& && n\mu \nabla y_i(f) = x(f) \\
(8) \qquad & \textstyle\sum_{k \in f_i} \lambda_k (r_k - \nabla y_i(f)) && + (n-i)\mu \nabla y_{i+1}(f) = x(f) \\
& \vdots && \\
& \textstyle\sum_{k \in f_n} \lambda_k (r_k - \nabla y_n(f)) && = x(f)
\end{aligned}
$$

where $x(f)$ is a scalar equal to the identical components of the vector $x(f)$. Since the coefficients of the $\nabla y_i(f)$ form a diagonal matrix, these equations are extremely easy to solve for the unknowns $\nabla y_1(f)$, $\nabla y_2(f)$, ..., $x(f)$. The first and third paragraphs of an example near the beginning of Sec. 5 comprise a numerical example of (8).

All the elements of the vector $x(f)$ are identical so that equation (6) holds with equality for any state and action and is therefore

bypassed. Equation (7) becomes

$$(9) \qquad \Sigma_{k\in a}\lambda_k(r_k-\nabla y_i(f)) \geq \Sigma_{k\in f_i}\lambda_k(r_k-\nabla y_i(f))$$

since $r(a) + Q_i(a)y(f) = \Sigma_{k\in a}\lambda_k(r_k-\nabla y_i(f)) + (n-i)\mu\nabla y_{i+1}(f)$, and from (8) $x(f) = \Sigma_{k\in f_i}\lambda_k(r_k-\nabla y_i(f)) + (n-i)\mu\nabla y_{i+1}(f)$. Hence we can construct the sets $G(f,i)$ from (9) alone. The algorithm for our queueing model then reduces to picking an initial policy f, solving (8), and then checking condition (9). Either f is optimal or we obtain a new policy from (9) and repeat.

4. QUALITATIVE RESULTS

In this section we establish two qualitative results. The first (Theorem 4.4) is that if it is optimal to serve a customer of class k when i servers are free, it is optimal to serve a customer of class k when j servers are free if $j \geq i$. The second (Theorem 4.6) is that if f maximizes $x(f)$ over $f \in F'$ then f maximizes $f \in F$ and we can limit our attention to the set F'. To avoid uninteresting cases we assume there is at least one server and one customer class.

Let f^* be the policy obtained by the algorithm of Sec. 3 and therefore optimal over the set F'. It follows from (9) and the fact that $G(f^*,i)$ is empty that for $i \geq 1$,

$$\Sigma_{k\in f_i^*}\lambda_k[r_k-\nabla y_i(f^*)] = H(\nabla y_i(f^*))$$

where

$$H(x) = \lambda_1(r_1-x) \quad \text{if} \quad x > r_1$$

$$= \Sigma_{k=1}^{m}\lambda_k[r_k-x]^+ \quad \text{if} \quad x \leq r_1.$$

The first equation defining $H(\cdot)$ follows from our requirement that f_i^* is not empty. We note that $H(x)$ is a strictly decreasing function of x for all x.

Lemma 4.1. For any $f \in F'$, $0 < x(f) < \Sigma_{k=1}^{m} \lambda_k r_k$, where $x(f)$ is the common element of the vector $x(f)$.

Proof: From (3) the vector $x(f) = P^*(f)r(f)$. The lemma follows from the inequalities $\lambda_1 r_1 \leq r_i(f) \leq \Sigma_{k=1}^{m} \lambda_k r_k$ $i \geq 1$, $r_o(f)=0$, and $0 < p_{io}^*(f) <$ for all i. The inequalities concerning p_{io}^* follow from the fact that all states communicate for $f \in F'$.

Lemma 4.2. For all i, $i = 1, 2, \ldots, n$, $\nabla y_i(f^*) \geq 0$.

Proof: Assume the contrary, that $\nabla y_j(f^*) < 0$ for some j. We now prove by induction that $\nabla y_\ell(f^*) < 0$ for $j \leq \ell \leq n$. It holds for $\ell = j$ by assumption. Suppose it holds for $j, j+1, \ldots, \ell$. From the $\ell+1$ equation of (8) we have $H(\nabla y_\ell(f^*)) + (n-\ell)\mu \nabla y_{\ell+1}(f^*) = x(f^*)$ or

$$\nabla y_{\ell+1}(f^*) = (x(f^*) - H(\nabla y_\ell(f^*))/(n-\ell)\mu. \tag{10}$$

By the induction hypothesis $\nabla y_\ell(f^*) < 0$ so that $H(\nabla y_\ell(f^*)) > \Sigma_{k=1}^{m} \lambda_k r_k \geq x(f^*)$ from Lemma 4.1 which proves $\nabla y_{\ell+1}(f^*) < 0$ and implies $\nabla y_n(f^*) < 0$. From the $n+1$ equation of (8) we have $x(f^*) = H(\nabla y_n(f^*)) > \Sigma_{k=1}^{m} \lambda_k r_k$ since $\nabla y_n(f^*) < 0$, which contradicts Lemma 4.1.

Let $\nabla^2 y_i(f^*) = \nabla y_i(f^*) - \nabla y_{i-1}(f^*)$.

Lemma 4.3. For any i, $i = 2, 3, \ldots, n$, $\nabla^2 y_i(f^*) \leq 0$.

Proof: Assume the contrary, that $\nabla^2 y_j(f^*) > 0$ for some j. We now prove by induction that $\nabla^2 y_\ell(f^*) > 0$ for $j \leq \ell \leq n$. It holds for $\ell = j$ by assumption. Suppose it holds for $j, j+1, \ldots, \ell$. As in (10)

$$\nabla y_{\ell}(f^*) = (x(f^*) - H(\nabla y_{\ell-1}(f^*)))/(n-\ell+1)\mu$$

(11)

$$\nabla y_{\ell+1}(f^*) = (x(f^*) - H(\nabla y_{\ell}(f^*)))/(n-\ell)\mu \ .$$

By the induction hypothesis $\nabla y_{\ell}(f^*) > \nabla y_{\ell-1}(f^*)$. The function $H(\cdot)$ is strictly decreasing so that (11) implies $\nabla y_{\ell+1}(f^*) > \nabla y_{\ell}(f^*)$ unless $\nabla y_{\ell+1}(f^*) < 0$ and the denominators adversely affect our inequality. However, this case is impossible by Lemma 4.2. Therefore the induction is validated and $\nabla^2 y_n(f^*) > 0$. The n+1 and n equations of (8) are $H(\nabla y_n(f^*)) = x(f^*) = H(\nabla y_{n-1}(f^*)) + \mu \nabla y_n(f^*)$. Since $\nabla y_n(f^*) > \nabla y_{n-1}(f^*)$, $\mu \nabla y_n(f^*) < 0$ which contradicts Lemma 4.2.

Theorem 4.4. The policy f^* has the property that if $k \in f^*_{i-1}$, then $k \in f^*_i$ for $i = 1, 2, \ldots, n$.

Proof: The theorem is nearly an immediate consequence of Lemma 4.3. For $i > 1$, $k \in f^*_{i-1}$ if and only if $k = 1$ or $r_k \geq \nabla y_{i-1}(f^*)$ (compare with the definition of $H(\cdot)$). From Lemma 4.3 $\nabla^2 y_i(f^*) \leq 0$ so that $k \in f^*_{i-1}$, $k \neq 1$, implies $r_k \geq \nabla y_{i-1}(f^*) \geq \nabla y_i(f^*)$ and $k \in f^*_i$. Since $k = 1 \in f^*_i$ this completes the $i > 1$ case. The $i = 1$ case is immediate since $f^*_0 = \phi$.

Lemma 4.5. For any i, $i = 1, 2, \ldots, n$, $\nabla y_i(f^*) \leq r_1$.

Proof: Assume the contrary that for some j, $\nabla y_j(f^*) > r_1$. This implies using Lemma 4.3 that $\nabla y_1(f^*) > r_1$. As in (10) we have $\nabla y_2(f^*) = (x(f^*) - H(\nabla y_1(f^*)))/(n-1)\mu > x(f^*)/(n-1)\mu$. From the first equation of (8), $\nabla y_1(f^*) = x(f^*)/n\mu$. Hence $\nabla y_2(f^*) > \nabla y_1(f^*)$ since $x(f^*) > 0$ from Lemma 4.1. This contradicts Lemma 4.3 and completes the proof.

Theorem 4.6. The policy f^* is optimal over the set F.

Proof: The policy f^* is optimal unless there is some $a \in A_i \setminus A_i'$ such that $a \in G(f,i)$, since $a \notin G(f,i)$ if $a \in A_i'$ by the optimality of f^* over F'. The only $a \in A_i \setminus A_i'$ is the empty set which we now show does not lie in G(f,i). Since all elements of $x(f^*)$ are identical, equation (6) holds with equality for the action "serve no one." From Lemma 4.5 $\nabla y_i(f^*) \leq r_1$ so that $\Sigma_{k\in\phi}\lambda_k(r_k-\nabla y_i(f^*)) = 0 \leq \lambda_1(r_1-\nabla y_i(f^*))$. Since the action $\{1\} \notin G(f,i)$, $\lambda_1(r_1-\nabla y_i(f^*)) \leq \Sigma_{k\in f_i^*}\lambda_k(r_k-\nabla y_i(f^*))$. These two inequalities imply that the action "serve no one" does not lie in G(f,i) and completes the proof.

The finite horizon version of this model is considered in [4] and the analogous result to Theorem 4.4 is obtained for that case. There it is also shown that if it is optimal to serve a customer of class k at time t' when i servers are free it is optimal to serve a customer of class k when i servers are free for all $t \geq t'$.

The result (Theorem 4.4) that we are more eager to serve customers when more servers are free is quite intuitive. However, it does depend on our assumption that an arriving customer desires only one server, as the following example shows.

Example: There are two servers, two classes of customers and $\lambda_1 = \lambda_2 = \mu = 1$. Customer class 1 has the reward structure $r_1(2) = 10$, $r_1(1) = 0$, and $r_1(0) = 0$ where the argument of $r_1(\cdot)$ is the number of servers assigned. Customer class 2 has the reward structure $r_2(2) = 3$, $r_2(1) = 3$, and $r_2(0) = 0$. If we let our policy be that of serving customer class 1 only with 2 servers when 2 servers are free and serving customer class 2 only with 1 server when 1 server

is free, then (4) becomes

$$\begin{aligned} -2y_o(f) + 2y_1(f) &= x_1(f) \\ 3 + y_o(f) - 2y_1(f) + y_2(f) &= x_2(f) \\ 10 + y_o(f) - y_2(f) &= x_3(f) \ . \end{aligned}$$

As before, all elements of x(f) are equal (4 1/3). The vector y(f) satisfies $y_o(f) = y_o(f)$, $y_1(f) = 2\ 1/6 + y_o(f)$, and $y_2(f) = 5\ 2/3 + y_o(f)$. It can be verified that this policy is optimal from (6) and (7), yet a customer of type 2 is served when 1 server is free but not when 2 servers are free, since $3 > y_1(f) - y_o(f) = 2\ 1/6$ and $3 < y_2(f) - y_1(f) = 3\ 1/2$.

5. A FURTHER EXAMINATION OF THE FUNCTIONS $\nabla y_i(\cdot)$

We have seen from Sections 3 and 4 that the actions taken depend entirely on the value of the $\nabla y_i(f)$ which intuitively is the expected cost to the system for having one of the i free servers become busy for a random length of time whose c.d.f. is $1-e^{-\mu t}$ when using the policy f. The main result of this section is Theorem 5.1, which gives us a further interpretation of $\nabla y_i(f)$. It will be shown in Section 6 how this theorem leads to an approximation in the case where service times have arbitrary distributions.

Theorem 5.1. For $f \in F'$, $\nabla y(f) = \int_0^\infty e^{-\mu s} P(s,f^1)(r(f)+Q(f)y(f) - r(f^1) - Q(f^1)y(f))ds$ where f^1 is the policy defined by $f^1_i = f_{i-1}$, the matrix $P(s,f^1)$ is (1-n) by (0-n), and the coordinates of $\nabla y(f)$ run from 1 to n.

Example of Theorem 5.1: In this example there are 2 servers

and 2 classes of customers. The parameters of the problem are $\lambda_1 = \lambda_2 = \mu = 1$, $r_1 = 5$, and $r_2 = 2$. We let f be the policy $f_2 = \{1,2\}$, $f_1 = \{1\}$, and $f_o = \phi$.

The policy f^1 is therefore $f_2^1 = \{1\}$, and $f_1^1 = \phi$. We solve for $P(s,f^1)$ and obtain

$$p_{11}(s,f^1) = 1/2 + 1/2\ e^{-2s},\ p_{12}(s,f^1) = 1/2 - 1/2\ e^{-2s},$$

$$p_{21}(s,f^1) = 1/2 - 1/2\ e^{-2s},\ \text{and}\ p_{22}(s,f^1) = 1/2 + 1/2\ e^{-2s}.$$

The equations of Theorem 5.1 are:

$$\nabla y_1(f) = \int_0^\infty e^{-s}((1/2+1/2e^{-2s})(5-\nabla y_1(f)) + (1/2-1/2e^{-2s})(7-2\nabla y_2(f)-5+\nabla y_2(f)))\ ds$$

$$\nabla y_2(f) = \int_0^\infty e^{-s}((1/2-1/2e^{-2s})(5-\nabla y_1(f)) + (1/2+1/2e^{-2s})(7-2\nabla y_2(f)-5+\nabla y_2(f)))\ ds$$

which when integrated gives us

$$\nabla y_1(f) = 3\ 1/3 - 2/3\ \nabla y_1(f) + 2/3 - 1/3\ \nabla y_2(f)$$

$$\nabla y_2(f) = 5/3 - 1/3\ \nabla y_1(f) + 4/3 - 2/3\ \nabla y_2(f)\ .$$

The solution of these equations is $\nabla y_1(f) = 2\ 1/8$ and $\nabla y_2(f) = 1\ 3/8$.

The equations (8) for this problem are

$$2\ \nabla y_1(f) = x(f)$$

$$5 \qquad -\nabla y_1(f) + \nabla y_2(f) = x(f)$$

$$7 \qquad -2\ \nabla y_2(f) = x(f)$$

and the solution is $x(f) = 4\ 1/4$, $\nabla y_1(f) = 2\ 1/8$, and $\nabla y_2(f) = 1\ 3/8$.

Interpretation of Theorem 5.1.

The interpretation of the theorem is that $\exp(-\mu s)$ represents the probability that the server which became available at time 0 is still unavailable at time s, and the remaining expression of the integrand is the expected loss recorded at time s if that server is unavailable. This loss depends on time only through the transition matrix $P(s,f^1)$. The components of the vector $(r(f)+Q(f)y(f) - r(f^1)-Q(f^1)y(f))$ represent the non-optimality of using action f_{i-1} when in state i rather than action f_i (see (7) of Section 3). If the decision vector f is not optimal, then some of these components could be negative.

Summary of Previous Results

Before beginning the proof of Theorem 5.1, it is necessary to present some results from the theory of continuous time Markov decision processes. Interestingly, results from the finite horizon case as well as the infinite horizon case are needed.

Let π be a piecewise constant policy defined on the interval $[0,t]$, where $\pi(u)\varepsilon F$ is the decision vector which applies at time u when using the policy π. We define the vector function $\psi(\cdot)$ by the differential equations

$$-\frac{d}{du}\psi = r(\pi(u)) + Q(\pi(u))\ \psi \tag{12}$$

$0 \leq u \leq t$, with the terminal condition $\psi(t) = 0$.

From [5, Theorem 2] we have

$$\psi(u) = \int_0^{t-u} P(s,f)\, r(f) \tag{13}$$

if π in (12) is the stationary policy f. As noted in [5], (13) gives us the interpretation of $\psi_i(u)$ as the expected return that will be obtained in time length (t-u) using the policy f when the system begins in state i. We extend modestly the results of [5] for stationary policies and obtain the limiting form of (13) using (3) and (5) of Section 3. This substitution gives us

$$\lim_{t\to\infty}\{\psi(u)-x(f)(t-u)-y(f)\} = 0 \ . \tag{14}$$

From both [5, Eq. (6)] and [3, p. 817] we have that for any two piecewise constant policies π^1, π^2,

$$V(t,\pi^2) - V(t,\pi^1) = \int_0^t P(u,\pi^1)(r(\pi^2(u)) + Q(\pi^2(u))\,\psi(u) - r(\pi^1(u)) - Q(\pi^1(u))\,\psi(u)) \tag{15}$$

where $\psi(\cdot)$ is given by (12) using the policy π^2 and $V(t,\pi) = \int_0^t P(u,\pi)\, r(\pi(u))$, the vector of expected returns up to time t using the policy π.

Outline of the Proof

It will be shown that $\nabla y(f)$ equals $\lim_{t\to\infty}\{V(t,f)-V(t,f)'\}$ where the prime indicates the coordinates go from 0 to n-1 rather than 1 to n. We set $V(t,f)'$ equal to the expected value of the conditioned vector $V(t,f|s)'$ defined as the vector of expected returns up to time t using the policy f conditioned on the fact that one of the

unavailable servers will become available exactly at time s and all other servers follow the exponential distribution as before. After the one server becomes free at time s, it also follows the exponential law. We show (Lemma 5.3) that $V(t,f|s)'$ is equal to $V(t,f^s)$ for all t where the vector $V(t,f^s)$ runs through the states 1, 2, ..., n, and the policy f^s is given by

$$f^s = f^1 \qquad t \leq s$$

$$f^s = f \qquad t \geq s \ .$$

We then compare $V(t,f^s)$ with $V(t,f)$ using (15) and the limiting result (14).

Proof of Lemmas 5.2 and 5.3.

Let $p_{ij}(t,f|s)$ be the probability of being in state j at time t given the system is in state i at time 0, the policy f is used, and one of the (n-i) servers becomes free exactly at time s.

Lemma 5.2. For $t < s$, $0 \leq i, j, \leq n-1$

$$p_{ij}(t,f|s) = p_{i+1,j+1}(t,f^s)$$

and $$p_{i,j+1}(s,f|s) = p_{i+1,j+1}(s,f^s) \ .$$

Proof: We prove the first equation by considering the relevant differential equations of the form (1) for $t < s$

$$\frac{d}{dt} P(t,f|s)' = P(t,f|s)'Q(f|s)' \quad \text{and}$$

$$\frac{d}{dt} P(t,f^s) = P(t,f^s)\, Q(f^s)$$

with initial conditions $P(0,f|s)' = I = P(0,f^s)$. As before the prime indicates coordinates 0 through n-1 and the unprimed notation indicates coordinates 1 through n. In the first case the state n is never reached since one server will not be free until time s and in the other case the state 0 is never reached since $f_1^1 = f_0 = \phi$. Hence both these states can be eliminated from consideration in their respective cases.

The solution to these differential equations is unique so that we will have proved the first part of the lemma if $Q_{ij}(f|s) = Q_{i+1,j+1}(f^s)$, for $0 \leq i,j \leq n-1$. We have

$$Q_{i,i-1}(f|s) = Q_{i+1,i}(f^s) \text{ since } f_i = f_{i+1}^1 = f_{i+1}^s(t), t < s,$$

and $Q_{i,i+1}(f|s) = Q_{i+1,i+2}(f^s)$ since in the conditioned case one of the (n-i) busy servers is certain to remain busy. The other elements of the Q matrices are zero, except for the diagonal elements which satisfy the desired condition since all rows of an infinitesimal generator matrix must sum to zero. The second equation follows by definition of the conditioning from the first equation using continuity considerations which completes the proof.

Lemma 5.3. For all $0 \leq t < \infty$, $V(t,f|s)' = V(t,f^s)$.

Proof: For $t < s$, $V(t,f|s)' = \int_0^t P(u,f|s)'\, r(f) = \int_0^t P(u,f^s)\, r(f^1) = V(t,f^s)$ since Lemma 5.2 implies $P(u,f|s)'\, r(f) = P(u,f^s)\, r(f^1)$ for all $u < s$. For $t \geq s$

$$V_i(t,f|s) = V_i(s,f|s) + \Sigma_j p_{ij}(s,f|s) V_j(t-s,f)$$

and $V_{i+1}(t,f^s) = V_{i+1}(s,f^s) + \Sigma_j\, p_{i+1,j}\,(s,f^s)V_j(t-s,f)$.

From the first part of this lemma $V_i(s,f|s) = V_{i+1}(s,f^s)$, and the second equation of Lemma 5.2 says that $p_{ij}(s,f|s) = p_{i+1,j}(s,f^s)$ which completes the proof.

Proof of Theorem 5.1.

From the representation (5) and the definition of $\nabla y(f)$,

$$\nabla y(f) = \lim_{t\to\infty} \int_0^t [P(u,t)r(f) - P^*(f)r(f)] - [P(u,f)'r(f) - P^*(f)'r(f)]$$

where $P(u,f)$ is an 1-n by 0-n matrix and $P(u,f)'$ is a 0-n-1 by 0-n matrix. By assumption $f\varepsilon F'$ so that all states form one ergodic chain which implies the components of $x(f) = P^*(f)r(f)$ are equal so that

$$\nabla y(f) = \lim_{t\to\infty} \int_0^t P(u,f)r(f) - P(u,f)'r(f)$$

$$= \lim_{t\to\infty} V(t,f) - V(t,f)' \ .$$

We condition the vector $V(t,f)'$ on the time one of its unavailable servers will return, which gives us

$$\nabla y(f) = \lim_{t\to\infty} V(t,f) - \int_0^\infty \mu e^{-\mu s}\, V(t,f|s)'$$

$$= \lim_{t\to\infty} \int_0^\infty \mu e^{-\mu s}\, \{V(t,f) - V(t,f^s)\}$$

using Lemma 5.3,

$$= \lim_{t\to\infty} \int_0^\infty \mu e^{-\mu s} \int_0^s P(u,f^1)(r(f)$$

$$+ Q(f)\psi(u) - r(f^1) - Q(f^1)\psi(u))$$

using (15) where $\psi(\cdot)$ is based on the policy f. Here we have substituted f^1 for $f^s(u), u < s$. The integral stops at s because after s the policies f^s and f are identical, causing the terms in brackets to

cancel. We can substitute for $\psi(u)$ using (14) and the fact that all components of $x(f)$ are identical to obtain

$$\nabla y(f) = \int_0^\infty \mu e^{-\mu s} \int_0^s P(u,f^1)(r(f) + Q(f)y(f)$$

$$- r(f^1)\ Q(f^1)y(f))$$

$$= \int_0^\infty e^{-\mu s}\ P(s,f^1)\ (r(f) + Q(f)y(f)$$

$$-r(f^1) - Q(f^1)y(f))$$

using well-known relationships between a density function and its cumulative distribution function which completes the proof.

6. ARBITRARY SERVICE TIMES

In this section we consider the same model as Section 1, except that instead of requiring that all customer classes be served according to a common exponential distribution we permit a general service distribution with c.d.f. $W_k'(\cdot)$ for customer class k. Unfortunately, it appears that this problem cannot be solved optimally, and two approximating procedures will be presented which are compared by simulation.

The Exponential Approximation

The exponential approximation method consists of approximating the general service problem by an exponential service problem. We then solve the latter by the algorithm of Section 3, and use that answer for the general service problem. Since more of our work involves the exponential problem, we reserve the unprimed notation for that case.

Let λ_k', r_k', and $1/\mu_k'$, $k = 1, \ldots, m$, be the arrival rate, reward, and mean service time associated with customer class k in the general service problem. The parameters of the exponential service problem are obtained from

$$\lambda_k' = \lambda_k'\mu/\mu_k'$$

(16)

$$r_k = r_k'\mu_k'/\mu$$

with μ, the common exponential service rate, arbitrary. The transformations (16) have the virtues that $r_k\mu = r_k'\mu_k'$ and $\lambda_k r_k = \lambda_k' r_k'$.

The exponential problem is then solved to obtain the optimal policy f. Our decision rule for the general service problem is to serve a customer of class k when i servers are free if, and only if, $k \in f_i$. Interestingly, the optimal f will not depend on the value of μ chosen. The value of μ does affect the approximation based on the method of Theorem 5.1, and therefore should be reasonable (some intelligently weighted average of the μ_k').

The Approximation Based on Theorem 5.1.

This approach begins literally where the exponential approach finishes. Instead of accepting the policy f, we turn to Theorem 5.1 and compare the relevant reward r_k' with the expected loss from having one of i free servers unavailable for a random length of time with c.d.f. $W_k'(\cdot)$. This loss we approximate by the i coordinate of

(17)
$$\int_0^\infty (1-W_k'(s))\, P(s,f^1)\, \{r(f) + Q(f)y(f) - r(f^1) - Q(f^1)y(f)\} \text{ where } f_i^1 = f_{i-1}.$$

The $(1-W_k'(s))$ term represents the probability that the server used to serve a customer of class k is still unavailable at time s. The rest of the expression represents the expected loss rate at time s from having the servers unavailable. Unfortunately, this expected loss rate applies to the exponential approximation, rather than the actual general service problem.

In order to keep the computational requirements at a very modest level, we approximate the expected loss rate at time s for coordinate i by the exponential form

$$C_{1i} + C_{2i} \exp(-C_{3i}s) \quad . \tag{18}$$

The vectors C_1, C_2, and C_3 are obtained from our knowledge of the expected loss rate at $s = 0$, as $s \to \infty$, and that for $W_k'(s) = 1-\exp(-\mu s)$ the loss should integrate to $\nabla y_i(f)$. The specific equations are

$$C_1 + C_2 = r(f) + Q(f)y(f) - r(f^1) - Q(f^1)y(f), \text{ and} \tag{19}$$

$$\begin{aligned} C_1 &= P^*(f^1)\{r(f)+Q(f)y(f) - r(f^1)-Q(f^1)y(f)\} \\ &= P^*(f^1)\{x(f)-r(f^1)\} \text{ since } P^*(f^1)Q(f^1) = 0 \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{([6, Theorem 5])} \\ &= x(f) - x(f^1) \text{ since } x(f) \text{ has all coordinates equal.} \end{aligned} \tag{20}$$

Also $\int_0^\infty \exp(-\mu s)\{C_{1i} + C_{2i} \exp(-C_{3i}s)\} = \nabla y_i(f)$, which integrates to

$$C_{1i}/\mu + C_{2i}/(\mu+C_{3i}) = \nabla y_i(f) \quad . \tag{21}$$

The equations (19), (20), and (21) are solved for C_{1i}, C_{2i}, and C_{3i}. If C_{3i} turns out to be negative, the exponential form (18) is invalid and C_{2i} and C_{3i} are set equal to zero, while C_{1i} remains unaltered. We return to (17) and serve a customer of class k when i servers are free if, and only if,

$$r'_k \geq \int_0^\infty (1-W_k(s))(C_{1i}+C_{2i}\exp(-C_{3i}s))$$

$$= C_{1i}/\mu'_k + (C_{2i}/C_{3i})\,(1-W^*_k(-C_{3i}))$$

where $W^*_k(-C_{3i})$ is the moment generating function of $W'_k(\cdot)$ evaluated at $-C_{3i}$.

The policies using the Theorem 5.1 approximation and the exponential approximation will tend to differ in the following way: The Theorem 5.1 method will be more eager to serve customers when the number of free servers i is relatively large and the customer mean service time is relatively short, or when the number of servers i is low and the customer mean serve time is relatively long. The exponential approximation method will be more eager to serve customers in the exact opposite situations. The reason for this is that the Theorem 5.1 method incorporates the effect that the loss per unit time from having a server unavailable is usually an increasing (decreasing) function of time if the state i is high (low), rather than a constant function of time which is the implicit assumption of the exponential method.

A Comparison by Simulation

A total of 500 individual simulation runs were made to compare

the two methods. In each simulation run there were five servers and ten customer classes. The service time for each customer class was Gamma distributed with mean $1/\mu_k$ and shape parameter a_k. Figure 1 details the simulation.

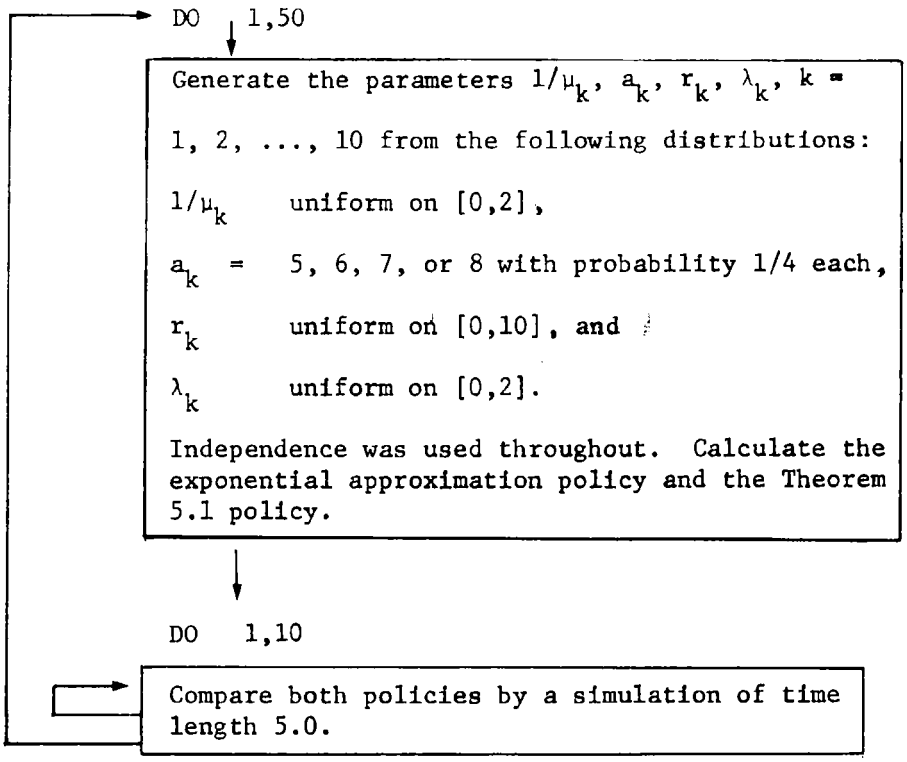

Fig. 1 -- Flow chart of the simulation

Let R_j, j = 1, 2, ..., 50, be the rewards received using the Theorem 5.1 policy minus the rewards received by the exponential policy during the ten simulations for the jth set of parameters. The results of the simulation were that $\bar{R} = \Sigma_j R_j/50 = 10.2$ and $(\Sigma_j (R_j-\bar{R})^2/50\cdot 49)^{\frac{1}{2}} = 3.2$ so that a t-test would give preference to the Theorem 5.1 method with confidence greater than .995.

7. ACKNOWLEDGMENT

It is a pleasure to thank my adviser, Professor Arthur F. Veinott, Jr., for his close examination of my work which resulted in the revision of many of the proofs.

REFERENCES

1. Chung, K. L. (1967), Markov Chains with Stationary Transition Probabilities (2d ed.), Springer, Berlin.

2. Howard, R. A. (1960), Dynamic Programming and Markov Processes, M.I.T. Press, Cambridge, Massachusetts.

3. Martin-Lof, A. (1967), "Optimal Control of a Continuous Time Markov Chain with Periodic Transition Probabilities," Operations Research, Vol. 15, pp. 872-882.

4. Miller, B. (1967), "Finite State Continuous Time Markov Decision Processes with Applications to a Class of Optimization Problems in Queueing Theory," Ph.D. dissertation, Stanford University, Stanford, California, 1967.

5. Miller, B. (1967), "Finite State Continuous Time Markov Decision Processes with a Finite Planning Horizon," The RAND Corporation, P-3569, to appear in the SIAM Journal on Control, Vol. 6, No. 2.

6. Miller, B. (1967), "Finite State Continuous Time Markov Decision Processes with an Infinite Planning Horizon," The RAND Corporation, RM-5425-PR, to appear in the Journal of Mathematical Analysis and Applications.

7. Tainter, M. (1964), "Some Stochastic Inventory Models for Rental Situations," Management Science, Vol. 11, pp. 316-326.

8. Whisler, W. (1967), "A Stochastic Inventory Model for Rented Equipment," Management Science, Vol. 13, pp. 640-648.