On the Pathwise Optimal Bernoulli Routing Policy for Homogeneous Parallel Servers

Ger Koole
INRIA Sophia Antipolis
B.P. 93, 06902 Sophia Antipolis Cedex
France

Mathematics of Operations Research **21**:469–476, 1996

Abstract

A long-standing conjecture on the optimal Bernoulli routing policy is proven to be true. For the case of equal exponential service times it is shown that splitting equally among the queues minimizes the departure times in a stochastic pathwise sense. A new technique is used, showing that certain distributional properties related to Schur convexity propagate forward in time.
Keywords: parallel queues, Bernoulli routing, majorization, pathwise optimality
Note: This research was supported by a European HCM grant.

1 Introduction

Consider a system with m queues, each queue having a single exponential server with parameter μ. Customers arrive at the system according to a general arrival process. We consider the class of Bernoulli policies R^p, with $p = (p_1, \dots, p_m)$ a probability distribution, where R^p assigns each arriving customer to queue k with probability p_k, $1 \leq k \leq m$. We prove that R^{p^*} with $p^* = (\frac{1}{m}, \dots, \frac{1}{m})$ (which we call the Equal Splitting Policy, or ESP) minimizes the departure process of the system in a stochastic pathwise sense. For $m = 2$, the optimality of the ESP holds also for various other objective functions. Although strongly related, we show that our result is weaker than optimality in the weak Schur convex sense.

Several papers have studied Bernoulli routing. Chang, Chao and Pinedo [2] and Chang [1] show that the ESP minimizes in expectation functions of the workload which are increasing, symmetric, L-subadditive and convex in each variable. Note that this does not imply stochastic minimization of the workload, as indicator functions are not convex. Jean-Marie and Gün [7] and Chang [1] consider parallel queues with resequencing, i.e., systems in which customers have to leave the system in the same order as in which they entered it. In [1] a result similar to the one for the system without resequencing is obtained, while [7] mainly is concerned with stationary response times. See Combé and Boxma [3] for a recent result and references on the model with general and unequal service time distributions.

The outline of the paper is as follows. In section 2 two lemmas, which are the basis of our results, are shown for $m = 2$. The first considers just a single system with assignment vector $(p, 1-p)$, with $p \geq \frac{1}{2}$. It states that there are, in some stochastic sense, more customers in queue 1 than in queue 2 at any time. This property of distributions resulting from Bernoulli routing is used in a second lemma to compare two systems with different assignment vectors, resulting in a class of functions which are minimized at each point in time by Bernoulli routing.

In section 3 only the total number of customers in the system is considered. First it is shown that for this case the optimality at t can be seen to hold jointly over all t as well. Using coupling this can be generalized to m queues, leading to the fact that the ESP minimizes the departure times in a stochastic pathwise sense. The section is ended with some notes related to Schur convexity.

2 The Basic Result

In this section we confine ourselves to two queues. We use the notation R^p for the policy with assignment vector $(p, 1-p)$.

Let $\{a_n\}_{n=1}^{\infty}$ be the sequence of arrival events (possibly a sample from a stochastic point process). The departures can be seen to be generated by a Poisson process with rate 2μ. At each event of this process, a queue is selected, each with probability $\frac{1}{2}$, and a departure will take place at that queue if there is a customer present. If there is no customer present, nothing happens. Condition on this Poisson process, with event times $\{d_n\}_{n=1}^{\infty}$. Now order both types of events, resulting in an event list $\{b_n\}_{n=1}^{\infty}$, where each event is either an arrival or a (potential) departure. Let $Q^p(n) = (Q_1^p(n), Q_2^p(n))$ be the queue lengths directly after the nth event, in a system with assignment probabilities p to queue 1 and $1-p$ to queue 2, and initial state $Q^p(0)$.

Define for all $i, j, s \in \mathbb{N}_0$

$$A(i,j,s) = \{(x,y) \in \mathbb{N}_0^2 \mid x \le i, y \le j, x+y \le s\}.$$

Now let

$$P_n^p(i,j,s) = \mathbb{P}((Q_1^p(n), Q_2^p(n)) \in A(i,j,s)).$$

Take $P_0^p(i,j,s) = 1$ for all i, j, s, which corresponds to starting with an empty system.

The first lemma states a property of a single system, in terms of the sets $A(i,j,s)$. Intuitively speaking, lemma 2.1 says (for $k = 0$) that if the assignment probability to queue 1 is bigger than $\frac{1}{2}$, then states with more customers in queue 1 than in queue 2 are more likely to occur than states with more customers in queue 2.

Lemma 2.1 *For $p \ge \frac{1}{2}$ and for all $i, j, k, s, n \ge 0$*

$$P_n^p(i+j, i+k, s) \ge P_n^p(i, i+j+k, s).$$

Proof. We will use induction on n. Note that the assertion is true for $n = 0$. Assume the lemma holds up to n. Let us first consider the case that the $(n+1)$th event is a (potential) departure. Then we have

$$P_{n+1}^p(i,j,s) = \tfrac{1}{2} P_n^p(i+1, j \wedge s, s+1) + \tfrac{1}{2} P_n^p(i \wedge s, j+1, s+1). \tag{1}$$

This equation can be explained as follows. Let us pay attention to $P_n^p(i+1, j \wedge s, s+1)$, which takes into account departures from queue 1. It is easily seen that

$$A(i+1, j \wedge s, s+1) = \{(x,y) \in \mathbb{N}_0^2 \mid 1 \le x \le i+1, y \le j, x+y \le s+1\} \cup A(0,j,s).$$

The first set consists of all states from which a departure leads into $A(i,j,s)$. Besides that, it is possible that queue 1 was already empty, resulting in a transition from a state to itself. This explains the second set and therefore the first term of (1). The second term is similar.

We apply (1) to $P^p_{n+1}(i+j,i+k,s)$. Note that we can reduce the values of i, j and k until $i+j \le s$ and $i+k \le s$, without changing $P^p_{n+1}(i+j,i+k,s)$ or $P^p_{n+1}(i,i+j+k,s)$, so we assume that these inequalities hold. Consider first the case $k>0$. Then we have

$$\begin{aligned} P^p_{n+1}(i+j,i+k,s) &= \tfrac{1}{2}P^p_n(i+j+1,i+k,s+1)+\tfrac{1}{2}P^p_n(i+j,i+k+1,s+1) \\ &\ge \tfrac{1}{2}P^p_n(i+1,i+j+k,s+1)+\tfrac{1}{2}P^p_n(i,i+j+k+1,s+1) \\ &\ge \tfrac{1}{2}P^p_n(i+1,(i+j+k)\wedge s,s+1)+\tfrac{1}{2}P^p_n(i\wedge s,i+j+k+1,s+1) \\ &= P^p_{n+1}(i,i+j+k,s), \end{aligned}$$

where the first equality follows because $i+j \le s$ and $i+k \le s$, the first inequality follows by induction.

Now consider the case $k=0$. Assuming $i+j \le s$ and similarly $s \le 2i+j$, we have, by using equation (1),

$$\begin{aligned} P^p_{n+1}(i+j,i,s) &= \tfrac{1}{2}P^p_n(i+j+1,i,s+1)+\tfrac{1}{2}P^p_n(i+j,i+1,s+1) \\ &= \tfrac{1}{2}P^p_n(i+j,i,s+1)+\tfrac{1}{2}P^p_n(i+j+1,i+1,s+1). \end{aligned} \tag{2}$$

The condition $s \le 2i+j$ is needed to exclude the point $(i+j+1,i+1)$ from $A(i+j+1,i+1,s+1)$, as is easily verified by drawing the sets. Because by induction

$$P^p_n(i+j,i,s+1) \ge P^p_n(i,i+j,s+1)$$

and

$$P^p_n(i+j+1,i+1,s+1) \ge P^p_n(i+1,i+j+1,s+1),$$

$P^p_{n+1}(i+j,i,s) \ge P^p_{n+1}(i,i+j,s)$ follows by using equation (2).

Now consider an arrival event. Define $P^p_n(i,j,s)=0$ if i, j or s is negative. As $p \ge \frac{1}{2}$, we can think of an arrival event being constructed from two successive Bernoulli experiments in the following way. With probability $2(1-p)$ the first experiment leads to a "symmetric" arrival as second experiment that sends the arriving customer to each queue with probability $\frac{1}{2}$. With probability $2p-1$ the first experiment leads to an "asymmetric" arrival, which is an arrival at queue 1. Indeed, the total probability of sending an arrival to queue 1 is correct under this construction, as $\frac{1}{2}(2(1-p))+(2p-1)=p$. Now condition also on the type of arrival, symmetric or asymmetric. Let us start with the symmetric arrivals. In general, we have that

$$P^p_{n+1}(i,j,s)=\tfrac{1}{2}P^p_n(i-1,j,s-1)+\tfrac{1}{2}P^p_n(i,j-1,s-1). \tag{3}$$

This leads to, for the case $k>0$,

$$P^p_{n+1}(i+j,i+k,s)=\tfrac{1}{2}P^p_n(i+j-1,i+k,s-1)+\tfrac{1}{2}P^p_n(i+j,i+k-1,s-1).$$

We have

$$P^p_n(i+j-1,i+k,s-1) \ge P^p_n(i-1,i+j+k,s-1)$$

and

$$P^p_n(i+j,i+k-1,s-1) \ge P^p_n(i,i+j+k-1,s-1).$$

Indeed, the first inequality follows by induction for $i>0$, and holds trivially for $i=0$. The second inequality follows by induction in all cases.

In case $k=0$, we can assume $j>0$, as the case $j=0$ follows trivially. The case $i=0$ is similar to that above. Assume therefore $i>0$. Now we have

$$\begin{aligned} P^p_{n+1}(i+j,i,s) &= \tfrac{1}{2}P^p_n(i+j-1,i,s-1)+\tfrac{1}{2}P^p_n(i+j,i-1,s-1) \\ &= \tfrac{1}{2}P^p_n(i+j-1,i,s-1)+\tfrac{1}{2}P^p_n((i-1)+(j+1),i-1,s-1). \end{aligned}$$

Now $P^p_{n+1}(i+j,i,s) \geq P^p_{n+1}(i,i+j,s)$ follows easily by induction.

Finally, consider an arrival at queue 1. We can assume $i>0$, as $P^p_{n+1}(0,j+k,s)=0$. Then

$$\begin{aligned} P^p_{n+1}(i+j,i+k,s) &= P^p_n(i+j-1,i+k,s-1) \\ &\geq P^p_n(i-1,i+j+k,s-1) = P^p_{n+1}(i,i+j+k,s). \end{aligned}$$

□

Based on the previous lemma, we can now compare two systems with different assignment parameters. The lemma basically states that when comparing the systems under R^p and R^q, for $q \geq p \geq \frac{1}{2}$, the system under R^p puts more probability mass on sets $A(i,k,s)$ with $k \leq i$, i.e., it puts more mass on sets with a surplus of queue 2 customers.

Lemma 2.2 *For $q \geq p \geq \frac{1}{2}$ and for all $i,j,s,n \geq 0$,*

$$P^p_n(i,i+j,s) \geq P^q_n(i,i+j,s).$$

Proof. The proof is similar to that of lemma 2.1. If an equation holds for both P^p and P^q, we omit the superscript. Assume first that the $(n+1)$th event is a (potential) departure. Assume also, without loss of generality, that $i+j \leq s$. By (2) and the induction hypothesis we have

$$\begin{aligned} P^p_{n+1}(i,i+j,s) &= \tfrac{1}{2}P^p_n(i,i+j,s+1)+\tfrac{1}{2}P^p_n(i+1,i+j+1,s+1) \\ &\geq \tfrac{1}{2}P^q_n(i,i+j,s+1)+\tfrac{1}{2}P^q_n(i+1,i+j+1,s+1) \\ &= P^q_{n+1}(i,i+j,s). \end{aligned}$$

Now consider an arrival event. As in the proof of lemma 2.1, we distinguish between symmetric and asymmetric arrivals, although their definitions are somewhat different. Under both systems the symmetric arrival occurs with probability $2(1-q)$, the asymmetric arrival with probability $2q-1$. If there is an asymmetric arrival, this customer joins queue 1 in the system operated by R^q. Some computations show for the system with policy R^p, and an asymmetric arrival, this customer should join queue 1 with probability $p' = (p+q-1)/(2q-1)$. This leads to an overall probability of joining queue 1 under R^p of $\frac{1}{2}2(1-q)+p'(2q-1)=p$. We condition also on the type of arrival.

Consider a symmetric arrival. First we consider the case $j>0$. Then, by (3)

$$P_{n+1}(i,i+j,s) = \tfrac{1}{2}P_n(i-1,i+j,s-1)+\tfrac{1}{2}P_n(i,i+j-1,s-1).$$

By differentiating between $i=0$ and $i>0$, and by induction, $P^p_{n+1}(i,i+j,s) \geq P^q_{n+1}(i,i+j,s)$ follows.

Now consider the case $j=0$. As in (2), we have

$$P_{n+1}(i,i,s) = \tfrac{1}{2}P_n(i-1,i-1,s-1)+\tfrac{1}{2}P_n(i,i,s-1).$$

By induction the inequality follows.

Finally, consider an asymmetric arrival. As argued above, under R^p (R^q) the arriving customer joins queue 1 with probability p' (1). The case $i = 0$ is easy, as $P^q_{n+1}(0, j, s) = 0$. Thus assume $i > 0$. Then

$$\begin{aligned}
P^p_{n+1}(i, i+j, s) &= p' P^p_n(i-1, i+j, s-1) + (1-p')P^p_n(i, i+j-1, s-1) \\
&\geq p' P^p_n(i-1, i+j, s-1) + (1-p')P^p_n(i-1, i+j, s-1) \\
&= P^p_n(i-1, i+j, s-1) \geq P^q_n(i-1, i+j, s-1) = P^q_{n+1}(i, i+j, s),
\end{aligned}$$

where the first inequality follows by lemma 2.1, the second follows by induction. $\square$

It is immediate from lemma 2.2 that $P^{\frac{1}{2}}_n(i, i, s) \geq P^q_n(i, i, s)$ for $q \geq \frac{1}{2}$. Obviously, by symmetry the same holds for $q \leq \frac{1}{2}$. This leads us to the following. Let $Q^p(t) = (Q^p_1(t), Q^p_2(t))$ be the queue length vector at time t, under the policy with assignment probability p.

Theorem 2.3 *The function $f(Q^p(t))$ is maximized in expectation by the ESP, for all t, and all f of the form $f(x) = \sum_k \alpha_k I\{x \in A(i_k, i_k, s_k)\}$, with $\alpha_k \geq 0$.*

Proof. We show that the result holds for every sequence $\{c_n\}$. Let $c_n < t \leq c_{n+1}$. Then $Q^p(t) = Q^p(n)$. Because $\mathbb{E} f(Q^p(t)) = \sum \alpha_k \mathbb{P}(Q^p(t) \in A(i_k, i_k, s_k)) = \sum \alpha_k P^p(i_k, i_k, s_k)$, the result follows from lemma 2.2. $\square$

An obvious choice for f is $f(Q(t)) = I\{Q(t) \in A(s, s, s)\} = \mathbb{P}(Q(t) \leq s)$, leading to the result that the ESP minimizes the total number of customers in the system stochastically at any time t.

The next section is devoted to a more systematic investigation of the allowable cost functions. Among other things we show that the optimality of the ESP also holds in the stochastic pathwise sense, i.e., jointly over t.

The generalization to more than two queues is non-trivial. The collection of sets of the form $A(x, s) = \{y \in \mathbb{N}^m_0 \mid y_n \leq x_n, \sum_n y_n \leq s\}$ (which would be the obvious generalization to m dimensions) is not closed in the sense that if there is a departure, $P^p_{n+1}(x, s)$ cannot be written in terms of these sets, and therefore theorem 2.3 cannot be generalized directly to more than two queues. However, the main result of the paper, the pathwise optimality of the ESP, will be proven for multiple queues on the basis of theorem 2.3, using coupling arguments, in section 3.

Remark 2.4 The condition that the system is initially empty is unnecessarily restrictive; any initial distribution P^p_0 satisfying $P^p_0(i+j, i+k, s) \geq P^p_0(i, i+j+k, s)$ for all s and all i, j and k will do.

Remark 2.5 Another method to compare two systems (for example, two systems operated by different Bernoulli policies) using distributions explicitly is the one described in Stoyan [10], section 4.2. There are two crucial differences with the current method. First, in [10] the set of all probability distributions on the state space is used. We manage to restrict this set to those satisfying lemma 2.1. Second, Stoyan [10] assumes in equation (4.2.4) that the two systems to be considered are ordered for each initial state. For models with complete state observations this is indeed often the case; e.g., the shortest queue policy is optimal for each initial state ([11]). For the current model that does not hold: if queue 2 contains more customers than queue 1 it is obviously better to send more customers to queue 1, in order to minimize the number of customers in the system in the near future. This is precisely where lemma 2.1 comes in: it states that states where the ESP behaves worse are less likely to occur, given an initially empty system. In Massey [9] a method related to Stoyan's is described.

3 On Pathwise Optimality and Stochastic Orderings

It was shown in the previous section that the ESP minimizes stochastically the number of customers in the system at any time t, and therefore it maximizes the number of departures. Here we will prove that this result also holds jointly over all t, i.e., $\mathbb{P}(Q_1^p(1)+\ldots+Q_m^p(1) \leq s_1, \ldots, Q_1^p(n)+\ldots+Q_m^p(n) \leq s_n)$ is maximized by the ESP, for all n and $s_1, \ldots, s_n$. In the proof it is first shown for $m=2$, based on the results of the previous section. Then, using coupling arguments, we generalize it to m queues. After that we will pay attention to the departure process.

Theorem 3.1 *For all n and $s_1, \ldots, s_n$,*

$$\mathbb{P}(Q_1^p(1) + \ldots + Q_m^p(1) \leq s_1, \ldots, Q_1^p(n) + \ldots + Q_m^p(n) \leq s_n)$$

is maximized by $p = p^*$, i.e., the ESP minimizes the total number of customers jointly over time.

Proof. Define $Q_T^p(k) = Q_1^p(k) + \ldots + Q_m^p(k)$. First we consider the case $m = 2$. For simplicity, assume that $n = 2$. As $Q^p(2) = f(Q^p(1), U)$, for some appropriate f (depending on the type of transition), and U a r.v. with distribution P_U governing the transition, we have

$$\mathbb{P}(Q_T^p(1) \leq s_1, Q_T^p(2) \leq s_2) =$$

$$\sum_x \sum_u P_1^p(x) P_U(u) I\{x_1 + x_2 \leq s_1\} I\{f_1(x,u) + f_2(x,u) \leq s_2\},$$

where we use the fact that U is independent of the state of the system.

Now define the defective distribution $\tilde{P}_1^p$ by $\tilde{P}_1^p(x) = P_1^p(x) I\{x \leq s_1\}$. Let $\tilde{Q}^p(1)$ have distribution $\tilde{P}_1^p$, and $\tilde{Q}^p(2) = f(\tilde{Q}^p(1), U)$. It is readily seen that

$$\mathbb{P}(Q_T^p(1) \leq s_1, Q_T^p(2) \leq s_2) = \mathbb{P}(\tilde{Q}_T^p(2) \leq s_2).$$

Thus, generalizing to $n \geq 2$, by replacing P_n^p by $\tilde{P}_n^p = P_n^p I\{\cdot \leq s_n\}$ after each transition, we find the distribution of $\mathbb{P}(Q_T^p(1) \leq s_1, \ldots, Q_T^p(n) \leq s_n)$. To prove the optimality of the ESP in this case, we have to prove that the lemmas of the previous section hold for $\tilde{P}_n^p$ instead of P_n^p. That is easily checked, because $\tilde{P}_n^p(i,j,s) = P_n^p(i,j,s \wedge s_n)$. This proves the theorem for the case of 2 queues.

Now consider the general case with m queues. Assume we have 2 vectors with assignment probabilities p and q, with $p_n = q_n$ for $n \neq l, r$ and $q_l \geq p_l \geq p_r \geq q_r$. As events, we distinguish between an arrival at queue l or r (with probability $p_1 + p_2$), and arrivals at queue n, $n \neq l, r$ (each having probability $p_n = q_n$). Similarly, one event deals with departures from queue l or r (probability $\frac{2}{m}$), and there is a separate departure event for each other queue (each with probability $\frac{1}{m}$), giving a total of $2(m-1)$ types of events. By conditioning on all these separate events, we see that under R^p and R^q all events at queues other than l and r are exactly the same, and thus $\sum_{i \neq l,r} Q_i^p(k) = \sum_{i \neq l,r} Q_i^q(k)$ for all k. Combining this with the result for 2 queues gives us that $\mathbb{P}(Q_T^p(1) \leq s_1, \ldots, Q_T^p(n) \leq s_n) \geq \mathbb{P}(Q_T^q(1) \leq s_1, \ldots, Q_T^q(n) \leq s_n)$. Repetition of this argument for different l and r proves the theorem. □

Let $D^p(n)$ be the total number of departures directly after the nth event. It follows immediately that $\mathbb{P}(D^p(1) \geq d_1, \ldots, D^p(n) \geq d_n)$ is maximized by $p = p^*$. Then, when comparing R^p and R^{p^*}, we can couple the sample paths such that there are always less customers in the system operated by R^{p^*}. In such a case, we say that R^{p^*} is better in a pathwise sense. Summarizing, we have the following.

Corollary 3.2 *The ESP maximizes the number of departures pathwise.*

We will end this section with some notes related to Schur convexity. Let us first define the weak majorization ordering. We say that x is weakly majorized by y, $x \prec_w y$, if $\sum_{i=1}^k x_{[i]} \leq \sum_{i=1}^k y_{[i]}$ for $k = 1, \ldots, m$, where $x_{[i]}$ is the ith largest element of x.

The functions $\phi \in \mathcal{C}$ preserving this ordering (i.e., the functions for which $x \prec_w y \Rightarrow \phi(x) \leq \phi(y)$) are called weak Schur convex (Marshall and Olkin [8]).

If $x \prec_w y$ then x is more balanced and/or smaller than y. In routing models, x is typically a state preferable to y. Indeed, for several routing models policies are found which minimize the queue lengths or the workloads in a weak Schur convex sense ([11, 6, 5, 4]).

The current result is strongly related to weak Schur convexity, as we have $A(i, i, s) = \{x \mid x \prec_w (i, s - i)\}$ (assuming that $i \leq s \leq 2i$).

It is worth noticing that the current result states that R^p is better than R^q if $p \prec q$ (which means that $p \prec_w q$ and $\sum_n p_n = \sum_n q_n$), as was pointed out in [2].

For random variables X and Y we say that $X \prec_w Y$ for r.v.'s X and Y if $\phi(X) \leq_s \phi(Y)$ for all weak Schur convex ϕ, and $\leq_s$ the usual stochastic ordering. For routing models there are often optimality results for all weak Schur convex functions. Consider for example the model where routing decisions are allowed to depend on the queue lengths, and let X_t and Y_t be the queue length vector at t under shortest queue routing and an arbitrary policy. For this case it is shown in [11, 6] that $X_t \prec_w Y_t$ for all t. However, our result is weaker. Indeed, take $g \in \mathcal{C}$ defined by $g(x) = 1 - I\{x \prec_w (3, 3) \text{ or } x \prec_w (2, 0)\}$. It is easily checked that this is a weak Schur convex function. The current result would imply that the ESP minimizes g if it could be written as a convex combination of functions of the form $I\{x \in A(i, i, s)\}$. However, $g = I\{x \in A(2, 2, 4)\} + I\{x \in A(3, 3, 3)\} - I\{x \in A(2, 2, 3)\}$, which does not fall within our class of functions. On the other hand, as convexity of the cost functions is not needed, the result is stronger than most other often used stochastic orderings. I conjecture that the result also holds for all weak Schur convex functions. To prove this in a similar manner would imply proving the lemma's for all weak Schur convex 0-1 functions, which would complicate the proofs considerably; the sets as they are now seem to be the minimum needed to get the induction working.

Acknowledgement. I would like to thank Rhonda Righter for stimulating discussions.

References

[1] C.S. Chang. A new ordering for stochastic majorization: Theory and applications. *Advances in Applied Probability*, 24:604–634, 1992.

[2] C.S. Chang, X. Chao, and M. Pinedo. A note on queues with Bernoulli routing. In *Proceedings of the 29th Conference on Decision and Control*, pages 897–902, 1990.

[3] M.B. Combé and O.J. Boxma. Optimization of static traffic allocation policies. *Theoretical Computer Science*, 125:17–43, 1994.

[4] D.J. Daley. Certain optimality properties of the first-come first-served discipline for $G|G|s$ queues. *Stochastic Processes and their Applications*, 25:301–308, 1987.

[5] S.G. Foss. *Extremal Problems in Queueing Theory.* PhD thesis, Novosibirsk State University, 1982. In Russian.

[6] A. Hordijk and G.M. Koole. On the assignment of customers to parallel queues. *Probability in the Engineering and Informational Sciences*, 6:495–511, 1992.

[7] A. Jean-Marie and L. Gün. Parallel queues with resequencing. *Journal of the ACM*, 40:1188–1208, 1993.

[8] A.W. Marshall and I. Olkin. *Inequalities: Theory of Majorization and its Applications.* Academic Press, 1979.

[9] W.A. Massey. Stochastic orderings for Markov processes on partially ordered spaces. *Mathematics of Operations Research*, 12:350–367, 1987.

[10] D. Stoyan. *Comparison Methods for Queues and other Stochastic Models.* Wiley, 1983.

[11] J. Walrand. *An Introduction to Queueing Networks.* Prentice-Hall, 1988.