MATHEMATICS OF OPERATIONS RESEARCH
Vol. 11, No. 4, November 1986
Printed in U.S.A.

STOCHASTIC COMPARISONS FOR NON-MARKOV PROCESSES

WARD WHITT

AT & T Bell Laboratories

A technique is developed for comparing a non-Markov process to a Markov process on a general state space with many possible stochastic orderings. Two such comparisons with a common Markov process yield a comparison between two non-Markov processes. The technique, which is based on stochastic monotonicity of the Markov process, yields stochastic comparisons of the limiting distributions and the marginal distributions at single time points, but not the joint distributions. These stochastic comparisons are obtained under the condition that the one-step transition probabilities (in discrete time) or infinitesimal transition rates (in continuous time) are appropriately ordered, uniformly in the extra information needed to add to the state to make the non-Markov process Markov. The technique is illustrated by establishing an inequality in traffic theory concerning the blocking in multi-server service facilities when each customer requires service from several facilities simultaneously.

1. Introduction. It is often of interest to make stochastic comparisons for non-Markov processes. One way to do this, exploiting established comparison methods for Markov processes, is to make stochastic comparisons of the transition probabilities (or transition rates for continuous-time processes) that hold uniformly in the extra information needed to add to the state to make the non-Markov process Markov. This technique has been applied to compare semi-Markov processes by Sonderman [15], general counting processes by Whitt [17] and generalized birth-and-death processes (non-Markov jump processes on the integers that move up or down one step at a time) by Smith and Whitt [14].

In all the applications above, the stochastic order relation has been the standard stochastic order determined by all increasing sets, as in Kamae, Krengel and O'Brien [4], or a stronger form such as the monotone-likelihood-ratio ordering in Theorem 5 of [14]. However, in many applications, especially when the state space is a product space such as R^n, such strong stochastic comparisons do not hold, and we must look for weaker notions of stochastic order. One way to obtain many different stochastic orderings for Markov processes on general state spaces is to exploit stochastic monotonicity of Markov processes. Stochastic monotonicity for Markov processes was introduced by Daley [3]. Its power for comparing Markov processes on general state spaces with many different stochastic orderings is expressed in Kester [6] and §4.2 of Stoyan [16]. The technique has also been applied to Markovian queueing networks and studied further by Massey [7–10].

The purpose of this paper is to combine the two techniques above to provide conditions for establishing many different stochastic orderings for non-Markov processes. Our results can be viewed as extensions of the theorems for Markov

*Received March 22, 1985.
AMS 1980 subject classification. Primary: 60G07 Secondary: 60G99.
IAOR 1973 subject classification. Main: Stochastic processes.
OR/MS Index 1978 subject classification. Primary: 565. Probability/Distribution Comparisons.
Key words. Stochastic order; stochastic dominance; stochastic monotonicity; comparison of stochastic processes; Markov processes; queueing theory; blocking.

0364-765X/86/1104/0608$01.25

processes on general state spaces in §4.2 of Stoyan [16]. In fact, our results can also be viewed as applications of the theorems in Stoyan [16], as we indicate in the remarks after Theorem 1. Furthermore, to a large extent, our results can also be viewed as an application of Theorem 3.5 of Massey [10]. (The main results here and in [10] were obtained concurrently and independently.) However, the point of view here is quite different, because the process in which we are initially interested is not Markov.

The stochastic comparison method here has the desirable property that both the theorems here and their application to problems of interest (e.g., §7) are established relatively easily, while the problems of interest are often quite difficult otherwise. Since the theorems are relatively easy to prove and apply, the primary contribution is the formulation of the theorems and the demonstration that they indeed have significant applications. In this regard, Massey deserves credit for recognizing the value of this general stochastic comparison approach for analyzing Markovian queueing networks.

We discovered this stochastic comparison approach during an investigation of a specific problem in traffic theory: the problem of blocking when several facilities are required simultaneously [18]. During the development of a software package in the Operations Research Department of AT & T Bell Laboratories to assist in the design of packet-switched communication networks, the need arose to approximately describe the blocking (percentage of failed attempts) in setting up virtual circuits [12, 13]. Analysis of the blocking is difficult because a circuit typically requires the simultaneous possession of limited resources associated with several different facilities (transmission links, memory buffers, etc.). Moreover, there is competition for the resources not only from demands for circuits on the same path, but also from demands for different circuits that use only some of the same facilities. Hence, even without waiting or alternate routing (which were not considered), the blocking is complicated. This multiple-facility blocking model for communication networks has also been studied by Burman, Lehoczky and Lim [2] and D. P. Heyman (personal communication). A special case of this model to represent database locking has also been studied by Mitra and Weinberger [11].

A standard approximation for the blocking in such complex settings is based on assuming that the facilities are independent: The approximate probability of no blocking for each customer is thus the product of the probabilities of no blocking in the required facilities, where the offered load at each facility is the sum of the offered loads of all classes requiring a server there, and the blocking at each facility is computed with the classical Erlang loss formula, which is displayed in (11) here. This approximation has long been regarded as conservative, but there has apparently been no proof. The methods here enable us to prove that the approximation is indeed an upper bound. We describe this application in §7. It is also discussed in much greater detail in [18]. Another contribution in [18] is an improved approximation, called the reduced-load approximation, which seems very promising.

The rest of this paper is organized as follows. In §2 we specify the stochastic order relations to be considered, which are defined directly on the space of all probability measures on the state space rather than via an ordering on the state space itself (as in [4] and [10]). In §3 we define stochastic monotonicity for Markov processes in the setting of §2. §4 contains the comparison results for discrete-time processes and §5 contains the corresponding comparison results for continuous-time processes. §6 extends §4 to the situation in which the ordering is only defined for a subset of all probability measures. Finally, the traffic-theory application is discussed in §7.

2. Integral stochastic order relations. Let S be a general state space and let $\mathscr{P}(S)$ be the space of all probability measures on S. Let $\leqslant$ be an order relation on $\mathscr{P}(S)$

defined by

$$P_1 \leqslant P_2 \quad \text{if } P_1 f \leqslant P_2 f \quad \text{for all } f \in \mathscr{F} \quad \text{where} \tag{1}$$

$$P_i f = \int_S P_i(ds) f(s) \tag{2}$$

and $\mathscr{F}$ is a class of $\mathscr{P}(S)$-integrable real-valued functions on S, i.e., that are integrable with respect to all $P \in \mathscr{P}(S)$. Often, as in Massey [7–10], the functions in $\mathscr{F}$ will be indicator functions of increasing measurable sets A in some class $\mathscr{A}$, so that $P_1 \leqslant P_2$ if $P_1(A) \leqslant P_2(A)$ for all $A \in \mathscr{A}$. However, even if the functions f are indicator functions, the sets A need not be increasing. Moreover, the functions in $\mathscr{F}$ could also be quite different, e.g., a subset of convex functions to represent a variability ordering. Further discussions of variability orderings and other interesting classes $\mathscr{F}$ is contained in Stoyan [16]. Because of the way the ordering $\leqslant$ is defined on $\mathscr{P}(S)$ by (1), we call it an *integral stochastic ordering.*

In other definitions of stochastic orderings, as in [16], the functions f in $\mathscr{F}$ are not required to be integrable with respect to all P. Then we would specify the order by having (1) hold for all f such that both integrals are well defined. In our general setting this convention could lead to difficulties, so we do not adopt it. For example, suppose that $\mathscr{F}$ contains only the two functions f_1 and f_2. Suppose that f_1 is integrable with respect to P_2 but not integrable with respect to P_1 and P_3 while f_2 is integrable with respect to P_1 and P_3 but not integrable with respect to P_2. Then $P_1 \leqslant P_2$ and $P_2 \leqslant P_3$ by default, but in general we do not have $P_1 \leqslant P_3$, so transitivity would fail.

Obviously, the relation $\leqslant$ defined by (1) is reflexive and transitive, but not necessarily a partial order because the antisymmetric property need not hold: $P_1 \leqslant P_2$ and $P_2 \leqslant P_1$ together do not necessarily imply that $P_1 = P_2$. Of course, the relation $\leqslant$ being a partial order is equivalent to $\mathscr{F}$ being a determining class [1], which is not necessary to assume.

To treat limiting distributions, we will also want the space of all probability measures $\mathscr{P}(S)$ to be endowed with a definition of convergence $\Rightarrow$ for sequences of probability measures. Then we assume only that the convergence and the order relation are compatible in the sense that the order relation is closed: If $P_{1n} \leqslant P_{2n}$ for all n and $P_{in} \Rightarrow P_i$ as $n \to \infty$ for each i, then $P_1 \leqslant P_2$.

It is significant that we have assumed neither an order relation nor a topology for the underlying space S. Of course, as in Massey [10], S will often be endowed with an order relation so that $\mathscr{F}$ is a family of indicator functions of increasing sets, but this is not necessary. Also, S will typically be endowed with a topology so that $\Rightarrow$ corresponds to weak convergence as in [1], but this is not necessary either. So the present framework is much more general than the theory of stochastic order in Kamae, Krengel and O'Brien [4], where S is a Polish space (metrizable as a complete separable metric space) endowed with a closed partial order and $\mathscr{F}$ is the set of indicator functions of all increasing subsets with respect to this order. Of course, in applications it is rarely a restriction to assume that underlying space S is Polish with a closed partial order. The most important departure from [4] here is that the order relation on $\mathscr{P}(S)$ need not be determined by the indicator functions of all increasing subsets of S, so that for any given order relation and topology on S there are many possible order relations $\leqslant$ on $\mathscr{P}(S)$.

The present setting is closer to Massey [7–10]. The most important departure from [10] is that the order relation on $\mathscr{P}(S)$ need not be determined by a subset of increasing subsets of S. For example, interesting variability ordering applications are possible if S is a compact subset of R^n and $\mathscr{F}$ is a subset of all real-valued convex functions on S. However, our motivating application in §7 is actually in the setting considered by Massey [7–10]. Then $S = R^n$ with the usual topology and order relation on S ($\mathbf{x}_1 \leqslant \mathbf{x}_2$ for $\mathbf{x}_i = (x_{i1}, \ldots, x_{in})$ if $x_{1j} \leqslant x_{2j}$ for each j), and $\mathscr{F}$ is the set of indicator functions of all lower sets $L(\mathbf{x}) = \{\mathbf{y} \in R^n\colon \mathbf{y} \leqslant \mathbf{x}\}$. Then $P_1 \leqslant P_2$ if and only if $F_1(\mathbf{x}) \leqslant F_2(\mathbf{x})$ for all $\mathbf{x}$, where F_i is the cdf of P_i. This ordering is the weak* ordering, denoted by $\langle x \rangle^*$, in Massey [10].

We conclude this section by indicating how a measurable mapping g from one space S to another S' may induce order for image measures. Suppose that orderings $\leqslant$ on $\mathscr{P}(S)$ and $\mathscr{P}(S')$ are determined by sets of real-valued functions $\mathscr{F}$ and $\mathscr{F}'$, respectively. For any probability measure P in $\mathscr{P}(S)$, let Pg^{-1} be the image measure in $\mathscr{P}(S')$ defined as usual by $(Pg^{-1})(A') = P(g^{-1}(A'))$ for each measurable subset A' of S'. The following elementary proposition motivates us to call g *isotone* with respect to $(S, \mathscr{F})$ and $(S', \mathscr{F})$ if $f' \circ g \in \mathscr{F}$ for all $f' \in \mathscr{F}'$; see Proposition 2.9 of Massey [10].

PROPOSITION 1. *If $f' \circ g \in \mathscr{F}$ for each $f' \in \mathscr{F}'$ and if $P_1 \leqslant P_2$ in $(\mathscr{P}(S), \mathscr{F})$, then $P_1 g^{-1} \leqslant P_2 g^{-1}$ in $(\mathscr{P}(S'), \mathscr{F}')$.*

3. Stochastically monotone Markov processes. A key ingredient in our approach is stochastic monotonicity for Markov processes; see [3], [5–10] and [16]. From §4.2 of Stoyan [16], it is clear that stochastic monotonicity applies in the general framework of §2.

Let $X \equiv \{X(n), n = 0, 1, \ldots\}$ be a discrete-time Markov process on the state space S with stationary transition kernel $K(s, A)$. The process X and its kernel K are said to be *stochastically monotone* with respect to $(\mathscr{P}(S), \leqslant)$ if $P_1 K \leqslant P_2 K$ whenever $P_1 \leqslant P_2$. Both the condition and the conclusion involve the same integral stochastic order relation $\leqslant$ on $\mathscr{P}(S)$.

As in §1 of Keilson and Kester [5], it is easy to see that stochastic monotonicity is preserved under many basic operations.

PROPOSITION 2. *Let K_i be stochastically monotone Markov transition kernels on S for each i.*

(a) *If $\{p_i, i \geqslant 1\}$ is a probability mass function, then $\sum_i p_i K_i$ is a stochastically monotone Markov transition kernel on S.*

(b) *The product (iterated operator) $K_n K_{n-1} \cdots K_1$ is a stochastically monotone Markov transition kernel on S for each n.*

4. Non-Markov processes. Let $X_i \equiv \{X_i(n), n = 0, 1, \ldots\}$ be discrete-time stochastic processes with values in S for $i = 1, 2$. We assume that X_2 is Markov with transition kernel K_2, but we do not assume that X_1 is Markov. Our goal is to obtain stochastic bounds for the general stochastic process X_1 in terms of the Markov process X_2.

As in [15] or Theorem 5 of [14], we assume that the evolution of X_1 can be described by one-step transition probabilities if we include additional information, which is represented by a discrete-time stochastic process Y_1 with state space S'. In particular, we assume that the process $(X_1, Y_1) \equiv \{(X_1(n), Y_1(n)), n = 0, 1, \ldots\}$ is a discrete-time Markov process with product state space $S \times S'$ and transition kernel $K_1 \equiv K_1((s, s'), A)$ for $A \subseteq S \times S'$. Let $\pi\colon S \times S' \to S$ be the projection map defined by $\pi((s, s')) = s$, so that $\pi(X_1(n), Y_1(n)) = X_1(n)$ and $P\pi^{-1}$ is the marginal distribution on S for each probability measure P on $S \times S'$. We assume that the product space $S \times S'$ is endowed with the usual product σ-field. (If topologies are defined on S and S' and $S \times S'$ is endowed with the product topology then S and S' should be separable metric spaces to ensure that the Borel σ-field is the product σ-field; p. 225 of Billingsley [1].)

As in [15] or Theorem 5 of [14], we obtain stochastic comparisons between $X_1(n)$ and $X_2(n)$ for all n by having the transition probabilities appropriately ordered on S for all initial points $s \in S$, uniformly in the extra information $s' \in S'$. Let P_1 and P_2 be initial probability measures on $S \times S'$ for X_1 and on S for X_2, respectively.

THEOREM 1. *Let $\mathcal{P}(S)$ be endowed with an order relation $\leqslant$ determined by a set $\mathcal{F}$ of $\mathcal{P}(S)$-integrable functions. If*

(i) K_2 is stochastically monotone,

(ii) $K_1((s, s'), \cdot)\pi^{-1} \leqslant K_2(s, \cdot)$ for all $(s, s') \in S \times S'$, and

(iii) $P_1\pi^{-1} \leqslant P_2$,

then $(P_1K_1^n)\pi^{-1} \leqslant P_2K_2^n$ for all n.

PROOF. Let $n = 1$. For $f \in \mathcal{F}$,

$$[(P_1K_1)\pi^{-1}]f = P_1(K_1\pi^{-1})f = \int_S\int_{S\times S'} P_1(ds, ds')(K_1\pi^{-1})((s, s'), ds'')f(s'')$$

$$\leqslant \int_S\int_{S\times S'} P_1(ds, ds')K_2(s, ds'')f(s'') \quad \text{by (ii)}$$

$$\leqslant \int_S\int_S (P_1\pi^{-1})(ds)K_2(s, ds'')f(s'') = (P_1\pi^{-1})(K_2f)$$

$$\leqslant \int_S\int_S P_2(ds)K_2(s, ds'')f(s'') = P_2K_2f \quad \text{by (i) and (iii).}$$

Extend to all n by induction. ■

COROLLARY. *If conditions* (i) *and* (ii) *of Theorem* 1 *hold, if the partial order is closed and if $(X_1(n), Y_1(n))$ and $X_2(n)$ have limiting distributions P_i^* as $n \to \infty$ for each i, then $P_1^*\pi^{-1} \leqslant P_2^*$.*

PROOF. Apply Theorem 1 with $P_1 = P_1^*$ and $P_2 = P_1^*\pi^{-1}$ so that condition (iii) holds. Then $P_1^*\pi^{-1} \leqslant P_2K_2^n$ for all n, so that $P_1^*\pi^{-1} \leqslant P_2^*$ by closure. ■

REMARKS. (1) Let $\leqslant$ for random variables represent the order relation applied to their distributions. Then condition (iii) states that $X_1(0) \leqslant X_2(0)$ and the conclusion states that $X_1(n) \leqslant X_2(n)$ for all n. Note that the stochastic order obtained is for the one-dimensional marginal distributions, not the joint distributions involving two or more values of n simultaneously.

(2) We obtain comparisons between two non-Markov processes by applying Theorem 1 twice with the same stochastically monotone Markov process in the middle.

(3) For two Markov processes on S, Theorem 1 reduces to Theorem 4.2.5a of Stoyan [16]. In fact, Theorem 1 can also be obtained as an application of Theorem 4.2.5a of [16] if, instead of projecting P_1 and K_1 onto S with π, we extended $\mathcal{F}$ and K_2 to $S \times S'$. This is done by defining an order on $\mathcal{P}(S \times S')$ by (1) with the set $\hat{\mathcal{F}}$ of all real-valued functions $\hat{f}$ on $S \times S'$ such that $\hat{f}(s, s') = f(s)$ for all $(s, s') \in S \times S'$, for some $f \in \mathcal{F}$. Corresponding to K_2, we define the Markov transition kernel $\hat{K}_2$ on $S \times S'$ by $\hat{K}_2((s, s'), A \times A') = K_2(s, A)1_{A'}(s')$ for all $(s, s') \in S \times S'$ and measurable subsets $A \subseteq S$ and $A' \subseteq S'$, where $1_{A'}(s')$ is the indicator function of the set A', i.e., $1_{A'}(s') = 1$ if $s' \in A'$ and 0 otherwise. With these definitions, it is not difficult to formulate a comparison result for the two Markov processes determined by K_1 and $\hat{K}_2$

with state space $S \times S'$ that is equivalent to Theorem 1. In this setting we can directly apply Theorem 4.2.5a of [16].

(4) In the special setting of Massey [7–10] in which the functions in $\mathscr{F}$ are indicator functions of increasing sets, Theorem 1 is also a discrete-time analogue of Theorem 3.5 of Massey [10]. The second half of Theorem 3.1 of [8] and Theorems 5.4, 6.4 and 6.5 of [9] are applications of Theorem 3.5 of [10]. Massey's work illustrates that Theorem 1 and the later results can be extended: the function π need not be a projection map.

The integral stochastic ordering $\leqslant$ on $\mathscr{P}(S)$ defined by (1) is rather special, being determined by the linear operation of integration with respect to functions in $\mathscr{F}$. The following example shows that Theorem 1 is not valid with arbitrary orderings on $\mathscr{P}(S)$, such as are discussed at the beginning of §4.2.1 of [16].

EXAMPLE 1. We show that Theorem 1 here and Theorem 4.2.5a of [16] are not correct with arbitrary orderings on $\mathscr{P}(S)$, even if both processes are Markov and the ordering on $\mathscr{P}(S)$ is a partial ordering. Let $S = \{0,1\}$ and let K_2 be the identity map, which corresponds to a stochastically monotone Markov process for any ordering on $\mathscr{P}(S)$. Let $K_1(s,\{0\}) = 1$ for all s. Let $P_1 \leqslant P_2$ mean that either (i) the cardinality of the support of P_1 is strictly greater than the cardinality of the support of P_2 or (ii) that the cardinality of the supports are equal and $P_1(\{0\}) \geqslant P_2(\{0\})$. It is easy to see that $\leqslant$ is a partial order on $\mathscr{P}(S)$, $K_1(s,\cdot) \leqslant K_2(s,\cdot)$ for all s, but $PK_1 > PK_2$ for all P with support S.

5. Continuous-time processes. As in §4.2 of [16] and the other sources, the comparison results for discrete-time processes extend easily to continuous-time processes when the continuous-time Markov processes are uniformizable jump processes. Uniformizable means that the rate of transitions out of any state is uniformly bounded. As in [15] and p. 69 of [16], the uniformizability assumption can often be subsequently relaxed.

In this section, let $X_i \equiv \{X_i(t), t \geqslant 0\}$ be a continuous-time jump stochastic process with the same general state space S for $i = 1, 2$. Let X_2 be a Markov jump process with exponential holding times in each state having means uniformly bounded away from zero (the uniformizability), stationary probability transition function $K_2(t)$ and transition rate function (infinitesimal generator) $Q_2 \equiv Q_2(s, A)$, i.e.,

$$K_2(h)(s, A) = P(X_2(t+h) \in A | X_2(t) = s) = hQ_2(s, A) + o(h) \tag{3}$$

for $s \notin A$ where $o(h)$ is a quantity that converges to zero after dividing by h for each s and A as $h \to 0$. Let (X_1, Y_1) be a continuous-time Markov jump process with state space $S \times S'$, exponential holding times having means uniformly bounded away from zero, probability transition function $K_1(t)$ and transition rate function $Q_1 \equiv Q_1((s, s'), A)$ for $A \subseteq S \times S'$.

The key to applying Theorem 1 is the following representation of $K_i(t)$ due to uniformization

$$K_i(t) = \sum_{n=1}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} (K_{i\lambda})^n, \tag{4}$$

where $K_{i\lambda}^n$ is the nth product of the discrete-time probability transition function

$$K_{i\lambda} = (I + \lambda^{-1} Q_i) \tag{5}$$

with I the identity map and λ sufficiently large so that $K_{i\lambda}$ is nonnegative for $i = 1$ and 2. One way to prove (4) and the following Theorem 2 is to simultaneously generate potential transitions of both processes with a Poisson process having intensity λ and appropriately select the real transitions of each process by independent thinning with probabilities determined by Q_i; see [15], [17], and references there. If we use the Poisson process to construct both (X_1, Y_1) and X_2, then it suffices to compare the two discrete-time processes embedded at the epochs of transitions of the Poisson process. These discrete-time processes have probability transition functions $K_{1\lambda}$ and $K_{2\lambda}$.

Hence, we have the following result as a direct consequence of Theorem 1.

THEOREM 2. *Let $\mathscr{P}(S)$ be endowed with an integral order relation $\leqslant$ determined by a set $\mathscr{F}$ of $\mathscr{P}(S)$-integrable functions. Let $K_{i\lambda}$ be the discrete-time probability transition functions in* (5) *obtained under the assumption of uniformizability. If*
(i) *$K_{2\lambda}$ is stochastically monotone,*
(ii) *$K_{1\lambda}((s, s'), \cdot)\pi^{-1} \leqslant K_{2\lambda}(s, \cdot)$ for all $(s, s') \in S \times S'$, and*
(iii) *$P_1\pi^{-1} \leqslant P_2$,*
then $P_1K_1(t) \leqslant P_2K_2(t)$ for all t.

REMARKS. (1) Conditions (i) and (ii) in Theorem 2 are relatively easy to check, because they involve the intensities Q_i instead of the continuous-time probability transition functions $K_i(t)$. For example, from (5) we see that (ii) holds if and only if

$$Q_1((s, s'), \cdot)\pi^{-1}f \leqslant Q_2(s, \cdot)f \tag{6}$$

for all $s \in S$, $s' \in S'$ and $f \in \mathscr{F}$. If $\mathscr{F}$ is the set of indicator functions of sets A in $\mathscr{A}$, then (6) reduces to

$$Q_1((s, s'), A \times S') \leqslant Q_2(s, A) \tag{7}$$

for all $s \in S$, $s' \in S'$ and $A \in \mathscr{A}$, which in turn is equivalent to

$$\begin{aligned} Q_1((s, s'), A \times S') &\leqslant Q_2(s, A), \qquad s \in A^c, \\ Q_1((s, s'), A^c \times S') &\geqslant Q_2(s, A^c), \qquad s \in A, \end{aligned} \tag{8}$$

for all $s \in S$, $s' \in S'$ and $A \in \mathscr{A}$.

(2) Condition (i) in Theorem 2 is also relatively easy to check, as is illustrated later in (14) in the proof of Theorem 5.

Stochastic monotonically for continuous-time kernels has several equivalent representations if, in addition to (3), we have

$$K_2(h)(s, \cdot)f = hQ_2(s, \cdot)f + o(h) \tag{9}$$

for each $s \in S$ and $f \in \mathscr{F}$. Formula (9) holds for example if (3) holds uniformly in A.

PROPOSITION 3. *If* (9) *holds, then the following are equivalent:*
(i) *$K_{2\lambda}$ in* (5) *is stochastically monotone for all λ sufficiently large,*
(ii) *$K_2(t)$ is stochastically monotone for all t,*
(iii) *$P_1Q_2(s, \cdot)f \leqslant P_2Q_2(s, \cdot)f$ for all $f \in \mathscr{F}$ whenever $P_1 \leqslant P_2$.*

PROOF. (i) implies (ii) by (4) and Proposition 2. If (ii) holds and $P_1 \leqslant P_2$, then

$$P_1(K_2(h)(s, \cdot) - I)f \leqslant P_2(K_2(h)(s, \cdot) - I)f + (P_2f - P_1f). \tag{10}$$

When we divide by h in (10) and let $h \to 0$, we obtain (iii). If (iii) holds and $P_1 \leqslant P_2$, then

$$P_1K_{2\lambda}(s,\cdot)f = P_1(I+\lambda^{-1}Q_2(s,\cdot))f = P_1f + \lambda^{-1}P_1Q_2(s,\cdot)f$$
$$\leq P_2f + \lambda^{-1}P_2Q_2(s,\cdot)f = P_2(I+\lambda^{-1}Q_2(s,\cdot))f = P_2K_{2\lambda}(s,\cdot)f. \quad \blacksquare$$

When (iii) of Proposition 3 holds, we say that Q_2 is stochastically monotone (although it is not an operator on $\mathscr{P}(S)$). The following elementary proposition is often convenient for establishing stochastic monotonicity. It shows that it is possible to consider different transitions separately. This approach has been exploited by Massey; see Theorem 4.1 of [10].

PROPOSITION 4. *Let Q and Q_i, $i \geq 1$, be transition rate functions of continuous-time jump processes on S. If*
(i) *Q is uniformizable,*
(ii) *Q_i is stochastically montone for each, and*
(iii) *$Q = \Sigma_i Q_i$,*
then Q is stochastically monotone.

REMARKS. (1) Massey has observed that in Proposition 4 it suffices for Q to be the generator of a contraction semigroup which is the strong operator limit of uniformizable monotone generators.

(2) Theorem 2 obviously has a corollary paralleling the corollary to Theorem 1.

6. A subset of probability measures. In this section we extend Theorem 1 to the situation in which the integral order relation $\leq$ is only defined on a subset $\mathscr{P}_s(S)$ of $\mathscr{P}(S)$. This may help because now the functions in $\mathscr{F}$ only need to be integrable with respect to all P in the subset $\mathscr{P}_s(S)$. For the Markov transition kernel K_2 to be stochastically monotone we now require that $PK_2 \in \mathscr{P}_s(S)$ for all $P \in \mathscr{P}_s(S)$ in addition to $P_1K_2 \leq P_2K_2$ for all P_1 and P_2 in $\mathscr{P}_s(S)$ with $P_1 \leq P_2$.

THEOREM 3. *Let the subset $\mathscr{P}_s(S)$ of $\mathscr{P}(S)$ be endowed with an integral order relation $\leq$ determined by a set $\mathscr{F}$ of $\mathscr{P}_s(S)$-integrable real-valued functions. If*
(i) *K_2 is stochastically monotone,*
(ii) *$(P_1K_1)\pi^{-1} \in \mathscr{P}_s(S)$ whenever $P_1\pi^{-1} \in \mathscr{P}_s(S)$,*
(iii) *$(P_1K_1)\pi^{-1} \leq (P_1\pi^{-1})K_2$ for all $P_1 \in \mathscr{P}(S \times S')$ with $P_1\pi^{-1} \in \mathscr{P}_s(S)$,*
(iv) *$P_1\pi^{-1} \in \mathscr{P}_s(S)$, $P_2 \in \mathscr{P}_s(S)$ and $P_1\pi^{-1} \leq P_2$,*
then $(P_1K_1^n)\pi^{-1} \in \mathscr{P}_s(S)$, $P_2K_2^n \in \mathscr{P}_s(S)$ and $(P_1K_1^n)\pi^{-1} \leq P_2K_2^n$ for all n.

REMARK. A sufficient condition for (iii) is condition (ii) of Theorem 1.

PROOF. Apply induction. The result is trivially true for $n = 0$ by (iv). Suppose the result has been established for n and consider $n + 1$. By induction, $(P_1K_1^n)\pi^{-1} \in \mathscr{P}_s(S)$ and $P_2K_2^n \in \mathscr{P}_s(S)$. For $f \in \mathscr{F}$,

$$[(P_1K_1^{n+1})\pi^{-1}]f = (P_1K_1^n)(K_1\pi^{-1})f$$
$$\leq (P_1K_1^n)\pi^{-1}(K_2f) \quad \text{by (ii) and (iii)}$$
$$\leq (P_2K_2^n)(K_2f) = (P_2K_2^{n+1})f$$

by (i) and the induction assumption. ∎

7. A traffic theory application. In this section we describe the traffic-theory application mentioned in the introduction. This application is investigated rather extensively in [18], so we will be brief here.

The problem is to describe the blocking when service is required from several facilities simultaneously. The model has n multi-server service facilities without extra waiting room and c customer classes. Service facility i has s_i servers. Customers from class j arrive according to a Poisson process with rate λ_j and immediately request service from one server at each facility in a subset A_j of the n service facilities. If all servers are busy in any of the required facilities, the request is blocked (lost without generating retrials). Otherwise, service begins immediately in all the required facilities. All servers working on a given customer from class j start and free up together. The service time for class j at all facilities has a general distribution with finite mean μ_j^{-1}. We assume that the c arrival processes and all the service times are mutually independent.

Let $b(A)$ be the probability that all servers are busy in at least one facility in subset A (at an arbitrary time in steady state). Thus $b(i) \equiv b(\{i\})$ is the probability that all servers are busy at facility i. Since Poisson arrivals see time averages [19], $b(A_j)$ is also the blocking probability for class j.

Let $B(s, \alpha)$ be the classical Erlang blocking formula associated with the M/G/s/loss service system with s servers and offered load α, defined by

$$B(s,\alpha) = (\alpha^s/s!) \Big/ \sum_{k=0}^{s} (\alpha^k/k!), \tag{11}$$

where, as usual, the offered load α is the arrival rate multiplied by the expected service time. Let C_i be the set of all classes that request service from facility i, i.e.,

$$C_i = \{ j : i \in A_j \}. \tag{12}$$

Let $\hat{\alpha}_i$ the offered load at facility i (not counting blocking elsewhere), defined by

$$\hat{\alpha}_i = \sum_{j \in C_i} \alpha_j, \tag{13}$$

where $\alpha_j = \lambda_j/\mu_j$ is the offered load of class j to the system as a whole.

In [18] we have applied the theory in the previous sections to show that a standard approximation for the blocking probability $b(A)$, which has long been regarded as conservative, is indeed an upper bound.

THEOREM 4. *For each subset A, $b(A) \leqslant 1 - \prod_{i \in A}(1 - B(s_i, \hat{\alpha}_i))$.*

To prove Theorem 4, we compare a continuous-time non-Markov process X_1 to a continuous-time Markov process X_2 using Theorem 2. First, however, we simplify the model. In Theorem 4 and Corollary 4.2 of [18] it is shown that the model possesses an insensitivity property: The stationary distribution of the number of customers of each class in service depends on the service-time distributions only through their means, and it depends on the arrival rates λ_j and service rates μ_j for each class only through their ratios $\alpha_j = \lambda_j/\mu_j$. (Also see Burman et al. [2] for the insensitivity result.) So henceforth we assume all service-time distributions are exponentially distributed with mean one.

The non-Markov process here is $X_1 \equiv \{(X_{11}(t), \ldots, X_{1n}(t)), t \geqslant 0\}$ where $X_{1i}(t)$ represents the number of busy servers at the facility i at time t. The additional information is contained in the stochastic process $Y_1 \equiv \{(Y_{11}(t), \ldots, Y_{1c}(t)), t \geqslant 0\}$ where $Y_{1j}(t)$ is the number of class j customers in service at time t. Obviously, Y_1 is a Markov process, so that (X_1, Y_1) is a Markov process too. The state spaces S and S' for X_1 and Y_1 are finite.

The Markov process to which we will compare X_1 is $X_2 \equiv \{(X_{21}(t), \ldots, X_{2n}(t)), t \geqslant 0\}$ where $X_{2i}(t)$ represents the number of busy servers in a standard $M/M/s_i$/loss system and the marginal processes are mutually independent. We have proved Theorem 4 by establishing the following more general stochastic comparison. Only the comparison of the stationary distributions, part (b) below, is discussed in [18]. Let X_i^* have the stationary distribution on S of X_i for $i = 1, 2$. (For $i = 1$, existence and uniqueness follows from Theorem 4 of [18].)

THEOREM 5. (a) *If* $P(X_1(0) \leq \mathbf{k}) \geq P(X_2(0) \leq \mathbf{k})$ *for all* $\mathbf{k} \equiv (k_1, \ldots, k_n)$, *then* $P(X_1(t) \leq \mathbf{k}) \geq P(X_2(t) \leq \mathbf{k})$ *for all* $\mathbf{k}$ *and* t.

(b) *For any n-tuple* $\mathbf{k} = (k_1, \ldots, k_n)$, $P(X_1^* \leq \mathbf{k}) \geq P(X_2^* \leq \mathbf{k})$.

OUTLINE OF THE PROOF AND DISCUSSION. We apply Theorem 2 using the partial order on $\mathscr{P}(S)$ generated by the class $\mathscr{F}$ of indicator functions of sets in the class $\mathscr{A}$ of all lower subsets of S; i.e., $A \in \mathscr{A}$ if $A = \{\mathbf{k}' \in S: \mathbf{k}' \leq \mathbf{k}\}$ for some $\mathbf{k} \in S$, with the usual partial order relation on R^n. This is the weak* order, denoted by $\langle \cdot \rangle^*$, in Massey [10]. It is well known that these lower sets are a determining class for $\mathscr{P}(R^n)$, so that this order relation is a partial order. It is also well known that this partial order is strictly weaker than the standard stochastic order in [4]. Moreover, Example 6 of [18] shows that the extension of Theorem 5(b) to stochastic order based on all increasing sets is not true, so that there is indeed a need for a different approach.

Since S is finite, weak convergence of probability measures is equivalent to convergence for all subsets. Hence, the partial order relation is closed, and part (b) follows from part (a). (A direct proof of (b) is given in [18].)

Finally, it is relatively easy to verify conditions (i) and (ii) of Theorem 2. For example,

$$PK_{2\lambda}(\{\mathbf{k}': \mathbf{k}' \leq \mathbf{k}\}) = \sum_i p_i^{\pm} P(\{\mathbf{k}': \mathbf{k}' \leq \mathbf{k} \pm \mathbf{e}_i\}) = \left(1 - \sum_i p_i^{\pm}\right) P(\{\mathbf{k}': \mathbf{k}' \leq \mathbf{k}\}) \tag{14}$$

where $\mathbf{e}_i$ is an n-tuple of all 0's except a 1 in the ith place and $p_i^{\pm}$ is a probability. (The permissible values of $\mathbf{e}_i$ depend on $\mathbf{k}$.) From (14), it is immediate that $K_{2\lambda}$ is stochastically monotone with respect to the lower-set ordering. Condition (ii) is easily established by verifying (8); see [18] for more details. By Proposition 4, it suffices to consider the transitions due to arrivals and departures separately. ∎

Acknowledgements. My work on the traffic-theory application was motivated first by discussions with D. D. Sheng and then again by discussions with D. P. Heyman. Sheng and D. R. Smith obtained a weaker bound on the blocking probability $b(A)$ in the case of two facilities by different methods; see the appendix of [12]. I am also grateful to W. A. Massey for a preliminary version of [10] and his comments on this work.

References

[1] Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.

[2] Burman, D. Y., Lehoczky, J. P. and Lim Y. (1984). Insensitivity of Blocking Probabilities in a Circuit-Switching Network. *J. Appl. Probab.* **21** 850–859.

[3] Daley, D. J. (1968). Stochastically Monotone Markov Chains. *Z. Warsch. Verw. Gebiete* **10** 305–317.

[4] Kamae, T., Krengel, U. and O'Brien, G. L. (1977). Stochastic Inequalities on Partially Ordered Spaces. *Ann. Probab.* **5** 899–912.

[5] Keilson, J. and Kester, A. (1977). Monotone Matrices and Monotone Markov Processes. *Stochastic Process. Appl.* **5** 231–241.

[6] Kester, A. (1977). Preservation of Cone Characterizing Properties of Markov Chains. Ph.D. dissertation, University of Rochester.

[7] Massey, W. A. (1984a). An Operator Analytic Approach to the Jackson Network. *J. Appl. Probab.* **21** 379–393.
[8] ______. (1984b). Open Networks of Queues: Their Algebraic Structure and Estimating their Transient Behavior. *Adv. in Appl. Probab.* **16** 176–201.
[9] ______. (1985a). A Family of Bounds for the Transient Behavior of a Jackson Network. *J. Appl. Probab.* **22**, 543–549.
[10] ______. (1985b). Stochastic Orderings for Markov Processes on Partially Ordered Spaces. *Math. Oper. Res.*
[11] Mitra, D. and Weinberger, P. J. (1984). Probabilistic Models of Database Locking: Solutions, Computational Algorithms and Asymptotics. *J. Assoc. Comput. Mach.* **31** 855–878.
[12] Monma, C. L. and Sheng, D. D. (1985). Backbone Network Design and Performance Analysis: A Methodology for Packet Switching Networks. unpublished work.
[13] Sheng, D. D. (1983). Performance Analysis Methodology for Packet Network Design. *IEEE Global Telecommunications Conf. GLOBECOM '83*, **13.5.1–13.5.5**, 456–460.
[14] Smith, D. R. and Whitt, W. (1981). Resource Sharing for Efficiency in Traffic Systems. *Bell System Tech. J.* **60** 39–55.
[15] Sonderman, D. (1980). Comparing Semi-Markov Processes. *Math. Oper. Res.* **5** 110–119.
[16] Stoyan, D. (1983). *Comparison Methods for Queues and Other Stochastic Models*. D. J. Daley (Ed.), Wiley, New York.
[17] Whitt, W. (1981). Comparing Counting Processes and Queues. *Adv. in Appl. Probab.* **13** 207–220.
[18] ______. (1985). Blocking When Service Is Required from Several Facilities Simultaneously. *AT&T Bell Lab. Tech. J.* **64**, 1807–1856.
[19] Wolff, R. W. (1982). Poisson Arrivals See Time Averages. *Oper. Res.* **30** 223–231.

AT & T BELL LABORATORIES (HO 3K304), HOLMDEL, NEW JERSEY 07733