# Technical Research Report

## A Lower Bounding Result for the Optimal Policy in an Adaptive Staffing Problem

*by A. Assad, M. Fu, J-S. Yoo*

**T.R. 98-7**

# A Lower Bounding Result for the Optimal Policy in an Adaptive Staffing Problem


**Arjang A. Assad**
College of Business and Management
University of Maryland
College Park, MD 20742
aassad@mbs.umd.edu
fax: (301) 314-9157

**Michael C. Fu**[1]
College of Business and Management
Institute for Systems Research
University of Maryland
College Park, MD 20742
mfu@umd5.umd.edu
fax: (301) 314-9157

**Jae-sin Yoo**
College of Business and Management
University of Maryland
College Park, MD 20742
jyoo@wam.umd.edu
fax: (301) 314-9157


May 1997


**Abstract**

We derive a lower bound for the staffing levels required to meet a projected load in a retail service facility. We model the queueing system as a Markovian process with non-homogeneous Poisson arrivals. Motivated by an application from the postal services, we assume that the arrival rate is piecewise constant over the time horizon and retain such transient effects as build-up in the system. The optimal staffing decision is formulated as a multiperiod dynamic programming problem where staff is allocated to each time period to minimize the total costs over the horizon. The main result is the derivation of a lower bound on the staffing requirements that is computed by decoupling successive time periods.



**Keywords**: dynamic programming, staffing, service operations



[1]M.C. Fu is supported in part by the National Science Foundation under Grant No. NSF EEC 94-02384.

## 1 Introduction

Queueing theory is frequently used to determine the staffing required to meet a desired level of service. Standard analytical formulas derived for such classical queueing systems as the $M/M/s$ queue are used to see how the desired service characteristics improve as the number of servers $s$ allocated to the system is changed. For example, using the classical $M/M/s$ queue formulas, one can compute the number of servers required to ensure that no more than $100p$ percent of the customers should wait more than $t$ minutes to be served. Studies of applications that use this approach include the following: Foote (1976) and Deutsch and Mabert (1980) for bank tellers; Larson (1972) for emergency telephone line operators; Andrews and Parsons (1989) and Quinn, Andrews, and Parsons (1991) for telephone agents and trunk lines at L.L.Bean; Agnihothri and Taylor (1993) for telephone operators scheduling appointments for hospital patients; and Khan and Callahan (1993) for staffing in an hospital laboratory providing outpatient services.

The present study was motivated by the determination of staffing needs for agents at post office facilities as part of a larger project for the U.S. Postal Service (USPS) in which the first author participated. Typically, a post office has several windows at which postal agents serve customers. The post office can react to changes in the demand (arrivals of customers) by opening or closing windows. Thus, given the level of demand for a specific period, the staffing decision is to allocate the number of servers (for window service) to meet a pre-determined service criterion. The staffing has to respond to changes in the arrival rate of customers by time of the day or the day of the week.

To cite an example, in a large post office, the arrival rate was found to go from a low of approximately 100 customers per hour in the early morning to a peak of 230 customers per hour between 1 to 2 PM, declining back to the low level by 5 PM. Correspondingly, over the course of the same day, the number of agents (clerks) varied between 5 and 9, and the mean service time for customers was 2.3 minutes. The USPS desired the staffing computation to incorporate the effect of the time-varying arrival rate as well as possible build-ups in the facility. Generally, the post office recorded arrival rates over 15 or 30-minute time intervals and had conducted preliminary studies to evaluate the effect of adaptively changing staffing levels every 15 or 30 minutes based on the number of customers in the system (Assad 1992).

As stated by Green, Kolesar, and Svoronos (1991), a standard stationary analysis can be used if the arrival rate fluctuations are mild. However, their experiments with queueing systems with exponential service times and sinusoidal Poisson arrivals show that stationary analysis can lead to significant errors in estimating delays as the amplitude and frequency of the periodic rate increase. This stream of research has continued to evaluate various approximation schemes designed to handle the time-varying arrival rates (Green and Kolesar 1997). Focusing more directly on staffing, Jennings et al. (1996) propose an approximation that can be used to determine the number of servers $s(t)$ as a function of time to meet a target probability of experiencing a nonzero wait in an $M_t/GI/s_t$ system with non-homogeneous Poisson arrivals. In this study, we consider a Poisson arrival process in which the arrival rate $\lambda(t)$ is piecewise constant and specifically consider staffing requirements. In contrast to the sources cited above, we consider the adaptive multi-period problem where the staffing decision depends upon the state of the system at the beginning of each time period. This allows for the model to consider build-up and the spill-over effect of congestion from one period to the next, which the stationary analysis ignores by assuming steady-state conditions in each time period independently. The purpose of this note is to derive a lower bound for the optimal staffing requirements that is computed based on a decoupling of the successive time periods.

# 2 The Dynamic Programming Formulation

## 2.1 Problem Formulation

We seek to determine the optimal staffing policy across multiple time periods in a retail service facility where the total cost function comprises the cost of servers allocated and the customer delay costs over the planning period. The time horizon of length $T$ is divided into $N$ periods indexed by $n$ ($n = 1, ..., N$). Period $n$ has duration $\tau_n$. We let $t_n = t_{n-1} + \tau_n$ with $t_1 = 0$ so that $t_n$ marks the beginning of period $n$.

We assume that customers arrive at the service facility following a Poisson process with arrival rate $\lambda_n$ over time period $n$, and that each customer requires an exponential service time with parameter $\mu$. We define the state of the system at the beginning of period $n$ as the number of customers in the system and denote it by the random variable $X_n$. For convenience, we assume that the system starts in an arbitrary but known state ($X_1 = i_1$). The staffing decision is to allocate $s_n$ servers to period $n$ based on the initial state $X_n$ and the arrival rate for that period (provided as input). If period $n$ starts with $i$ customers and uses $s$ servers, we can calculate the time-averaged mean system size $\overline{m}_n(i; s)$ over time period $n$. We denote the staffing policy by $\mathbf{S}$: it specifies $s_n$ as a function of $X_n$.

The cost incurred in each period has two components. First, using a server incurs a cost of $r$ per person per time unit, reflecting the labor cost rate for each staff person. Next, there is a waiting cost of $d$ per time unit for each person residing in the system. Let $k = r/d$ denote the conversion ratio between unit wait and server costs. The total cost for period $n$ starting in state $i$ is

$$C_n(i;s) = [d\overline{m}_n(i;s) + rs]\tau_n = \overline{c}_n(i;s)d\tau_n,$$

where the cost

$$\overline{c}_n(i;s) = [\overline{m}_n(i;s) + ks]$$

can be interpreted as the rate at which costs are incurred in a period with initial state $i$ and $s$ servers if waiting costs are scaled so that $d = 1$.

Using the preceding notation, we can state the optimization problem as

$$\min_{\mathbf{S}} \sum_{n=1}^{N} E[C_n(X_n; s_n)], \tag{1}$$

with the minimization performed over all staffing policies $\mathbf{S}$, with initial condition $X_1 = i_1$.

Evaluation of the cost function requires computation of the time-averaged quantity $\overline{m}_n(i;s)$. Given the initial state of the system $i$ at the beginning of period $n$ and the number of servers $s$ used, the system evolves as a transient $M/M/s$ queue with the initial state $i$ specified. Using numerical integration, we trace the trajectory of the mean system size, say $m_t(i;s)$, as a function of time over period $n$ to compute $\overline{m}_n(i;s) = \frac{1}{\tau_n}\int_0^{\tau_n} m_t(i;s)dt$. For convenience, henceforth, we shall assume that $\tau_n = \tau$ for all $n$. If all time periods are equal, $C_n$ and $\overline{c}_n$ differ by the same multiplicative factor for all $n$, and the minimization problem (1) can be restated with $C_n$ replacing $\overline{c}_n$.

## 2.2 The Optimality Equation

We use dynamic programming to obtain the optimal adaptive staffing problem. The key observation in formulating the dynamic program is that the transition from $X_n$ to $X_{n+1}$ is governed by Markovian transition probabilities obtained from the transient behavior of the multi-server queue in operation during period $n$. Thus, if the system starts in state $i$ at the beginning of period $n$ and $s$ servers are used during the period, one can compute the transition probability $P_{ij}(s)$ of finding the system in state $j$ at the end of period $n$ or, equivalently, at the beginning of period $n+1$.

Define $f_n^*(i)$ as the optimal expected cost over periods $n$ through $N$ if the system starts period $n$ in state $i$. Let $X(i;s)$ denote the state at the end of a period, given that the period starts in state $i$ and uses $s$ servers. We also define $f_n(i;s) = \overline{c}_n(i;s) + \min_{\widehat{s}} E[f_{n+1}(X(i;s);\widehat{s})]$, which is the total expected cost for stages $n$ through $N$, when the system starts period $n$

with $i$ customers, uses $s$ servers in period $n$, and optimal decisions are made thereafter. By definition of $s^*(\cdot)$, $f_n^*(i) = \min_s\{f_n(i;s)\}$, and the value of $\widehat{s}$ yielding the minimum is $s_{n+1}^*(X(i,s))$. Using backward recursion, we have the optimality equations:

$$f_N^*(i) = \min_s \overline{c}_N(i;s), \tag{2}$$

$$f_n^*(i) = \min_s \left\{\overline{c}_n(i;s) + E\left[f_{n+1}(X(i;s), s_{n+1}^*(X(i;s)))\right]\right\}, n = 1, ..., N-1 \tag{3}$$

$$= \min_s \left\{\overline{c}_n(i;s) + E\left[f_{n+1}^*(X(i;s))\right]\right\}, n = 1, ..., N-1. \tag{4}$$

Thus, the optimal policy is computed as follows:

$$s_N^*(i) = \arg\min_s \overline{c}_N(i;s). \tag{5}$$

$$s_n^*(i) = \arg\min_s \left\{\overline{c}_n(i;s) + \sum_j P_{ij}(s) f_{n+1}^*(j)\right\}, n = N-1, ..., 1. \tag{6}$$

### 2.3 Single-Period Decoupled Staffing Procedure Approximation

This method chooses the number of servers in order to minimize the expected total period cost for each period for a given initial state, i.e.,

$$s_n(i) = \arg\min_s \overline{c}(i;s), \quad n = 1, \cdots, N.$$

Of course, this procedure is myopic in that it ignores the costs of subsequent periods. As in the optimal solution, the policy is a state-dependent vector for each period. Thus, the initial state and transient effects within each period are considered, unlike in steady-state staffing.

## 3 A Lower Bound on the Optimal Policy

Our main result is the following structural property for the optimal policy:

**Theorem** *The single-period staffing provides a lower bound for the optimal staffing policy, i.e., $s_n^*(i) \geq s_n(i)$ for all n=1,$\cdots$, N.*

The proof of the theorem uses a series of lemmas that rely on notions of stochastic ordering. Some of the lemmas can be derived as special cases of more general results found in the sizable literature on stochastic ordering (see for example Stoyan 1983, and Shaked and Shanthikumar 1994). However, to provide a more self-contained exposition, we have relied on a single key proposition, which allows us to derive all the necessary lemmas. We now review the definitions of stochastic ordering (the two definitions are equivalent for the case of random variables), followed by the two propositions used in the proofs of the results that follow.

**Definition**(Ross 1983, p.251)

$X \geq_{st} Y$ ($X$ is stochastically greater than $Y$) if $P(X > a) \geq P(Y > a) \forall a$. Thus, two random variables $X$ and $Y$ are equal in distribution, $X =_{st} Y$, if $P(X > a) = P(Y > a)\ \forall a$.

**Definition** (Ross 1983, p.256)

The random vector $X = (X_1, X_2, \cdots, X_n)$ is stochastically greater than the random vector $Y = (Y_1, Y_2, \cdots, Y_n)$, written $X \geq_{st} Y$, if for all increasing functions $f$, $E[f(X)] \geq E[f(Y)]$. The use of sample path proofs (e.g., Lemma 2) requires the following result, which could also serve as a third definition for stochastic ordering of random variables.

**Proposition 1** (Shaked and Shanthikumar 1994, p.5). $X \geq_{st} Y$ if and only if there exist two random variables $\widehat{X}$ and $\widehat{Y}$ defined on the same probability space such that

$$\widehat{X} =_{st} X, \quad \widehat{Y} =_{st} Y, \quad \text{and} \quad P\{\widehat{X} \geq \widehat{Y}\} = 1.$$

The key stochastic ordering result from the queueing literature that is used in our proofs is the following result:

**Proposition 2** (from Proposition 2 in Shanthikumar and Yao 1989, p.415). In a queueing network of $M$ nodes, with first-come-first-served queue discipline, define the following:

$$\begin{aligned} S_i(n) &= \text{service time of the } n\text{th job to initiate a service at node } i, \\ d_i(n) &= \text{epoch of the } n\text{th job to complete a service at node } i, \\ s_i &= \text{number of servers at node } i. \end{aligned}$$

Then the $n$th service completion epoch at node $i$, $d_i(n)$, is

(a) decreasing in the number of servers ($s_i$), and

(b) increasing in the job service times ($S_i$).

So in particular, if we consider a single-node system and let $D(t)$ denote the number of customer departures by epoch $t$, then the above proposition implies that $D(t)$ is increasing in the number of servers and decreasing in the service times.

**Lemma 1** If $X \geq_{st} Y$ and $f$ is an increasing function, then

(a) $f(X) \geq_{st} f(Y)$ and

(b) $E[f(X)] \geq E[f(Y)]$.

*Proof.*
(a) Letting $f^{-1}(a) = \inf\{x : f(x) > a\}$ for an increasing function $f$, then

$$P\{f(X) > a\} = P\{X > f^{-1}(a)\} \geq P\{Y > f^{-1}(a)\} = P\{f(Y) > a\}.$$

Hence, $f(X) \geq_{st} f(Y)$.
(b) This follows from (a) or the second definition of stochastic ordering. □

The following result can also be established as a special case of Sonderman (1979), Theorem 1.

**Lemma 2** $X(i;s) \geq_{st} X(i;s+1)$ for all $i$, $s$.

*Proof.* We compare two systems, one with $s$ servers and one with $s+1$ servers, given identical interarrival time and service time sequences. Specifically, let $A_1, A_2, A_3, \cdots$ be i.i.d interarrival random variables and $S_1, S_2, S_3, ...$ be i.i.d. service time random variables. For a system with $s$ servers and $\omega = \{A_1, A_2, A_3, ...\} \times \{S_1, S_2, S_3, \cdots\}$, let $N^{(s)}(t, \omega)$ be the number of customers in the system at $t$, $A^{(s)}(t, \omega)$ be the number of customer arrivals by $t$, and $D^{(s)}(t, \omega)$ be the number of customer departures by $t$. Let $N_0 = N^{(s)}(0, \omega)$ be the initial number of customers in the system. Then $X(i;s)$ corresponds to $N^{(s)}(\tau, \omega)$ with $N_0 = i$, where $\tau$ is the length of time period. Then we have for each $\omega$ and all $t$,

$$N^{(s)}(t, \omega) = N_0 + A^{(s)}(t, \omega) - D^{(s)}(t, \omega),$$

$$N^{(s+1)}(t, \omega) = N_0 + A^{(s+1)}(t, \omega) - D^{(s+1)}(t, \omega).$$

Since $A^{(s+1)}(t, \omega) = A^{(s)}(t, \omega)$, we have

$$N^{(s)}(t, \omega) - N^{(s+1)}(t, \omega) = D^{(s+1)}(t, \omega) - D^{(s)}(t, \omega).$$

From Proposition 2 part (a), $D(t)$ is increasing in $s$. Hence, applying Proposition 1, we have $D^{(s+1)}(t) \geq_{st} D^{(s)}(t) \Longrightarrow N^{(s+1)}(t) \leq_{st} N^{(s)}(t) \Longrightarrow X(i;s+1) \leq_{st} X(i;s)$. □

In the following three lemmas, we drop the period subscript $n$ for notational convenience, since a single period is being considered in isolation.

**Lemma 3** $\overline{c}(i;s)$ and $\overline{m}(i;s)$ are increasing in $i$ for all $s$.

*Proof.* Since $\overline{c}(i;s) = \overline{m}(i;s) + ks$, we need to show that $\overline{m}(i;s)$ is increasing in $i$. By definition, $\overline{m}(i;s) = \frac{1}{\tau}\int_0^\tau N(t)dt$. So it suffices to show that $N(t)$ is stochastically increasing in $i$. Let $N_{(i)}(t)$ denote the number of customers in the system at $t$, given that the system starts with $i$ customers at $t = 0$. We must show that $N_{(i)}(t) \leq_{st} N_{(i+1)}(t)$. We give a direct proof based on Proposition 2. Ross (1983) uses a somewhat different argument to establish this result for birth-death processes.

We construct two systems denoted throughout with superscripted "prime" and "double prime" such that $N'(t)$ and $N''(t)$ are equal in distribution to $N_{(i)}(t)$ and $N_{(i+1)}(t)$, respectively, as follows. Both begin empty with identical interarrival time distributions

$$A'_j = A''_j = 0, \quad j = 1, ..., i+1; \quad A_j \text{ i.i.d, } j > i+1,$$

and i.i.d. service time distributions with the single exception that

$$S'_1 = 0,$$

i.e., the first service time is 0 in the primed system.

Since $N(t) = A(t) - D(t)$ and the arrival processes are clearly equal in distribution, the proof reduces to showing

$$D'(t) \geq_{st} D''(t),$$

which follows from Proposition 2, part (b), since $D(t)$ is decreasing in service times and $S'_j \leq_{st} S''_j$. □

**Lemma 4** If $X \geq_{st} Y$, then $\overline{m}(X; s) \geq_{st} \overline{m}(Y; s+1)$ and $E[\overline{c}(X; s)] \geq_{st} E[\overline{c}(Y; s)]$ for all $s$.

*Proof.* By Lemma 3, $\overline{m}(i; s)$ and $\overline{c}(i; s)$ are both increasing in $i$ for all $s$. Hence by Lemma 1,

$$\overline{m}(X; s) \geq_{st} \overline{m}(Y; s) \quad \text{and} \quad E[\overline{c}(X; s)] \geq_{st} E[\overline{c}(Y; s)].$$

In the proof of Lemma 2, we established that $N^{(s+1)}(t) \leq_{st} N^{(s)}(t)$; hence,

$$\overline{m}(Y; s) \geq_{st} \overline{m}(Y; s+1).$$

Combining this with the first inequality above completes the proof. □

**Lemma 5** $\min_s \overline{c}(i; s) \leq \min_s \overline{c}(i+1; s)$ for all $i$ and $s$.

*Proof.* By definition of $s^*(\cdot), \overline{c}(i+1; s^*(i+1)) = \min_s \overline{c}(i+1; s)$. Clearly,

$$\min_s \overline{c}(i; s) \leq \overline{c}(i; s^*(i+1)) \leq \overline{c}(i+1; s^*(i+1)),$$

since by Lemma 3, $\overline{c}(i; s^*(i+1)) \leq \overline{c}(i+1; s^*(i+1))$. The result follows by definition of $s^*(\cdot)$.
□

**Lemma 6** $f_n(i;s)$ and $f_i^*(i)$ are increasing in $i$ for all $s$.

*Proof.* The proof proceeds by backward induction on $n$.

For $n = N$, $f_N(i;s)$ equals $\overline{c}_N(i;s)$, which is increasing in $i$ by Lemma 3.

Assuming the result holds for $f_{n+1}$, we establish it for $f_n$.

From the proof of Lemma 3, we know that $X(i;s)$ is stochastically increasing in $i$, so that

$$X(i;s) \geq_{st} X(i-l;s), \quad i > l > 0.$$

By the induction hypothesis for any arbitrary fixed value of $\widehat{s}$, $f_{n+1}(i;\widehat{s})$ is increasing in $i$, so Lemma 1 applies to give

$$E[f_{n+1}(X(i;s);\widehat{s})] \geq E[f_{n+1}(X(i-l;s);\widehat{s})].$$

Since the preceding relation is true for arbitrary $\widehat{s}$, taking minimums on both sides, we have

$$\min_{\widehat{s}} E[f_{n+1}(X(i;s),\widehat{s})] \geq \min_{\widehat{s}} E[f_{n+1}(X(i-l;s),\widehat{s})].$$

Hence, $\min_{\widehat{s}} E[f_{n+1}(X(i;s),\widehat{s})]$ is increasing in $i$. This shows that $f_n(i;s)$ is the sum of two functions that are increasing in $i$, establishing the first part of the result.

By definition, $f_n^*(i) = \min_s\{f_n(i;s)\}$, and $f_n(i;s)$ is increasing in $i$ as just shown. Taking minimum with respect to $s$ establishes the second part of the lemma. □

*Proof of Theorem.* We proceed by backward induction.

For $n = N$, $s_N^*(i) = s_N(i)$, since both values minimize $\overline{c}_N(i;s)$ with respect to $s$ by definition.

Assume the result holds for $n+1$, i.e., $s_{n+1}^*(i) \geq s_{n+1}(i)$, we establish it for $n$.

Consider $s_n(i) - l,\ l > 0$. We will show that the total cost increases over using a staffing level of $s_n(i)$. Specifically, we show $f_n(i;s_n(i)-l) \geq f_n(i;s_n(i)),\ l > 0$, by showing

$$\begin{aligned} &\overline{c}_n(i;s_n(i)-l) + \min_s E[\overline{c}_{n+1}(X(i;s_n(i)-l);s)] \\ &\geq \ \overline{c}_n(i;s_n(i)) + \min_s E[\overline{c}_{n+1}(X(i;s_n(i));s)],\ l > 0. \end{aligned} \tag{7}$$

By definition of $s_n(i)$ minimizing the single-period cost, we have

$$\overline{c}_n(i;s_n(i)-l) \geq \overline{c}_n(i;s_n(i)),\ l > 0. \tag{8}$$

By Lemma 2, $X(i;s_n(i)-l) \geq_{st} X(i;s_n(i)),\ l > 0$, so that by Lemma 4,

$$E[\overline{c}_{n+1}(X(i;s_n(i)-l);s)] \geq E[\overline{c}_{n+1}(X(i;s_n(i));s)].$$

Taking minimums on both sides over $s$,

$$\min_s E[\overline{c}_{n+1}(X(i;s_n(i)-l),s)] \geq \min_s E[\overline{c}_{n+1}(X(i;s_n(i)),s)]. \tag{9}$$

Hence, combining (8) and (9) establishes (7), and the result follows. □

# 4 Summary and Conclusions

The original motivation for the work reported here was staffing in the presence of transient effects. We found that the optimal policy for the dynamic programming formulation is bounded below by the single-period decoupled solution. Experiments testing the quality of the lower bound were performed by Yoo (1996) as part of a computational investigation of the optimal staffing problem. The test problems indicated that the expected staffing (resulting from the decoupled staffing procedure of Section 2.3) was within 5% of the expected staffing associated with the optimal staffing policy in about 95% of the test cases; similarly, the expected cost was within 3% of the optimal in about 80% of the test cases.

## References


[1] Agnihothri, S. and P. Taylor, "Staffing a Centralized Appointment Scheduling Department in a Lourdes Hospital," *Interfaces*, 21 (1991), 5, 1-11.

[2] Andrews, B. and H. Parsons, "L.L.Bean Chooses a Telephone Agent Scheduling System," *Interfaces*, 19 (1989), 6, 1-9.

[3] Assad, A., unpublished notes (1992).

[4] Deutsch, H. and V. Mabert; "Queueing Theory and Teller Staffing," *Interfaces*, 10 (1980), 5, 63-67.

[5] Foote, B., "A Queueing Study of Drive-in Banking," *Interfaces*, 6 (1976), 4, 31-37.

[6] Green, L., and P. Kolesar, "The Lagged PSA for Estimating Peak Congestion in Multiserver Markovian Queues with Periodic Arrival Rates," *Management Science*, 43 (1997), 80-87.

[7] Green, L., P. Kolesar, and A. Svoronos, "Some Effects of Nonstationarity on Multiserver Markovian Queueing Systems," *Operations Research*, 39 (1991), 502-511.

[8] Heyman, S.P. and Sobel, M.J., *Stochastic Models in Operations Research*, Vol.II, McGraw-Hill, New York (1984).

[9] Jennings, O. et al., "Server Staffing to Meet Time-Varying Demand," *Management Science*, 42 (1996), 1383-1394.

[10] Khan, M. and B. Callahan, "Planning Laboratory Staffing with a Queueing Model," European J. *Operational Research*, 67 (1993), 321-333.

[11] Larson, R., "Improving the Effectiveness of New York City's 911," in *Analysis of Public Systems*, MIT Press, Cambridge, Mass. (1972).

[12] Quinn, P., B. Andrews, and H. Parsons, "Allocating Telecommunications Resources at L.L.Bean, Inc.," *Interfaces*, 21 (1991), 75-91.

[13] Ross, S., *Stochastic Processes*, Wiley, New York, 1983.

[14] Shaked, M. and Shanthikumar, G., *Stochastic Orders and Their Applications*, Academic Press, California (1994).

[15] Shanthikumar, J. and Yao, D., "Stochastic Monotonicity in General Queueing Networks," *Journal of Applied Probability*, 26 (1989), 413-417.

[16] Sonderman, D., "Comparing Multi-Server Queues with Finite Waiting Rooms, II: Different Number of Servers," *Advances in Applied Probability*, 11 (1979), 448-455.

[17] Stoyan, D., *Comparison Methods for Queues and Other Stochastic Models*, John Wiley, New York (1983).

[18] Yoo, J., *Queueing Models for Staffing Service Operations*, PhD dissertation, University of Maryland, College Park, College of Business and Management (1996).