International Journal of Operations Research

International Journal of Operations Research Vol. 3, No. 1, 1−6 (2006)

Optimal Selection of Arrival and Service Rates in Tandem Queues

Hsing Luh[1, *] **and M.S. Moustafa**[2]

[1]Department of Mathematical Sciences, National ChengChi University, Taipei 116, Taiwan, R.O.C.

[2]Department of Science, The American University in Cairo, Egypt

Received June 2004; *Revised September 2005*; *Accepted October 2005*

Abstract—We consider n M/M/1 queues in series. At queue one the arrival and service rates are chosen in pair from a finite set whenever there are arrivals or service completions at any queue. Customers arriving to queue k (k = 1, 2, …, n − 1) must go on to queue k + 1 after finishing service at server associated queue k. Customers arriving to queue n leave the system after finishing service at the last server. Arrival and service rates are fixed at queues 2 to n. The objective is to minimize the expected discounted cost of the system over finite and infinite horizons. We show that there is a monotone hysteretic optimal policy in which the arrival and service rates are decreasing and increasing, respectively, in the queue length. In order to establish the result, we formulate the optimal control problem with an equivalent Linear Programming. We believe that many optimal control queueing problems, in which the dynamic programming formulation fails, can be treated successfully via Linear Programming techniques.

Keywords—Queueing networks, Markov decision processes, Stochastic linear programming, Sample-path arguments

1. INTRODUCTION

In last decade, many authors considered the optimal control of systems with more than one server. Rosberg, Varaiya and Walrand (1982) considered two M/M/1 service stations in tandem with uncontrollable arrival process and control of service rate $\mu \in [0, a]$ at first station only. They have showed that the optimal policy is characterized by switching curves. A generalization of the model given by Rosberg, Varaiya, and Walrand (1982) for controlling the service rates in a cycle of m queues has been considered by Weber and Stidham (1987). They also studied the control of arrivals to the first queue only and service rates at each queue of m queues in series.

Hajek (1984) has considered a general two-node model. Nodes 1 and 2 have Poisson arrivals at rates λ_1 and λ_2 respectively. A third stream of Poisson arrivals at rate λ can be routed to either queue. The nodes have fixed exponential service rates μ_1 and μ_2 respectively. There are two additional exponential servers, with rates v_{12} and v_{21}. The first of which serves queue 1 and sends jobs to queue 2. The second of which serves queue 2 and sends jobs to queue 1. Service completions by these servers can be "accepted" or "rejected". The jobs arriving at rate λ must be routed to one of the two nodes. All decisions are made dynamically as a function of the number of jobs in the two queues. Hajek (1984) uses an inductive proof to establish the existence of a monotonic switching curve.

Ghoneim and Stidham (1985) studied two exponential servers in series (with mean service rates μ_1 and μ_2), each with an infinite capacity queue. Arrivals to queue i are from a Poisson process with mean rate λ_i, i = 1, 2. Jobs arriving to queue 1 must go on to queue 2 after finishing service at server 1. Jobs arriving to queue 2 leave the system after finishing service at server 2. They have showed that λ_i are nonincreasing in the number of customers in either queue.

Moustafa (1992) considered two M/M/1 queues in series. At queue 1, the arrival and service rates are chosen in pair from a finite set. He showed numerically that the optimal policy is characterized by a switching curve, but he could not apply the induction proof to construct this structure of the optimal policy. The model considered by Moustafa (1992) is two node version of the model studied by Lu and Serfozo (1984). They used an inductive proof to show that there is a monotone hysteric optimal policy in which the arrival and service rates are decreasing and increasing respectively in queue length. A monotone hysteretic optimal policy defines a set of optimal policies which are specified by a range of queue sizes but its bounds are different when the number of customers is increasing or decreasing. The "hysteresis" refer to decisions that depend both on the current queue length and the current rates. For example, assuming there are only two pairs of arrival and service rates, it is optimal to choose the first pair of arrival and service rates when the queue size decreases to 7 from 10, but to choose the second pair when the queue size increases to 4 from 0. It mean that the bounds of the optimal policy are different when the number of customers is increasing or decreasing.

Typically, the control problem is formulated as a Markovian decision process and the tool of Dynamic

[*]Corresponding author's email: slu@nccu.edu.tw

Programming is used to establish the structure of the optimal policy. However, arguments based on Linear Programming (see Luh and Rieder, 2001) may be used in models where Dynamic Programming technique fails. Bertsimas and Niño -Mora (1999) presented new lower bounds on the achievable cost that emerge as the values of nonlinear programming problems over relaxed formulations of the system's achievable performance region.

In this paper, we use Linear Programming (LP) arguments to generalize the model considered by Moustafa (1992) and establish the structure of the optimal policy. We show that there is a monotone hysteretic optimal policy in which the arrival and service rates are decreasing and increasing, respectively, in the queue length. We formulate the optimal control problem with an equivalent Linear Programming, exploring the structure of optimal policies by studying its dual solutions. Our approaches extend Lu and Serfozo's (1984) results to more general cases and provide powerful new methodologies to optimal load distribution across a network of interconnected stations.

The paper is organized as follows: In section 2 we describe the queueing model of the system. In section 3 we provide the Linear Programming formulation of the optimal control problem. In section 4 we discuss the structure of the optimal policy. Finally, conclusion and future research are given in section 5.

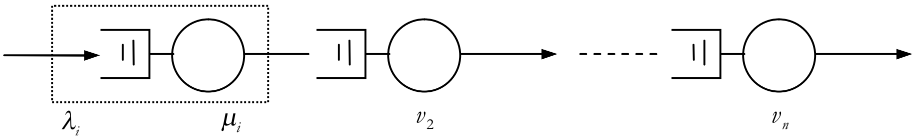

Figure 1. A series of queues in tandem where $(\lambda_i,\ \mu_i)$ is selected at station 1.

2. THE SYSTEM MODEL

We consider n M/M/1 queues in series as depicted in Figure 1. Customers arriving to queue k ($k = 1, 2, \ldots, n-1$) must go on to queue $k + 1$ after finishing service at server k. Customers arriving to queue n leave the system after finishing service at the last server. At the first queue, the arrival and service rates are chosen in pair $(\lambda_i,\ \mu_i)$ from a finite set $\phi = \{(\lambda_1,\ \mu_1), (\lambda_2,\ \mu_2), \ldots, (\lambda_m,\ \mu_m)\}$ with $\lambda_1 \geq \cdots \geq \lambda_m > 0$ and $0 < \mu_1 \leq \cdots \leq \mu_m$.

A switching cost s^i occurs when the i-th pair is substituted by another pair. We assume that s^i is increasing of i and $\lambda_1 + \mu_1, \ldots, \lambda_m + \mu_m$ are positive but not identical. At queues 2 to n, the service rates are fixed to v_j, $j = 2, 3, \ldots, n$. We consider selection decisions that are stationary and nonidling which means decisions have to be made at each decision point. Thus, the rates may be adjusted at decision epoch whenever the system state changes.

Our objective is to minimize the expected holding cost over finite and infinite horizons. Let x_k^l be the number of customers at queue l, $l = 1, 2, \ldots, n$, when the k-th transitions (i.e., arrival or service completion) occurs. However, since there exists m classes of customers at queue 1, x_k^l is a total number of customers with different classes. Now, with abusing a bit of notation, let x_k^i denote the number of customers at queue 1 for different class i, $i = 1, 2, \ldots, m$ but denote the number of customers at queue x_k^i for $i = m + 1, m + 2, \ldots, m + n - 1$. For finite horizon N, we consider the discounting factor $0 < \beta < 1$, the objective is given by

$$G_f = \min_{\phi} E \sum_{k=0}^{N} \beta^k \sum_{i=1}^{m+n-1} (\alpha_i x_k^i) \tag{1}$$

where α_i ($i = 1, 2, \ldots, m + n$) are the holding costs per customer with respect to each i. For infinite horizon, we consider

$$G_i = \lim_{\beta \to 1} (1 - \beta) \min_{\phi} E \left\{ \sum_{k=0}^{\infty} \beta^k \sum_{i=1}^{m+n-1} (\alpha_i x_k^i) \right\} \tag{2}$$

We can use success approximations in (1) and establish these properties for both (1) and (2). It follows from the theory of Markov decision processes that $G_f \to G_i$ as $N \to \infty$ and $\beta \to 1$. See, e.g., Puterman (1994). In this paper, we only consider the case of $\alpha_i = 1$ for all i to concentrate on the methodology itself.

3. LINEAR PROGRAMMING FORMULATION

Now, we describe the LP formulation whose construction is sampled on system at the times corresponding to either arrivals, real service completions, or virtual service completions. Let

$$\Omega = \{ A_1, A_2, \ldots, A_m, D_1, D_2, \ldots, D_m, U_{m+1}, U_{m+2}, \ldots, U_{m+n-1} \}$$

be the set of all transitions. Here A_i and D_i ($i = 1, 2, \ldots, m$) represent the arrival and departure respectively at queue 1 as the i-th pair of arrival and service rates are chosen. For $j = m + 1, m + 2, \ldots, m + n - 1$, U_j represent the event of departure at queue $j - m + 1$. Because we assume no idle between departure at queue $j - 1$ and the arrival at queue j, the event of departure at queue $j - 1$ is equivalent to arrival at queue j. Thus the event of arrival at queue j is not included in Ω.

Let $\omega_k \in \Omega$ represent the k-th transition of the queueing system. Denote by $\Omega^k = \Omega \times \Omega \times \ldots\ldots \times \Omega$ the sample space of the system having k transitions,

$1 \le k \le N$. Define $\omega^k = \omega_1\omega_2...\omega_k \in \Omega^k$ as *k* transitions and $Pr(\omega^k)$ as the probability distribution over Ω^k. Let $\xi_k^i(\omega_k)$ represent the change in the system state incurred by transition ω^k with $1 \le i \le m+n-1$. If $i \le m$, then the transition is either A_i or D_i, which denotes the change at queue 1. If $i > m$, then it denotes the state change at queue $i - m + 1$. The function $\xi_k^i(\omega^k)$ is given by for $i = 1, 2, \ldots, m$ (at the first queue)

$$\xi_k^i(\omega^k) = \begin{cases} 1 & \omega_k = A_i \\ -1 & \omega_k = D_i \quad i = 1, 2, \ ..., m \\ 0 & o.w. \end{cases}$$

It is given at queue 2 that

$$\xi_k^{m+i}(\omega^k) = \begin{cases} 1 & \omega_k = D_i \\ -1 & \omega_k = U_{m+1} \ \ i = 1, 2, \ ..., m \\ 0 & o.w. \end{cases}$$

and for queue $i - m+1$, $m + 1 < i < m + n$,

$$\xi_k^i(\omega^k) = \begin{cases} 1 & \omega_k = U_{i-1} \\ -1 & \omega_k = U_i \\ 0 & o.w. \end{cases}$$

Let z_k^i $(w^{k-1}w_k)$ denote the control variable that represents selection of (λ_i, μ_i) taken at the *k*-th transition instant, $i = 1, 2, \ldots, m$. We define

$$z_k^i(w^k) \in \{0,1\}, w_k = A_j \ \ or \ \ w_k^i = D_j, j = 1, \ 2, \ ..., \ m.$$

When A_j or D_j occurs we set $z_k^i(w^k) = 1$ if the *i*-th pair is chosen at queue 1. For other queues, it is assigned to 1 since we did not specify any control action on transitions at queue *l* ($l = 2, 3, \ldots, n$). This shall be explained in detail in next section. The evolution of the system is described by the following equation:

$$x_{k+1}^i(w^{k+1}) = x_k^i(w^k)\xi_{k+1}^i(w^{k+1})z_{k+1}^i(w^{k+1}),$$
$$i = 1, 2, ..., m+n-1.$$

Note both state variables *x* and control variables *z* are integer-valued. Principally, we shall construct an integer programming for this model. Instead, we derive a relaxed Linear Programming problem where all integer requirements are relaxed. The reason may be found in Luh and Viniotis (2002) in which the optimal solution of a similar formulation has been proved integer-valued. Suppose that the initial queue length is x_0, we may rewrite the cost incurred by policy *z* as

$$\sum_{k=0}^{N-1} \sum_{w^k \in \Omega^k} \Pr(w^k)\beta^\kappa \{ \sum_{i=1}^{m+n-1} x_k^i(w^k) \}$$

$$= (1 + \beta + ... + \beta^N) \sum_{i=0}^{N-1} x_0^i + \sum_{k=1}^{N-1} \sum_{w^k \in \Omega^k} \sum_{i=0}^{m+n-1} \gamma_k^i(w^k) z_k^i(w^k)$$

$$\gamma_k^i(w^k) = \sum_{j=k}^{N-1} \{ \sum_{w_{k+1} \cdot w_j} \Pr(w^k, ..., w_j)\xi_k^i(w^k) + s^i \} \beta^j \tag{3}$$

and *z* must be satisfied with the following three constraints.

(i) Nonnegative queue sizes
For every queue, we have nonnegative queue size constraints as follows.

$$0 \le x_k^i(w^k, z) \overset{\Delta}{=} x_0^i + \sum_{j=1}^{k} \xi_j^i(w^j) z_j^i(w^j) \tag{4}$$

for all *i*, *j*, and *k*.

(ii) Non-idling policy
A non-idling policy means that the server should not be idled whenever the associated queue is not empty. Namely, for all *i*,

$$\text{if } x_k^i(w^k) > 0 \text{ then } z_{k+1}^i(w^k U_i) > 0 \tag{5}$$

(iii) Assignment
For $D_r \in \Omega$ at queue 1, $r = 1, 2, \ldots, m$, set

$$\sum_{i=1}^{m} z_j^i(w^{j-1}D_r) = 1 \tag{6}$$

Since the initial queue sizes are given as parameters in the formulation, constraints (4) may be rearranged to place variables *z* on the left hand side and x_0^i on the right hand side of the inequality. Under the stability condition $\lambda_i < \mu_i$, $i = 1, 2, \ldots, m$, the queue sizes never blow up in a finite time. Therefore, it is easy to check constraints (4)-(6) are bounded and feasible.

Lemma 3.1 *The solution set of constraints (4)-(6) is consistent, and the optimal solution exists.*

Proof. Because $N < \infty$ and (3) is piecewise linear which was proved in Luh and Viniotis (2002), it is immediately clear.

Lemma 3.2 *The optimal solutions of minimizing (3) subjecting to (4)-(6) are integer valued.*

Proof. The proof may be found in Luh and Viniotis (2002). Therefore, we omit it here.

Lemma 3.3 *The optimal solutions $z_1^i(D_r)$ correspond to the optimal policies which assign customers to queue 1 from pair i when D_r occurs, $D_r \in \Omega$.*

Proof: From definition of *z*, we know the possible values of $z_1^i(D_r)$ is either zero or one. In constraints (iii), the optimal solution must have $z_1^i(D_r)=1$ for only one *i*. From (i), it is obvious that the optimal values of $z_1^i(D_r)$, $Dr\in\Omega$, have direct relationship with the assignments associated with pairs (λ_i, μ_i).

4. THE OPTIMAL SELECTION

The linear program (3), with constraints (4), (5), and (6) is the basis for the results we present in this section. To simplify the discussion, we first start to investigate the redundant constraints among (4). Since the state trajectory has to satisfy with (4), we have

$$x_0^i+\sum_{j=1}^{k}\xi_j^i(\omega^j)z_j^i(\omega^j)\geq 0.$$

Consider for $i = 1$ and any $\omega^{k+1}\in\Omega^{k+1}$, (4) becomes

$$x_0^i+\sum_{j=1}^{k}\xi_j^1(\omega^j)z_j^1(\omega^j)+z_{k+1}^1(\omega^k A_j)\geq 0$$
$$\text{if } \omega_{k+1}=A_j,\ j=1,2,\ldots,m,$$

and

$$x_0^1+\sum_{j=1}^{k}\xi_j^1(\omega^j)z_j^1(\omega^j)-z_{k+1}^1(\omega^k D_j)\geq 0$$
$$\text{if } \omega_{k+1}=D_j,\ j=1,2,\ldots,m.$$

That is

$$x_{k+1}^1(\omega^{k+1})=x_k^1(\omega^k)+z_{k+1}^1(\omega^k A_j)\geq 0$$

and

$$x_{k+1}^1(\omega^{k+1})=x_k^1(\omega^k)-z_{k+1}^1(\omega^k D_j)\geq 0$$

Hence, if

$$x_{k+1}^1(\omega^{k+1})\geq 0 \text{ for all } \omega^{k+1}\in\Omega^{k+1}$$

then it implies

$$x_k^1(\omega^k)\geq 0 \text{ for all } \omega^k\in\Omega^k$$

because of $z_{k+1}^1(\omega^k D_j)\geq 0$

By induction from $k = 1$ to $k = N − 1$, we know constraint (4) with $k = 1, 2, \ldots, N − 2$ is redundant if $k = N − 1$ is satisfied by (4). Furthermore, among the constraints (4) with $k = N − 1$, there are some other redundant constraints. Consider for any $\omega^N\in\Omega^N$,

$$x_N^1(\omega^N)=x_{N-1}^1(\omega^{N-1})+\xi_N^1(\omega^N)z_N^1(\omega^N).$$

If we have

$$x_{N-1}^1(\omega^{N-1})-z_N^1(\omega^{N-1}D_j)\geq 0 \text{ for any } 1\leq j\leq m,$$

then it is immediately satisfied with

$$x_{N-1}^1(\omega^{N-1})+z_N^1(\omega^{N-1}D_j)\geq 0$$

because of $z_N^1(\omega^{N-1}D_j)\geq 0$. Therefore, instead of considering $x_N^1(\omega^N)\geq 0$ for every $\omega^N\in\Omega^N$, we consider only

$$x_0^1+\sum_{j=1}^{N-1}\xi_r^1(\omega^r)z_r^1(\omega^r)-z_N^1(\omega^{N-1}D_j)\geq 0 \quad\text{when } \omega_N=D_j,\ j=1,2,\ldots,m. \tag{7}$$

for every $\omega^r\in\Omega^r$, $1\leq r\leq N-1, 1\leq i\leq m$.

In order to characterize the optimal solutions satisfying (4)-(7), its dual variables is introduced. Let dual variables *y* = (y_1, y_2, …, y_N) in which *y*'s components $y_k^i(\omega^k)\geq 0$, i = 1, 2, …, $m + n − 1$, k = 1, 2, …,$N − 1$ be defined associated with constraints (7). Rewrite the LP as

$$\min\sum_{k=1}^{N-1}\sum_{\omega^k\in\Omega^k}\sum_{k=1}^{m+n-1}\gamma_k^i(\omega^k)z_k^i(\omega^k)$$
$$-\sum_{k=1}^{N-1}\sum_{\omega^k\in\Omega^k}\sum_{k=1}^{m+n-1}y_k^i(\omega^k)[x_0^i+\sum_{j=1}^{k}\xi_k^i(\omega^k)z_k^i(\omega^k)]$$

subject to (4) and (6).
The cost function may be rewritten as

$$\sum_{k=1}^{N-1}\sum_{\omega^k\in\Omega^k}\sum_{k=1}^{m+n-1}c_k^i(\omega^k)z_k^i(\omega^k)+C \tag{8}$$

where

$$c_k^i(\omega^k)=\sum_{k=1}^{N-1}\sum_{\omega_{k+1}\ldots\omega_j}\{[Pr(\omega^k,\ldots,\omega_j)\xi_k^i(\omega^k)+s^i]\beta^j - y_j^i(\omega^j)\xi_j^i(\omega^j)\} \tag{9}$$

and C is a constant independent of *z*. In order to find the set of optimal solutions, we apply the complementary slackness theorem, namely, $\bar{z}$ is an optimal solution of (3)-(6) if and only if there exists $\bar{y}$ such that (a), (b) and (c) below hold.

(a) $\bar{z}$ is an optimal solution of (8) subjecting to (5) and (6).

(b) (Feasibility): $x_k^i(\omega^k,z)\geq 0$, for all *i* and *k*.

(c) (Complementary slackness): If $\bar{y}_k^i(\omega^k)$ >0, then $x_k^i(\omega^k, z) = 0$, for all *i* and *k*.

$$\bar{z}_k^i(\omega^k) = \begin{cases} 1 & \text{if } c_k^i(\omega^k) < 0 \\ 0 & \text{if } c_k^i(\omega^k) > 0 \\ \in [0,1] & \text{if } c_k^i(\omega^k) = 0 \end{cases}$$

Clearly, (8) with constraints (6) has become an assignment problem with variables *z* and associated linear cost functions *c* depending on pairs of service and arrival rates as well as the switching cost. Because of $\lambda_1 \geq \ldots \geq \lambda_m > 0$ and $0 < \mu_1 \leq \ldots \leq \mu_m$, if $s_i = 0$, we have

$$c_0^1 \leq \ldots \leq c_0^m.$$

With constraints (6) and discussion above, the optimal solution is clearly decided from pair 1 to pair *m* one by one. In addition, based on sensitivity analysis of (8), every optimal solution of $\bar{z}$ has a correspondent interval associated with its coefficient *c*. When the initial queue size vector is large enough so that all constraints (7) satisfy with strict inequality, it implies $\bar{y}$ = 0 by complementary slackness conditions (c). In this case, *c* is not changed according to (9) and the optimal solution $\bar{z}$ remains the same for $x_0 > L$, with some $L > 0$. It suffices to establish the following lemma.

Lemma 4.1 *The optimal solution will remain the same if the initial queue size is greater than constants* $\bar{L}_\ell$, ℓ = 1, 2, …, *n, for every queue* ℓ.

The proof is omitted because it is clear from the previous statement.

Lemma 4.2 *The optimal solution will remain the same if the initial queue size belongs to the interval* $[\underline{L}_\ell, \bar{L}_\ell]$, ℓ = 1, 2, …, *n, for every queue* ℓ.

Proof. Following Lemma 4.1, we only show x_0^1 shall be greater than a constant $\underline{L}_1$ to satisfy the constraint (6). Consider constraints (6), i.e., for every $\omega^{N-1} \in \Omega^{N-1}$,

$$x_0^1 + \sum_{j=1}^{N-1} \xi_j^1(\omega^j)\bar{z}_j^1(\omega^j) - \bar{z}_N^1(\omega^{N-1}) \geq 0.$$

Suppose $\bar{\omega}_N = \bar{\omega}_1\bar{\omega}_2\ldots\bar{\omega}_{N-1}D_i$ is given as a specific transition path such that $\bar{z}_j^1(\omega^j) \neq 0$, for some *j*, $1 \leq j \leq N$. Since the sum of some $\bar{z}_j(\omega^{j-1}D_i)$ is equal to 1, we have

$$x_0^1 \geq -\sum_{j=1}^{N-1} \xi_j^1(\bar{\omega}^j)\bar{z}_j^1(\bar{\omega}^j) - \bar{z}_N^1(\bar{\omega}_{N-1}D_i).$$

To restrict ℓ only to denote *i* = 1, 2, …, *m*, we have showed that with an optimal policy an interval corresponds to each class *i* of a pair (λ_i, μ_i). Combining with the fact that (8) is piecewise linear, it completes the proof.

Notice that these intervals may be ordered by a monotonic property of (λ_i, μ_i) but there maybe exists overlaps between adjacent intervals, e.g., $\underline{L}_i < \underline{L}_j < \bar{L}_i$, and $\underline{L}_j < \bar{L}_i < \bar{L}_j$, for $j = i + 1$.

Theorem 4.1 *The optimal policy for finite horizon in (1) has a monotone hysteretic structure.*

Proof. From Lemma 4.2, it suffices to show that the optimal policy will be switched from (λ_i, μ_i) to (λ_{i+1}, μ_{i+1}) and switched back from (λ_{i+1}, μ_{i+1}) to (λ_i, μ_i) with different bounds respectively. Since the coefficient of $C_1^i(D_r)$, $D_r \in \Omega$, is a function of λ_i or μ_i depending on the event at the decision epoch, the optimal policy may be assigned when the initial queue size is increasing to $\underline{L}_{i+1}$ or decreasing to $\bar{L}_i$ which depends on the initial queue size and the current service and arrival rates.

By induction on *N*, the results are easy to extend to a case of infinite horizon. We state the following theorem.

Theorem 4.2 *The optimal policy for infinite horizon in (2) has a monotone hysteretic structure.*

Proof. Applying Lippman (1975), Kumar and Meyn (1996), the result of finite horizon is able to extend to the case of infinite horizon.

5. DISCUSSION

The major contribution of this paper lies in the development of a new methodology for studying controlled queueing systems. We strongly believe that the LP based methodology has significantly higher potential than the traditional Dynamic Programming (DP) and Stochastic Dominance techniques. Since the proof is not possible to be provided by the DP approach, the strength of the approach stems from the following facts:

(a) It captures the *essence* of the dynamics and cost structure of the system (linear), without the burden of the (state-dependent) statistical descriptions;
(b) The sample path constraints of the system are conveniently ordered in the constraint matrix.

In general, the way the system parameters affect the parameters of the linear program are the key to the success of this approach. The parameters of the LP are very easily derived from the system parameters.

The second contribution of this paper is the derivation of the structure of the optimal policy in the important queueing problems, namely scheduling policies. With the study that has been done on deriving the necessary conditions of optimality and on reducing the original problem to a trivial assignment problem, the methodology presented in the paper is very general. One can use the approach to develop the structure of optimal policies for other abstract queueing network models as well.

6. CONCLUSIONS AND FURTHER RESEARCH

We have shown that the optimal selection of arrival and service rates at the first queue of *n* queues in series has a monotone hysteretic structure. We applied the linear programming arguments to establish this result. If our assumption of a_i for $i = 1, 2, \ldots, m$ is relaxed to consider more general cases it is clear that c_0^i may not have a monotonous property when the queue length increases, i.e., the monotone hysteretic optimal policies may not exist. However, this should be studied for future research.

We believe that many optimal control queueing problems, in which the dynamic programming formulation fails, can be treated successfully via Linear Programming techniques. It is an effective analysis tool for obtaining performance structure in stochastical optimization problems.

REFERENCES

1. Bertsimas, D. and Niño-Mora, J. (1999). Optimization of multiclass queueing networks with changeover times via the achieveable region approach: part ii, the multi-station case. *Mathematics of Operations Research*, 24(2): 331-361.
2. Ghoneim, H. and Stidham, S. (1985). Optimal control of arrivals to two queues in series. *European Journal of Operations Research*, 21: 399-409.
3. Hajeck, B. (1984). Optimal control of two interacting service stations. *IEEE Transactions on Automatic Control*, AC-29(6): 491-499.
4. Kumar, P.R. and Meyn, S.P. (1996). Duality and linear programs for stability and performance analysis of queueing networks and scheduling policies. *IEEE Transactions on Automatic Control*, 41(1): 4-16.
5. Lippman, S. (1975). Applying a new device in the optimization of exponential queueing systems. *Operations Research*, 23: 687-710.
6. Lu, F.V. and Serfozo, R.F. (1984). M/M/1 queueing decision processes with monotone hysteric optimal policies. *Operations Research*, 1117-1132.
7. Luh, H. and Rieder, U. (2001). Optimal control of arrivals in tandem queues of constant service time. *Mathematical Methods of Operations Research*, 53(3): 481-491.
8. Luh, H. and Viniotis, I. (2002). Threshold control policies for heterogeneous server systems. *Mathematical Methods of Operations Research*, 55(1) :121-142.
9. Moustafa, M.S. (1992). Optimal control of arrival and service rates for two M/M/1 queues in series. *Proceedings of the 1st World Congress of Nonlinear Analysis*, Tampa, Florida, USA, 19-26.
10. Puteramn, M. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley, New York.
11. Rosberg, Z., Varaiya, P., and Walrand, J.C. (1982). Optimal control of service in tandem queues. *IEEE Transactions on Automatic Control*, AC-27(3): 600-610.
12. Weber, R. and Stidham, S. (1987). Optimal control of service rates in networks of queues. *Advances in Applied Probability*, 19: 202-218.