The Annals of Applied Probability
1995, Vol. 5, No. 1, 49–77

ON POSITIVE HARRIS RECURRENCE OF MULTICLASS QUEUEING NETWORKS: A UNIFIED APPROACH VIA FLUID LIMIT MODELS[1]

BY J. G. DAI

Georgia Institute of Technology

It is now known that the usual traffic condition (the nominal load being less than 1 at each station) is *not* sufficient for stability for a multiclass open queueing network. Although there has been some progress in establishing the stability conditions for a multiclass network, there is no unified approach to this problem. In this paper, we prove that a queueing network is positive Harris recurrent if the corresponding fluid limit model eventually reaches zero and stays there regardless of the initial system configuration. As an application of the result, we prove that single class networks, multiclass feedforward networks and first-buffer–first-served preemptive resume discipline in a reentrant line are positive Harris recurrent under the usual traffic condition.

1. Introduction. We consider a network composed of d single server stations, which we index by $i = 1, \dots, d$. The network is populated by K classes of customers, and each class k has its own exogenous arrival process with interarrival times $\{\xi_k(n), n \geq 1\}$. We allow $\xi_k(n) = \infty$ for all n for some class k. In this case, the external arrival process for class k is null. We let $\mathscr{E}$ denote the set of classes with nonnull exogenous arrivals. Hereafter, whenever external arrival processes are under discussion, only classes with nonnull exogenous arrivals are considered. Class k customers require service at station $s(k)$, and their service times are $\{\eta_k(n), n \geq 1\}$. Upon completion of service at station $s(k)$, a class k customer becomes a customer of class l with probability P_{kl} and exits the network with probability $1 - \sum_l P_{kl}$, independent of all previous history. To be more precise about the last statement, let $\phi^k(n)$ be the routing vector for the nth class k customer who finishes service at station $s(k)$. The lth component of $\phi^k(n)$ is 1 if this customer becomes a class l customer and zero otherwise. Therefore, $\phi^k(n)$ is a K-dimensional "Bernoulli random variable" with parameter P'_k, where P_k denotes the kth row of $P = (P_{kl})$ and primes denote transpose. (All vectors are envisioned as column vectors.) We assume that $\phi^k = \{\phi^k(n), n \geq 1\}$ is iid, and $\phi^1, \dots, \phi^K$ are independent and are independent of the arrival processes and service processes. Such a routing mechanism is often called Bernoulli routing. The transition matrix $P = (P_{kl})$

Received October 1993; revised May 1994.
[1]Research supported in part by NSF Grants DMS-92-03524, DDM-92-15233, and two grants from the Texas Instruments Corporation.
AMS 1991 *subject classifications.* Primary 60K25, 90B22; secondary 60K20, 90B35.
Key words and phrases. Multiclass queueing networks, Harris positive recurrent, stability, fluid approximation.

is taken to be transient. That is,

$$I + P + P^2 + \cdots \text{ is convergent.} \tag{1.1}$$

Condition (1.1) implies that the expected number of visits to class l by a class k customer is finite. Therefore, all customers eventually leave the network. Hence the networks we are considering are open queueing networks. We assume that the waiting buffer at each station has infinite capacity. To avoid trivial complications, we assume that no two events can happen simultaneously. Hereafter, we will refer to such a network as a multiclass open queueing network, or simply a *multiclass network*. If there is only one class, say class 1, that has nonnull exogenous arrivals and the entries of the routing matrix take the form $P_{k,k+1} = 1$ for $k = 1, \dots, K-1$ and zero for all others, then the multiclass network is called a *reentrant line*; see [32]. An example of such a network is given at the end of this section.

So far, the sequences of interarrival times and service times are quite general. We need to put some assumptions on them. Throughout this paper, we make the following three assumptions on the network. We first assume that

$$\begin{array}{l}\xi_1, \dots, \xi_K, \eta_1, \dots, \eta_K \text{ are iid sequences and they} \\ \text{are mutually independent.}\end{array} \tag{1.2}$$

This independence assumption may be relaxed for some networks; see the remark following Proposition 2.1. Next, we put some moment assumptions on interarrival and service times. We assume that

$$\begin{aligned} &\mathsf{E}[\eta_k(1)] < \infty \quad \text{for } k = 1, \dots, K, \\ &\mathsf{E}[\xi_k(1)] < \infty \quad \text{for } k \in \mathscr{E}. \end{aligned} \tag{1.3}$$

Finally, we assume that interarrival times are unbounded and spread out. That is, for each $k \in \mathscr{E}$, there exist some integer $j_k > 0$ and some function $p_k(x) \geq 0$ on $\mathbb{R}_+$ with $\int_0^\infty p_k(x)\,dx > 0$, such that

$$\mathsf{P}\{\xi_k(1) \geq x\} > 0 \quad \text{for any } x > 0 \tag{1.4}$$

and

$$\mathsf{P}\left\{a \leq \sum_{i=1}^{j_k} \xi_k(i) \leq b\right\} \geq \int_a^b p_k(x)\,dx \quad \text{for any } 0 \leq a < b. \tag{1.5}$$

Let $C_i = \{k : s(k) = i\}$. The set C_i is called the *constituency* for station i. Let C be a $d \times K$ incidence matrix,

$$C_{ik} = \begin{cases} 1, & \text{if } s(k) = i, \\ 0, & \text{otherwise.} \end{cases} \tag{1.6}$$

Let $\alpha_k = 1/\mathsf{E}[\xi_k(1)]$ and $m_k = \mathsf{E}[\eta_k(1)]$ be the arrival rate and mean service time for class k customers, respectively. In light of assumption (1.1), $(I - P')^{-1}$ exists and

$$(I - P')^{-1} = (I + P + P^2 + \cdots)'.$$

Put $\lambda = (I - P')^{-1}\alpha$. One interprets λ_k as the *effective* arrival rate to class k. For each $i = 1, \ldots, d$ let

$$\rho_i = \sum_{k \in C_i} \lambda_k m_k. \tag{1.7}$$

We call ρ_i the *nominal workload* for server i per unit of time. In vector form, we have $\rho = CM\lambda$, where

$$M = \operatorname{diag}(m_1, \ldots, m_K). \tag{1.8}$$

The main objective of this paper is to establish a sufficient condition for the stability of a multiclass open queueing network under a variety of queueing disciplines. A queueing discipline at station i dictates which customer will be served next when server i completes a service. We say that a queueing discipline for a multiclass network is *stable* if the underlying Markov process describing the network dynamics is *positive Harris recurrent*, whose precise definition is given in Section 3. When there is no ambiguity in the underlying queueing discipline, we say that a queueing network is stable if the queueing discipline under discussion is stable. Our main theorem (Theorem 4.2) states that a queueing network is stable if the corresponding fluid limit model is stable. The latter phrase means that the fluid limit model eventually reaches zero and stays there regardless of the initial system configuration. To make the theorem practical, we give a systematic treatment of fluid limit models. As an application of the theorem, we prove that single class networks, multiclass feedforward networks and first-buffer–first-served (FBFS) preemptive resume discipline in a reentrant line are positive Harris recurrent under the usual traffic condition

$$\rho_i < 1, \qquad i = 1, \ldots, d, \tag{1.9}$$

or in vector form $\rho < e$, where $\rho = (\rho_1, \ldots, \rho_d)'$ and e is the d-dimensional vector of ones. (Vector inequalities are interpreted componentwise.)

Our theorem is a generalization of recent work by Rybko and Stolyar [47], who mainly considered the stability of multiclass Markovian networks with *discrete* state space. Their brilliant analysis, via fluid models, leads to the stability of a two station multiclass network under the first-in–first-out (FIFO) queueing discipline. They also showed that under a certain "bad" priority discipline, the two-station network was *not* stable under (1.9). However, in their paper, convergence to a fluid model was never addressed. In other words, they did not consider the kind of *fluid limit* that we do here. Rather, they worked with the formal *fluid analog* of the queueing network. Therefore, they needed a parallel analysis for the original network even if they proved a certain property for the fluid model. As far as the stability of a queueing network is concerned, our theorem is attractive in that readers can focus on the fluid limit itself without explicitly going back to the original queueing network. This work was initially motivated by recent work of Dupuis and Williams [17], where positive recurrence for d-dimensional reflecting Brownian motion in an orthant was investigated. By constructing a Lyapunov function, they showed that a semimartingale reflecting Brownian motion in an orthant is positive recurrent if all solutions of the corresponding deterministic Skorohod problem for a driving noiseless drift path are stable. In a similar philosophical vein, Malyshev has obtained conditions for positive recurrence of reflected random walks in the orthant by studying related deterministic dynamical systems; see [36] and the references therein.

Our description of a multiclass network is now quite standard. (The class of queueing networks described here is in fact an important special case of the setup in [24].) In his pioneering paper on queueing networks, Jackson [26] assumed that customers visiting or occupying any given station are essentially indistinguishable from one another, and that a customer completing service at station i will move next to station j with some fixed probability P_{ij}, independent of all previous history. Thus in Jackson's networks, each station serves a single customer class; hence, these networks have been called *single-class* networks. Jackson's model was extended by Baskett, Chandy, Muntz and Palacios [2] and Kelly [30] to the multiclass setting. In both cases, they assumed either special types of service disciplines or exponential distributions for interarrival and service times.

For a long time, research on stability for open queueing networks under general distributional assumptions has mostly been restricted to single class networks or generalized Jackson networks; see [4, 50, 38, 1, 7]. However, recently stability conditions for multiclass networks have received a lot of attention. In the *deterministic* setting, Kumar and his coauthors did the pioneering work on the stability of general multiclass networks under a variety of queueing disciplines, see [44, 34, 35, 32]. In particular, they proved that FBFS and last-buffer–first-served (LBFS) are stable under (1.9) in a deterministic reentrant line. They also found that $\rho_i < 1$ for all i is *not* sufficient for stability under certain priority disciplines. Rybko and Stolyar [47] found that this phenomenon still exists for stochastic networks. Even more surprising, Bramson [6] found a two station multiclass stochastic network, where FIFO discipline is not stable even if (1.9) holds. Seidman [48], independently, found a deterministic example showing a similar phenomenon. These examples demonstrate that to obtain a general criterion in terms of basic system parameters like (1.9) for stability of a multiclass network is very difficult. On the positive side, Kumar and Meyn [33] recently gave some sufficient conditions for the stability of multiclass networks with Poisson arrivals and exponential service times. Their sufficient condition, in terms of the solution to a linear program, implies, among other things, that moments of some processes associated with the network converge as time $t \to \infty$. More recently, while this paper was under review, Dai and Weiss [13] and Down and Meyn [16] showed that piecewise Lyapunov functions, initiated by Botvitch and Zamyatin [5], are powerful tools in characterizing stability regions for certain priority disciplines.

For readers who are not familiar with the recent literature on multiclass networks, let us consider an example of a two station multiclass network

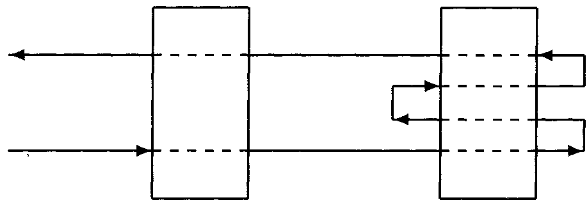

FIG. 1. *A variant of Bramson's two station multiclass network.*

pictured in Figure 1. It is a reentrant line. There is one type of customer arriving at station 1 from the outside according to a Poisson process with rate 1. An arriving customer visits stations 1, 2, 2, 2, 2, 1 and exits. We call customers in the kth stage of visit class k customers. The variants of this two station model were studied by Kumar [32], Whitt [51] and Bramson [6]. Obviously,

$$\rho_1 = m_1 + m_6, \qquad \rho_2 = m_2 + m_3 + m_4 + m_5.$$

Assume that all service times are exponential and $m_6 = 0.899$, $m_2 = 0.897$ and $m_1 = m_3 = m_4 = m_5 = 0.001$. Therefore the nominal loads are $\rho_1 = 0.90$ and $\rho_2 = 0.90$. One would hope that such a network is stable under FIFO discipline and that the long run average queue lengths are finite. Let T_n be the throughput time (total time in the network) of the nth customer. Let

$$\overline{T}_n = \frac{1}{n} \sum_{i=1}^{n} T_i.$$

We simulated this network starting from empty for one replication, which corresponds to some particular choice of seeds for random number generators. In this replication, we let n vary from 1000 to 1,000,000. The average throughput times (rounded to integers) are tabulated in the "exponential" row of Table 1. It seems that as n increases, $\overline{T}_n$ increases almost linearly. We conjecture that

$$\overline{T}_n \to \infty \quad \text{almost surely}$$

as $n \to \infty$. If we replace all exponential random variables by constants (their respective means), a similar phenomenon occurs, as evidenced in the "deterministic" row of Table 1. Figure 2 plots the queue lengths at both stations in 100 and 1000 units of simulation time. Clearly, the queue lengths oscillate with increasing magnitude. Most of the time, only one of the two servers is working. This mutual blocking apparently causes the instability. Readers are referred to [6] for more insight.

The following notation will be used in the rest of this paper. For a finite set A, $|A|$ denotes the cardinality of A. Let $\mathbb{R}_+ = [0, \infty)$. The set of nonnegative integers is $\mathbb{Z}_+$. $D_{\mathbb{R}}[0, \infty)$ is the space of right continuous functions on $\mathbb{R}_+$ having left limits on $(0, \infty)$, endowed with the Skorohod topology; see Ethier

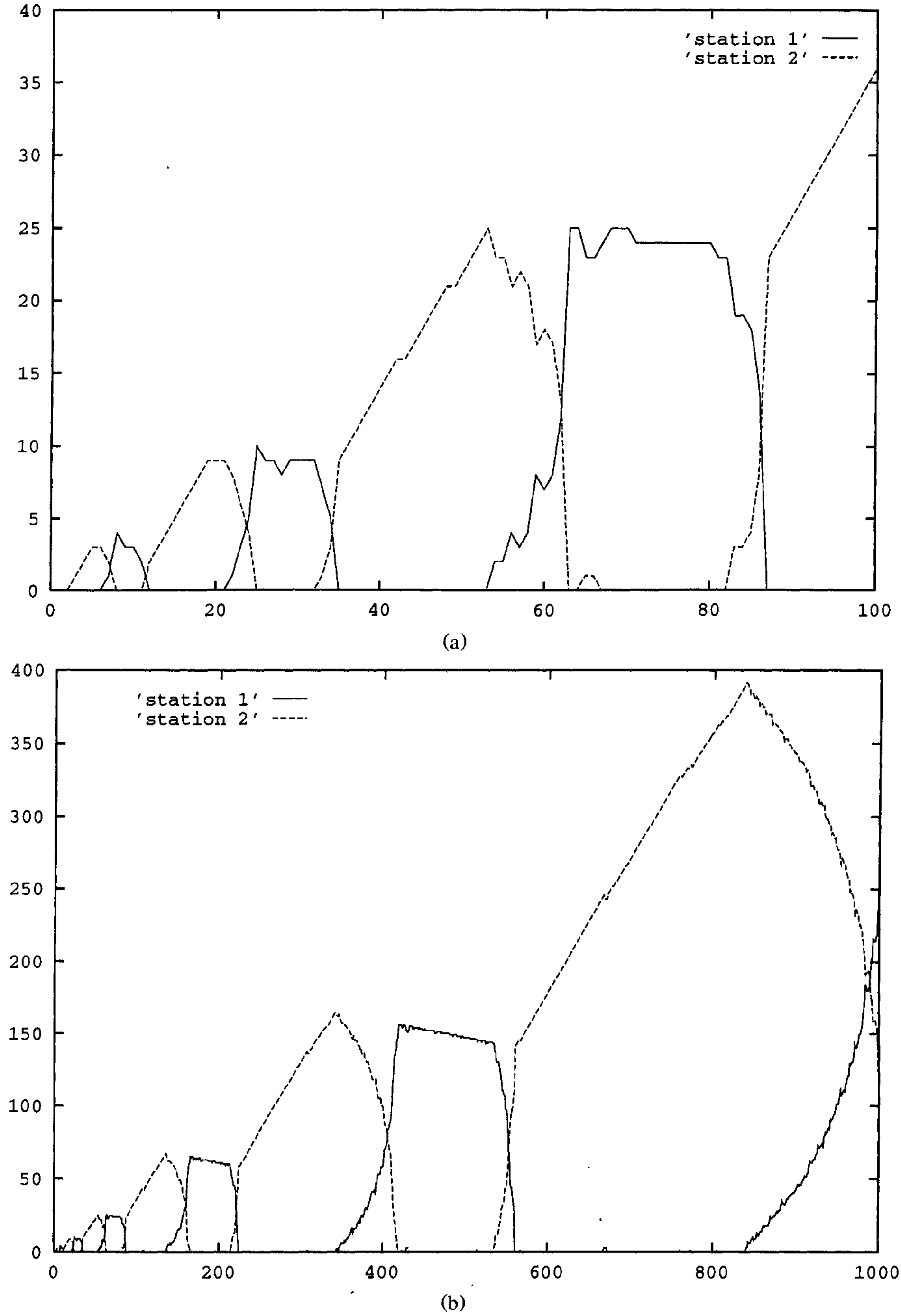

FIG. 2. *Queue lengths in* 100 *and* 1000 *units of time for the deterministic network.*

TABLE 1
Simulation estimates of average throughput time $\overline{T}_n$

n	1000	5000	10,000	50,000	100,000	1,000,000
Exponential	252	1352	2848	15,270	27,284	267,353
Deterministic	258	1346	2881	14,711	29,138	244,480

and Kurtz [18]. Let $C_{\mathbb{R}}[0,\infty)$ be the subset of continuous paths in $D_{\mathbb{R}}[0,\infty)$. The following fact is useful in this paper. Assume that $f_n \in D_{\mathbb{R}}[0,\infty)$ for each n and $f \in C_{\mathbb{R}}[0,\infty)$. Then $f_n \to f$ in the Skorohod topology if and only if for each $t > 0$,

$$\sup_{0 \le s \le t} |f_n(s) - f(s)| \to 0 \quad \text{as } n \to \infty. \tag{1.10}$$

When (1.10) holds for all $t \ge 0$, we say that $f_n \to f$ *uniformly on compact sets*, or simply $f_n \to f$ u.o.c. The symbol $\Rightarrow$ denotes weak convergence or convergence in distribution.

2. Preliminaries.

2.1. *State descriptions.* We now define a Markov process $X = \{X(t), t \ge 0\}$ that describes the dynamics of the queueing network. If all the interarrival time and service time distributions are exponential, then queue lengths, in addition to the order of customer arrivals in FIFO discipline, are sufficient for a state description. In general, we need to augment the previous state description with the remaining exogenous interarrival times and the remaining service times. Therefore, in general, the state $X(t)$ at time t is

$$X(t) = (\mathbb{Q}(t), U(t), V(t)),$$

where $\mathbb{Q}(t)$ captures how customers are lined up at each station, $U(t) = (U_k(t)\colon k \in \mathscr{E})' \in \mathbb{R}_+^{|\mathscr{E}|}$ and $V(t) = (V_1(t), \dots, V_K(t))' \in \mathbb{R}_+^K$. For $k \in \mathscr{E}$, $U_k(t)$ is the remaining time before the next class k customer will arrive from outside. For $k = 1, \dots, K$, $V_k(t)$ is the remaining service time for the class k customer that is in service, which is set to be zero if $Q_k(t) = 0$. Both $U(t)$ and $V(t)$ are taken to be right continuous. Let X be the set of all possible states that X can take. We use $|\mathbb{Q}(t)|$ to denote the total queue length in the network at time t. For a $u \in \mathbb{R}^K$, the norm of u is $|u| = \sum_{k=1}^K |u_k|$. We define the norm $|X(t)|$ of $X(t)$ to be the total queue length plus the total remaining interarrival time and total remaining service time at t. That is,

$$|X(t)| = |\mathbb{Q}(t)| + |U(t)| + |V(t)|.$$

The precise state description depends on a queueing discipline. Here we give some examples.

FIFO discipline. Under the FIFO discipline, customers at each station are served on the first-in–first-out basis. For station i define

$$\mathbb{Q}_i(t) = (k_{i,1}, k_{i,2}, \dots, k_{i,N_i(t)}), \tag{2.1}$$

where $k_{i,j}$ is the class number for the jth customer at station i and $N_i(t)$ is the queue length (including possibly the one being served) at station i. [If $N_i(t) = 0$, $\mathbb{Q}_i(t)$ is defined to be an empty list.] Then we can take

$$\mathbb{Q}(t) = (\mathbb{Q}_1(t), \dots, \mathbb{Q}_d(t)).$$

At each station i with $N_i(t) > 0$, there is only one customer that can receive partial service. Thus,

$$X(t) = (\mathbb{Q}(t), U(t), (V_{k_{i,1}(t)}(t), i = 1, \dots, d))$$

is the state at time t. Let $\mathbb{Z}_K = \{1, 2, \dots, K\}$ and $\mathbb{Z}_K^\infty$ be the set of finitely terminated sequences taking values in $\mathbb{Z}_K$. (Zero length sequence corresponds to the situation where there are no customers at the station.) It is evident that

$$\mathbb{Q}_i(t) \in \mathbb{Z}_K^\infty.$$

Hence

$$\mathsf{X} \subset (\mathbb{Z}_K^\infty)^d \times \mathbb{R}_+^{|\mathscr{E}|+d}.$$

We assume that X is endowed with the natural induced topology from $(\mathbb{Z}_K^\infty)^d \times \mathbb{R}_+^{|\mathscr{E}|+d}$.

Priority disciplines. Under priority disciplines, customers at each station are ranked according to their class designations. Classes to be served at station i are ranked as $|C_i|, \dots, 1$. The class with rank $|C_i|$ has the highest priority and the class with rank 1 has the lowest priority. When the server finishes service of a customer, the server picks a customer from a nonempty class with the highest rank. If there is no customer at the station, the server stays idle. To be concrete, within each class, customers are served in FIFO order. In the case that ties among classes are allowed, all customers with the same rank are served in FIFO order. There are two different modes for a priority discipline. In nonpreemptive mode, the server has to finish service of the current customer being served before choosing another customer to work on. In preemptive resume mode, when an arriving customer has rank higher than the one currently being served, the customer in service is preempted immediately until all customers with higher ranks are served. At that point, the server recommences service of the preempted customer until service is completed or a higher priority customer arrives. For preemptive resume priority discipline, we can take $\mathbb{Q}(t)$ as

$$\mathbb{Q}(t) = Q(t) = (Q_1(t), \dots, Q_K(t))' \tag{2.2}$$

and

$$X(t) = (\mathbb{Q}(t), U(t), V(t)),$$

where $Q_k(t)$ is the queue length for class k customers at time t. In this case we have

$$\mathsf{X} \subset \mathbb{Z}_+^K \times \mathbb{R}_+^{|\mathscr{E}|+K}.$$

For nonpreemptive priority discipline, at each station there is at most one customer that can receive partial service. Thus,

$$X(t) = (Q(t), U(t), k_i(t), V_{k_i(t)}(t), i = 1, \dots, d),$$

where $k_i(t)$ is the class number of the customer being served at station i at time t. If there are no customers at station i at time t, $k_i(t)$ is set to be zero and $V_{k_i(t)}(t) = 0$. In the nonpreemptive case,

$$\mathsf{X} \subset \mathbb{Z}_+^K \times \mathbb{R}_+^{|\mathscr{E}|+d} \times \mathbb{Z}_K^d.$$

Head-of-line processor sharing discipline. With this discipline, the lead customer in all customer classes visiting a station simultaneously receives service. Hence, we can take

$$X(t) = (\mathbb{Q}(t), U(t), V(t)),$$

where $\mathbb{Q}(t) = Q(t)$ as in (2.2). The state space

$$\mathsf{X} \subset \mathbb{Z}_+^K \times \mathbb{R}_+^{|\mathscr{E}|+K}.$$

Processor sharing discipline. With a processor-sharing discipline, all customers visiting a station simultaneously receive service. Let $\tilde{V}_{i,j}(t)$ be the remaining service time for the jth customer at station i at time t. [If $N_i(t) = 0$, $\tilde{V}_{ij}(t)$ is taken to be zero. We can envision that customers at each station are lined up according to the FIFO discipline.] The state is

$$X(t) = (\mathbb{Q}(t), U(t), \tilde{V}(t)),$$

where $\mathbb{Q}(t)$ is defined in (2.1), $\tilde{V}(t) = (\tilde{V}_1(t), \dots, \tilde{V}_d(t))$ and

$$\tilde{V}_i(t) = (\tilde{V}_{i,1}(t), \dots, \tilde{V}_{i,N_i(t)}(t)).$$

The state space

$$\mathsf{X} \subset (\mathbb{Z}_K^\infty)^d \times \mathbb{R}_+^{|\mathscr{E}|} \times (\mathbb{R}_+^\infty)^d,$$

where $\mathbb{R}_+^\infty$ is the space of all finitely terminated sequences in $\mathbb{R}_+$. The norm of $X(t)$ is still the total queue length plus the total remaining interarrival time and the total remaining service time.

REMARK. It turns out that the processor sharing discipline or any disciplines in which the number of customers receiving simultaneous service is unbounded will slightly alter the fluid limit model discussed in Section 4. Specifically, the statement and the proof of Theorem 4.1 need some modifications. However, all results in Section 3 are still valid without any modification. In order to keep a clean treatment in the paper, we shall *not* include these queueing disciplines in the discussion. However, extensions of Theorem 4.2 to cover these disciplines should be evident.

2.2. *Strong Markov property.* To establish that $X = \{X(t),\ t \geq 0\}$ is a strong Markov process, we follow Sections 2.3 and 2.4 of Kaspi and Mandelbaum [28]. Because of the independence assumption (1.2), we can check that X is Markov for each of the disciplines discussed in Section 2.1. As time t goes on, $U_k(t)$ and $V_k(\mathrm{t})$ decrease while the remainder of the state remains constant. As one of these residual processes reaches zero, a jump occurs for X. Hence $X = \{X(t),\ t \geq 0\}$ is a piecewise-deterministic Markov (PDM) process that conforms to Assumption 3.1 of Davis [15]. The following proposition follows from Davis ([15], page 362).

PROPOSITION 2.1. *$X = \{X(t),\ t \geq 0\}$ is a strong Markov process with state space* X.

REMARK. Condition (1.2) is mainly used to establish the Markov property for X. In certain networks, this condition can be relaxed—for example, for the reentrant line presented at the end of Section 1, we can weaken the assumption on the service times. For that network, it is enough to assume that $\{(\eta_1(n), \dots, \eta_K(n)),\ n \geq 1\}$ is iid. For fixed n, $\eta_1(n)$, ..., $\eta_K(n)$ may have arbitrary dependency. This feature is useful for certain applications, notably in computer communications and manufacturing systems. There the length of a computer message or the size of a manufacturing lot may be random. However, the service times in general are proportional to the message length or lot size, and therefore are positively correlated.

Let $\mathcal{B}_{\mathsf{X}}$ be the Borel σ-field of X. Having verified that X is a PDM, as in Section 2.4 of Kaspi and Mandelbaum [28], we may assume at our disposal the usual elements that constitute a Markovian environment for X. Formally, it will be assumed hereafter that $(\Omega, \mathcal{F}, \mathcal{F}_t, X_t, \theta_t, P_x)$ is a Borel right process on the measurable state space $(\mathsf{X}, \mathcal{B}_{\mathsf{X}})$. In particular, $X = \{X(t), t \geq 0\}$ has right-continuous sample paths, it is defined on $(\Omega, \mathcal{F})$ and is adapted to $\{\mathcal{F}_t, t \geq 0\}$, $\{\mathsf{P}_x,\ x \in \mathsf{X}\}$ are probability measures on $(\Omega, \mathcal{F})$ such that for all $x \in \mathsf{X}$,

$$\mathsf{P}_x\{X(0) = x\} = 1,$$

and

$$\mathsf{E}_x\left\{f(X \circ \theta_\tau) \mid \mathcal{F}_\tau\right\} = \mathsf{E}_{X(\tau)} f(X) \quad \text{on } \{\tau < \infty\}, \mathsf{P}_x\text{-a.s.},$$

where τ is any F_t-stopping time

$$(X \circ \theta_\tau)(\omega) = \{X(\tau(\omega)+t, \omega), t \geq 0\}$$

and f is any real-valued bounded measurable function (the domain of f is the space of X-valued right-continuous functions on $[0, \infty)$, equipped with the Kolmogorov σ-field generated by cylinders). Let $P^t(x, A)$, $A \in \mathrm{B}_{\mathrm{X}}$, be the transition probability of X. That is,

$$P^t(x, A) = \mathsf{P}_x(X(t) \in A).$$

A nonzero measure π on $(\mathrm{X}, \mathrm{B}_{\mathrm{X}})$ is *invariant* for X if π is σ-finite, and

$$\pi(A) = \int_S P^t(x, A)\, \pi(dx) \quad \text{for all } A \in \mathrm{B}_{\mathrm{X}}.$$

An invariant measure π is said to be unique if the only invariant measures for X are positive scalar multiples of π. We denote $\mathrm{X}^0 = \{x \in \mathrm{X} : |x| > 0\}$.

3. Harris positive recurrent chains. Let $\tau_A = \inf\{t \geq 0\colon X_t \in A\}$. The process X is *Harris recurrent* if there exists some σ-finite measure μ on $(\mathrm{X}, \mathrm{B}_{\mathrm{X}})$, such that whenever $\mu(A) > 0$ and $A \in \mathrm{B}_{\mathrm{X}}$,

$$\mathsf{P}_x\{\tau_A < \infty\} \equiv 1.$$

It is well known that if X is Harris recurrent, then an essentially unique invariant measure π exists; see, for example, [21]. If the invariant measure is finite, then it may be normalized to a probability measure. In this case X is called *positive Harris recurrent*.

REMARK 1. Harris recurrence implies an apparently stronger condition that

$$\mathsf{P}_x\left\{\int_0^\infty 1_A(X(s))\, ds = \infty\right\} \equiv 1,$$

whenever $\mu'(A) > 0$ for some σ-finite measure μ'. See [29] or [40], Theorem 1.1.

REMARK 2. Assume that $X = \{X(t),\ t \geq 0\}$ is positive Harris recurrent with the unique stationary probability distribution π. For any measurable function f on $(\mathrm{X}, \mathrm{B}_{\mathrm{X}})$, let

$$\pi(f) = \int_{\mathrm{X}} f(x)\, \pi(dx)$$

whenever the integral makes sense. It was proved in [12] that positive recurrence implies the following ergodic property: for every measurable f on X with $\pi(|f|) < \infty$,

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t f(X(s))\, ds = \pi(f), \quad \mathsf{P}_x\text{-a.s. for each } x \in \mathrm{X}.$$

See also Theorem 17.0.1(i) of [39] and Section 2.5C of [28]. Take $f(x)$ to be the number of class k customers when the state is in x. Because f is nonnegative, positive Harris recurrence for X implies that each customer class k,

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t Q_k(s)\,ds \to \pi(f) \quad \mathsf{P}_x\text{-a.s. for each } x \in \mathsf{X},$$

where $Q_k(t) = f(X(t))$ is the class k queue length at time t. Note that in general $\pi(f)$ may be infinite.

Suppose that a is a general probability distribution on $\mathbb{R}_+$, and define the Markov transition function K_a as

$$K_a(x, A) \equiv \int_0^\infty P^t(x, A)\,a(dt).$$

A nonempty set $A \in \mathcal{B}_{\mathsf{X}}$ is called *μ_a-petite* if μ_a is a nontrivial measure on $(\mathsf{X}, \mathcal{B}_{\mathsf{X}})$ and a is a probability distribution on $(0, \infty)$ satisfying

$$K_a(x, \cdot) \geq \mu_a(\cdot)$$

for all $x \in A$. The distribution a is called the *sampling distribution* for the petite set A; see Section 4.1 of Meyn and Tweedie [41]. The following result is from [42], Theorem 4.1.

LEMMA 3.1. *Let B be a closed petite set, suppose that $\mathsf{P}_x(\tau_B < \infty) \equiv 1$ and that for some $\delta > 0$,*

$$\sup_{x\in B} \mathsf{E}_x[\tau_B(\delta)] < \infty, \tag{3.1}$$

where $\tau_B(\delta) = \inf\{t \geq \delta : X(t) \in B\}$. Then X is positive Harris recurrent.

LEMMA 3.2. *Under assumptions (1.4) and (1.5) on interarrival distributions, $B = \{|x| \leq \kappa\}$ is a closed petite set for any $\kappa > 0$.*

PROOF. This lemma can be proved in a way similar to the proof of Lemma 3.7 of [38]. □

REMARK. This is the only place where conditions (1.4) and (1.5) are used.

THEOREM 3.1. *If there exists $\delta > 0$ such that*

$$\lim_{|x|\to\infty} \frac{1}{|x|} \mathsf{E}|X^x(|x|\delta)| = 0, \tag{3.2}$$

then (3.1) holds for $B = \{x \in \mathsf{X} : |x| \leq \kappa\}$ with some $\kappa > 0$. In particular, X is positive Harris recurrent.

REMARK. Here we use X^x to denote X starting from x. Note that $\mathsf{E}|X^x(t)|$ can be alternatively written as $\mathsf{E}_x|X(t)|$.

PROOF. Let $0 < \varepsilon < 1$, for example, 1/2. From (3.2) there exists $\kappa \geq 1$ such that

$$\frac{1}{|x|}\mathsf{E}|X^x(|x|\delta)| \leq 1 - \varepsilon \tag{3.3}$$

for all x such that $|x| > \kappa$. Let $B = \{x \in \mathsf{X} : |x| \leq \kappa\}$. It follows from (1.3) that, for some constant $b > 0$,

$$\mathsf{E}|X^x(|x|\delta)| \leq (1 - \varepsilon)|x| + b1_B(x) \tag{3.4}$$

for all x. Let

$$n(x) = \begin{cases} |x|\delta, & \text{if } x \notin B, \\ \delta, & \text{if } x \in B. \end{cases} \tag{3.5}$$

Because $\kappa \geq 1$, $n(x) \geq \delta$ for all $x \in \mathsf{X}$. It follows from (3.4) that

$$\mathsf{E}|X^x(n(x))| \leq (1 - \varepsilon)|x| + b1_B(x) \leq |x| - \frac{\varepsilon}{\delta}n(x) + \tilde{b}1_B(x)$$

for some $\tilde{b} > 0$ and all $x \in \mathsf{X}$. Proceeding exactly the same as in the proof of Theorem 2.1(ii) of [43], we have for each $x \in \mathsf{X}$,

$$\mathsf{E}_x[\tau_B(\delta)] \leq \frac{\delta}{\varepsilon}(|x| + \tilde{b}) < \infty$$

and

$$\sup_{x\in B} \mathsf{E}_x[\tau_B(\delta)] \leq \frac{\delta}{\varepsilon}\left(\sup_{x\in B}|x| + \tilde{b}\right) = \frac{\delta}{\varepsilon}(\kappa + \tilde{b}) < \infty.$$

Thus, $\mathsf{P}_x(\tau_B < \infty) \equiv 1$ and it follows from Lemmas 3.1 and 3.2 that X is positive Harris recurrent. □

4. Fluid limit models. In order to make Theorem 3.1 practical, one must have a systematic method to check limit (3.2). The fluid approximation developed in this section is a powerful tool to obtain such a limit. In particular, we will prove in the main theorem (Theorem 4.2) of this paper that a queueing discipline in the queueing network is stable if the corresponding *fluid limit model* is stable. In Sections 5–7, we will give examples to demonstrate how one can check the stability of fluid limit models.

The following lemma maybe hidden in some textbook. It is needed to prove Lemma 4.2. For completeness, we present a proof.

LEMMA 4.1. *Let $\{f_n\}$ be a sequence of nondecreasing functions on $\mathbb{R}_+$ and let f be a continuous function on $\mathbb{R}_+$. Assume that $f_n(t) \to f(t)$ for all rational $t \geq 0$. Then $f_n \to f$ u.o.c.*

PROOF. First, because f_n is nondecreasing and f is continuous, one can easily check that $f_n(t) \to f(t)$ for every $t \in \mathbb{R}_+$. Next, suppose that f_n does not converge to f uniformly on compact sets. Then there exist $\varepsilon > 0$, $t > 0$ and $\{t_{n_l}\}$ such that $t_{n_l} \le t$ and

$$|f_{n_l}(t_{n_l}) - f(t_{n_l})| \ge \varepsilon \quad \text{for all } l. \tag{4.1}$$

Because $\{t_{n_l}\}$ is bounded, we may assume that $t_{n_l} \to t_0 \le t$. Thus for any $\delta > 0$, t_{n_l} eventually is less than $t_0 + \delta$. Hence for l large enough,

$$\begin{aligned} f_{n_l}(t_{n_l}) - f(t_{n_l}) &\le f_{n_l}(t_0+\delta) - f(t_{n_l}) \\ &= f_{n_l}(t_0+\delta) - f(t_0+\delta) + f(t_0+\delta) - f(t_0) + f(t_0) - f(t_{n_l}). \end{aligned}$$

Therefore,

$$\limsup_{l\to\infty}(f_{n_l}(t_{n_l}) - f(t_{n_l})) \le f(t_0+\delta) - f(t_0).$$

Because f is continuous and δ is arbitrary, we have

$$\limsup_{l\to\infty}(f_{n_l}(t_{n_l}) - f(t_{n_l})) \le 0.$$

When $t_0 > 0$, one can similarly prove that

$$\liminf_{l\to\infty}(f_{n_l}(t_{n_l}) - f(t_{n_l})) \ge 0.$$

When $t_0 = 0$,

$$\liminf_{l\to\infty}(f_{n_l}(t_{n_l}) - f(t_{n_l})) \ge \lim_{l\to\infty}(f_{n_l}(0) - f(t_{n_l})) = 0.$$

Thus we have

$$\lim_{l\to\infty}(f_{n_l}(t_{n_l}) - f(t_{n_l})) = 0,$$

which contradicts (4.1). Hence the lemma is proved. □

Let $x = (\mathbb{Q}(0), U(0), V(0))$ be the initial state of the network under a specified queueing discipline. In this section, we attach a superscript x to a symbol to explicitly denote the dependence on the initial state x. In particular, $Q_k^x(t)$ is the queue length for class k customers at time t. For $\ell \in \mathscr{E}$ and $k = 1, \dots, K$, define

$$E_\ell^x(t) = \max\{r\colon\ U_\ell(0) + \xi_\ell(1) + \cdots + \xi_\ell(r-1) \le t\}, \qquad t \ge 0,$$

$$S_k^x(t) = \max\{r\colon\ \tilde{V}_k(0) + \eta_k(1) + \cdots + \eta_k(r-1) \le t\}, \qquad t \ge 0,$$

where $\tilde{V}_k(0) = V_k(0)$ if there is a class k customer that has nonzero remaining service time at time 0 and is set to be a fresh class k service time $\eta_k(0)$ otherwise, where $\eta_k(0)$ is independent of $\{\eta_k(n),\ n \ge 1\}$ and has the distribution as that of $\eta_k(1)$. (We make the convention that the maximum of an empty set is zero.) It is clear that $E_\ell^x(t)$ is the total number of exogenous arrivals to class k by time t. For $k = 1, \dots, K$ and each n, define

$$\Phi^k(n) = \sum_{i=1}^{n} \phi^k(i).$$

Note that $\Phi^k(n)$ does not depend on x. Now we present a functional strong law of large numbers for the processes defined.

LEMMA 4.2. *Let $\{x_n\} \subset \mathsf{X}$ with $|x_n| \to \infty$ as $n \to \infty$. Assume that*

$$\lim_{n\to\infty} \frac{1}{|x_n|} U(0) = \overline{U} \quad \textit{and} \quad \lim_{n\to\infty} \frac{1}{|x_n|} V(0) = \overline{V}.$$

Then as $n \to \infty$, almost surely

$$\frac{1}{|x_n|} \Phi^k([|x_n| t]) \to P'_k t, \quad \textit{u.o.c.}, \tag{4.2}$$

$$\frac{1}{|x_n|} E_k^{x_n}(|x_n| t) \to \alpha_k (t - \overline{U}_k)^+, \quad \textit{u.o.c.}, \tag{4.3}$$

$$\frac{1}{|x_n|} S_k^{x_n}(|x_n| t) \to \mu_k (t - \overline{V}_k)^+, \quad \textit{u.o.c.}, \tag{4.4}$$

where $[t]$ is the integer part of t and $\mu_k = 1/m_k$.

REMARK. The discussion in Section 2.3 of Chen and Mandelbaum [9] also outlined a proof for the lemma.

PROOF. First (4.2) follows from the classical functional strong law of large numbers. To prove (4.3) it suffices to prove that

$$\frac{1}{|x_n|} E_k^0(|x_n| t) \to \alpha_k t, \quad \text{u.o.c.} \tag{4.5}$$

almost surely because

$$\frac{1}{|x_n|} E_k^{x_n}(|x_n| t) = \frac{1}{|x_n|} E_k^0(|x_n| (t - U_k(0)/|x_n|)^+).$$

To see (4.5), first by the strong law of large numbers for renewal process [see, e.g., [46], Theorem 3.3.2(1)], one has that almost surely

$$\lim_{t\to\infty} E_k^0(t)/t \to \alpha_k.$$

It then follows that almost surely

$$\lim_{n\to\infty} E_k^0(|x_n| t)/|x_n| = \lim_{n\to\infty} t\, E_k^0(|x_n| t)/(|x_n| t) = \alpha_k t \equiv f(t) \quad \text{for each } t \geq 0.$$

Because $f(t)$ is continuous and $E_k^0(|x_n|t)/|x_n|$ is nondecreasing for each n, it follows from Lemma 4.1 that (4.5) holds. Similarly, one can prove that (4.4) holds. □

For $x = (\mathbb{Q}(0), U(0), V(0))$, let

$$U_k^x(t) = \sum_{i=1}^{E_k^x(t)+1} \xi_k(i) + U_k(0) - t.$$

It is the *residual life* process associated with the delayed renewal process $\{E_k^x(t),\ t \geq 0\}$.

LEMMA 4.3. *Let $\{x_n\} \subset \mathsf{X}$ with $|x_n| \to \infty$ as $n \to \infty$. Assume that*

$$\lim_{n\to\infty} \frac{1}{|x_n|} U_k(0) = \overline{U}_k \quad \text{and} \quad \lim_{n\to\infty} \frac{1}{|x_n|} V_k(0) = \overline{V}_k.$$

(a) *As $n \to \infty$ almost surely,*

$$\frac{1}{|x_n|} U_k^{x_n}(|x_n|t) \to (\overline{U}_k - t)^+ \quad \text{u.o.c.} \quad \text{and} \quad \frac{1}{|x_n|} V_k^{x_n}(|x_n|t) \to (\overline{V}_k - t)^+ \quad \text{u.o.c.}$$

(b) *For each fixed $t \geq 0$,*

$$\left\{ \frac{1}{|x_n|} U_k^{x_n}(|x_n|t), |x_n| \geq 1 \right\} \quad \text{and} \quad \left\{ \frac{1}{|x_n|} V_k^{x_n}(|x_n|t), |x_n| \geq 1 \right\}$$

are uniformly integrable.

PROOF. For each fixed $t \geq 0$, by the strong law of large numbers, we have

$$\frac{1}{|x_n|} \sum_{i=1}^{E_k^{x_n}(|x_n|t)+1} \xi_k(i) \to (t - \overline{U}_k)^+ \tag{4.6}$$

almost surely. Hence, (4.6) holds almost surely for all rational $t \geq 0$. Because $t \to (t - \overline{U}_k)^+$ is continuous, by Lemma 4.1, almost surely (4.6) holds u.o.c. Thus (a) is proved. To prove (b), note that

$$\mathsf{E}_{x_n}\left[\frac{1}{|x_n|} U_k^{x_n}(|x_n|t)\right] = \mathsf{E}_{x_n}\left[\frac{1}{|x_n|} \sum_{i=1}^{E_k^{x_n}(|x_n|t)+1} \xi_k(i)\right] + \frac{U_k(0)}{|x_n|} - t.$$

By Wald's identity, the right term in the preceding equation is equal to

$$\frac{1}{|x_n|}\frac{1}{\alpha_k} \mathsf{E}_{x_n}[E_k^{x_n}(|x_n|t) + 1] + \frac{U_k(0)}{|x_n|} - t.$$

By the elementary renewal theorem ([46], Theorem 3.3.3), the preceding expression converges to $(t - \overline{U}_k)^+ + \overline{U}_k - t$. The proof of (b) then follows from Chung ([11], Theorem 4.5.4). □

Now we describe a set of equations that capture most of the network dynamics. These equations were first derived by Harrison [23] to obtain Brownian models with scheduling capability. They were later refined by Harrison and Nguyen [24] for FIFO discipline. They are now quite standard in Brownian models of queueing networks. For notational convenience we assume that at most one customer within a customer class can have nonzero partial service time. Thus $S_l^x(t)$ is the total number of service completions from class l if class l is given t units of service time. Let $T_l^x(t)$ be the cumulative amount of service time that server $s(l)$ has spent on class l customers by time t. Then $S_l^x(T_l^x(t))$ is the total number of service completions from class l by time t. A fraction of these customers, equal to $\Phi_k^l(S_l^x(T_l^x(t)))$ of them, become class k customers. Thus, we have the following representation for the queue length processes:

$$(4.7)\quad Q_k^x(t) = Q_k^x(0)+E_k^x(t)+\sum_{l=1}^{K}\Phi_k^l(S_l^x(T_l^x(t)))-S_k^x(T_k^x(t)), \qquad k = 1,\dots,K.$$

For $i = 1\dots,d$, let

$$I_i^x(t) = t - \sum_{k\in C_i} T_k^x(t).$$

Then $I_i^x(t)$ is the cumulative amount of time that server i has been idle by time t. We assume that all queueing disciplines considered in this paper are *work-conserving*. That is, server i is idle only when there is no customer at station i. Hence, we have

$$(4.8)\qquad \int_0^\infty \left(\sum_{k\in C_i} Q_k^x(t)\right) dI_i^x(t) = 0, \qquad i = 1,\dots,d.$$

Put $T^x(t) = (T_1^x(t),\dots,T_K^x(t))'$, $I^x(t) = (I_1^x(t),\dots,I_d^x(t))'$ and

$$S^x(T^x(t)) = (S_1^x(T_1^x(t)),\dots,S_1^x(T_1^x(t)))'.$$

For a fixed work-conserving queueing discipline, in vector form, we have

$$(4.9)\qquad Q^x(t) = Q^x(0) + E^x(t) + \sum_{l=1}^{K}\Phi^l(S_l^x(T_l^x(t))) - S^x(T^x(t)),$$

$$(4.10)\qquad Q^x(t) \geq 0,$$

$$(4.11)\qquad T^x(t) \text{ is nondecreasing and } T^x(0) = 0,$$

$$(4.12)\qquad I^x(t) = et - CT^x(t) \text{ is nondecreasing},$$

$$(4.13)\qquad \int_0^\infty (CQ^x(t))\, dI^x(t) = 0,$$

(4.14) some additional conditions on $(Q^x(\cdot), T^x(\cdot))$ that are specific to the discipline,

where e is the d-dimensional vector of ones, C is the constituency matrix defined in (1.6) and for a vector function $f(t) = (f_1(t), \dots, f_d(t))'$,

$$\int f(t)\, dI^x(t) = \left(\int f_1(t)\, dI_1^x(t), \dots, \int f_d(t)\, dI_d^x(t) \right)'.$$

Conditions (4.9)–(4.13) are satisfied for any work-conserving discipline. Examples will be given for condition (4.14) for priority disciplines in Section 7.

THEOREM 4.1. *Consider a work-conserving queueing discipline. For almost all sample paths* ω *and any sequence of initial states* $\{x_n\} \subset \mathsf{X}$ *with* $|x_n| \to \infty$, *there is a subsequence* $\{x_{n_j}\}$ *with* $|x_{n_j}| \to \infty$ *such that*

$$\frac{1}{|x_{n_j}|}(Q^{x_{n_j}}(0), U^{x_{n_j}}(0), V^{x_{n_j}}(0)) \to (\overline{Q}(0), \overline{U}, \overline{V}), \tag{4.15}$$

$$\frac{1}{|x_{n_j}|}(Q^{x_{n_j}}(|x_{n_j}| t), T^{x_{n_j}}(|x_{n_j}| t)) \to (\overline{Q}(t), \overline{T}(t)) \quad u.o.c. \tag{4.16}$$

Furthermore, $(\overline{Q}, \overline{T})$ *satisfies the following:*

$$\overline{Q}(t) = \overline{Q}(0) + (\alpha t - \overline{U})^+ - (I - P)' M^{-1} (\overline{T}(t) - \overline{V})^+, \tag{4.17}$$

$$\overline{Q}(t) \geq 0, \tag{4.18}$$

(4.19) $\overline{T}(t)$ *is nondecreasing and starts from zero,*

(4.20) $\bar{I}(t) = et - C\overline{T}(t)$ *is nondecreasing,*

$$\int_0^\infty (C\overline{Q}(t))\, d\bar{I}(t) = 0, \tag{4.21}$$

(4.22) *some additional conditions on* $(\overline{Q}(\cdot), \overline{T}(\cdot))$ *that are specific to the discipline.*

REMARK. This theorem, in a slightly different form, was proved by Chen and Mandelbaum [10].

PROOF. Notice first that

$$\frac{1}{|x_n|}|Q^{x_n}(0)| \leq 1, \qquad \frac{1}{|x_n|}|U^{x_n}(0)| \leq 1, \qquad \frac{1}{|x_n|}|V^{x_n}(0)| \leq 1$$

for all n. Therefore, there exists a subsequence $\{x_{n_j}\}$ such that (4.15) holds. For notational convenience, we assume that as $n \to \infty$,

$$\frac{1}{|x_n|}(Q^{x_n}(0), U^{x_n}(0), V^{x_n}(0)) \to (\overline{Q}(0), \overline{U}, \overline{V}).$$

Let ω be a fixed sample path such that (4.2)–(4.4) hold. For $x \in \mathsf{X}^0$ and $k = 1, \dots, K$, it is easy to see that

$$\frac{1}{|x|}T_k^x(|x|\,t) - \frac{1}{|x|}T_k^x(|x|\,s) \le t - s \quad \text{for } t \ge s \ge 0.$$

Therefore there exists a subsequence $\{x_{n_j}\}$ such that

$$\frac{1}{|x_{n_j}|}T_k^{x_{n_j}}(|x_{n_j}|\,t) \to \overline{T}_k(t) \quad \text{u.o.c. as } j \to \infty$$

for some process $\overline{T}(t) = (\overline{T}_1(t), \dots, \overline{T}_K(t))'$. Furthermore, conditions (4.18)–(4.20) follow easily from (4.10)–(4.12). By Lemma 4.2 and (4.9), we have

$$\frac{1}{|x_{n_j}|}Q^{x_{n_j}}(|x_{n_j}|\,t) \to \overline{Q}(t),$$

where $\overline{Q}(t)$ satisfies (4.17). To prove (4.21), note first that (4.13) is equivalent to

$$\int_0^t \left(\sum_{k \in C_i} \frac{1}{|x_{n_j}|} Q_k^{x_{n_j}}(|x_{n_j}|\,s) \right) \wedge 1\, d\left(\frac{1}{|x_{n_j}|} I_i^{x_{n_j}}(|x_{n_j}|s) \right) = 0$$

$$\text{for each } t \ge 0 \text{ and } i = 1, \dots, d,$$

where $a \wedge b$ is the minimum of a and b. Letting $j \to \infty$ and using Lemma 4.4, we have

$$\int_0^t \left(\sum_{k \in C_i} \overline{Q}_k(s) \right) \wedge 1\, d\bar{I}_i(s) = 0 \quad \text{for each } t \ge 0 \text{ and } i = 1, \dots, d,$$

which is equivalent to (4.21). Additional condition (4.22) is justified for each discipline from (4.14) by passing through the limiting procedure similar to the one for obtaining (4.21). □

LEMMA 4.4. *Let $\{(z_n, y_n)\}$ be a sequence in $D_{\mathbb{R}}[0,\infty) \times C_{\mathbb{R}}[0,\infty)$. Assume that y_n is nondecreasing and (z_n, y_n) converges to $(z, y) \in C_{\mathbb{R}}[0,\infty) \times C_{\mathbb{R}}[0,\infty)$ u.o.c. Then for any bounded continuous function f,*

$$\int_0^t f(z_n(s))\, dy_n(s) \to \int_0^t f(z(s))\, dy(s) \quad \textit{u.o.c.}$$

PROOF. See Lemma 2.4 of [14].

DEFINITION 4.1. Let a queueing discipline be fixed. Any limit $(\overline{Q}(\cdot), \overline{T}(\cdot))$ in (4.16) is a *fluid limit* of the discipline. Any solution $(\overline{Q}(\cdot), \overline{T}(\cdot))$ to (4.17)–(4.22) is called a *fluid solution* of the discipline. We say that fluid limit model (resp. fluid model) of the queueing discipline is *stable* if there exists a constant $\delta > 0$ that depends on α, μ and P only, such that for any fluid limit (resp. fluid solution) with $|\overline{Q}(0)| + |\overline{U}| + |\overline{V}| = 1$, $\overline{Q}(\cdot + \delta) \equiv 0$.

Now we can summarize our main result.

THEOREM 4.2. *Let a queueing discipline be fixed. Assume that* (1.2)–(1.5) *hold. If the fluid limit model of the queueing discipline is stable, then the Markov chain X describing the dynamics of the network under the discipline is positive Harris recurrent.*

PROOF. First notice that

$$\mathsf{E}|X^x(t)| = \sum_{k=1}^K \mathsf{E}Q_k^x(t) + \sum_{k=1}^K \mathsf{E}[U_k^x(t) + V_k^x(t)]. \tag{4.23}$$

Assume that the fluid model is stable and $\delta \geq 1$ is a constant as in Definition 4.1. Let $\{x_n\} \subset \mathbf{X}$ be any sequence of initial states with $|x_n| \to \infty$ as $n \to \infty$. Because $\delta \geq 1$ and the fact that $|\overline{U}| + |\overline{V}| \leq 1$, by Theorem 4.1 and Lemma 4.3 (a), almost surely there is a subsequence $\{x_{n_j}\}$ such that

$$\lim_{j\to\infty} \frac{1}{|x_{n_j}|} Q_k^{x_{n_j}}(|x_{n_j}|\,\delta) = \overline{Q}_i(\delta) = 0,$$

$$\lim_{j\to\infty} \frac{1}{|x_{n_j}|} U_k^{x_{n_j}}(|x_{n_j}|\,\delta) = (\delta - \overline{U}_k)^+ + \overline{U}_k - \delta = 0,$$

$$\lim_{j\to\infty} \frac{1}{|x_{n_j}|} V_k^{x_{n_j}}(|x_{n_j}|\,\delta) = (\delta - \overline{V}_k)^+ + \overline{V}_k - \delta = 0.$$

It follows from Lemma 4.5 below and Lemma 4.3(b) that

$$\lim_{j\to\infty} \tfrac{1}{|x_{n_j}|} \mathsf{E}[Q_k^{x_{n_j}}(|x_{n_j}|\,\delta)] = 0, \tag{4.24}$$

$$\begin{aligned} &\lim_{j\to\infty} \tfrac{1}{|x_{n_j}|} \mathsf{E}[U_k^{x_{n_j}}(|x_{n_j}|\,\delta)] = 0 \quad \text{and} \\ &\lim_{j\to\infty} \tfrac{1}{|x_{n_j}|} \mathsf{E}[V_k^{x_{n_j}}(|x_{n_j}|\,\delta)] = 0. \end{aligned} \tag{4.25}$$

By (4.23)–(4.25), we have

$$\lim_{j\to\infty} \frac{1}{|x_{n_j}|} \mathsf{E}|X^{x_{n_j}}(|x_{n_j}|\,\delta)| = 0.$$

Because $\{x_n\}$ is an arbitrary sequence, we have that condition (3.2) holds and hence X is positive Harris recurrent. □

REMARK 1. Even if the fluid model is stable, the fluid solution in (4.17)–(4.22) may not be unique. This is not surprising because the equations governing the fluid model did not specify how the fluid is configured initially.

REMARK 2. Under the standard assumptions (1.2)–(1.5) the queueing network is stable if the corresponding fluid limit model is stable. In the sequel, we always assume that these standard assumptions hold.

LEMMA 4.5. $\{(1/|x|)Q^x(|x|\,\delta)\colon x \in \mathrm{X}$ *with* $|x| \geq 1\}$ *is uniformly integrable.*

PROOF. Because

$$\frac{1}{|x|}|Q^x(|x|\,\delta)| \leq \sum_{k\in\mathscr{E}} \frac{1}{|x|} E_k^x(|x|\,\delta) + 1,$$

it suffices to prove for each $k \in \mathscr{E}$,

$$\{E_k^0(t)/t, t \geq 1\} \tag{4.26}$$

is uniformly integrable. Because

$$E_k^0(n)/n \to \alpha_k \quad \text{almost surely}$$

and $\mathsf{E}[E_k^0(n)]/n \to \alpha_k$ by the elementary renewal theorem ([46], Theorem 3.3.3), it follows from [11], Theorem 4.5.4, that $\{E^0(n)/n,\ n \geq 1\}$ is uniformly integrable. Note that for each $t \geq 1$, there is a unique integer n such that $n-1 \leq t \leq n$ and

$$\frac{E_k^0(t)}{t} \leq \frac{n}{n-1}\,\frac{E_k^0(n)}{n}.$$

Hence $\{E_k^0(t)/t,\ t \geq 1\}$ is uniformly integrable and the lemma is proved. □

5. Generalized Jackson network. Consider a special case where $|C_i| = 1$ for $i = 1, \dots, d$. This is called a single class network because there is only one class of customers served at each station. In other words, all customers at a station are homogeneous in terms of service requirement and routing. This network is the same one as considered by Jackson [26] except that we allow general interarrival and service time distributions. Hence, this single class network is often called a generalized Jackson network.

THEOREM 5.1. *Assume that* (1.2)–(1.5) *hold in a generalized Jackson network. Then each of the work-conserving disciplines discussed in Section* 2.1 *is stable if* (1.9) *holds.*

REMARK. The stability question for this network was resolved by Borovkov [4], Sigman [50], Foss [20], Meyn and Down [38], Chang, Thomas and Kiang [7] and Baccelli and Foss [1] under various assumptions on interarrival and service time distributions.

Before proving the theorem, let us first introduce the dynamic complementarity problem (DCP). Let

$$R = (I - P')M^{-1}.$$

DEFINITION 5.1. Let x: $[0,\infty) \to \mathbb{R}^d$ be a continuous path with $x(0) \geq 0$. The *(continuous) dynamic complementarity problem* for $x(\cdot)$ and R is DCP$(x(\cdot), R)$: given $x(\cdot)$, find a pair of continuous paths $(z(\cdot), y(\cdot))$ such that

$$z(t) = x(t) + Ry(t) \quad \text{for } t \geq 0,$$

$$z(t) \geq 0 \quad \text{for } t \geq 0,$$

$$y(\cdot) \text{ is nondecreasing, } y(0) = 0$$

and

$$y_i(\cdot) \text{ increases only at times } t \text{ such that } z_i(t) = 0,\ i = 1, \dots, d.$$

The pair $(z(\cdot), y(\cdot))$ is a solution to the DCP$(x(\cdot), R)$.

The term DCP is from Mandelbaum [37]. It is sometimes called the *deterministic Skorohod problem.* When the solution $(z(\cdot), y(\cdot))$ is unique for each $x(\cdot)$, the mapping $x(\cdot) \to (z(\cdot), y(\cdot))$ is also called the *reflection mapping* with reflection matrix R; see [22]. Harrison and Reiman [25] proved that given an $x(\cdot)$ with $x(0) \geq 0$, there is a unique solution $(z(\cdot), y(\cdot))$ to the DCP$(x(\cdot), R)$. Mandelbaum [37], Bernard and El Kharroubi [3] and Kozyakin and Mandelbaum [31] studied the DCP for much more general reflection matrix R. The reflection mapping has played an important role in the heavy traffic analysis of queueing networks; see [45], [27] and [9].

A path $x(\cdot)$ is said to be regular at t if it is differentiable at t. We use $\dot{x}(t)$ to denote the derivative of $x(\cdot)$ at a regular point t. The following lemma resembles a monotonicity result in [19] and [49].

LEMMA 5.1. *Let $(z(\cdot), y(\cdot))$ be the solution to the* DCP$(x(\cdot), R)$ *and let $s \geq 0$ be fixed. Assume that* (1.9) *holds and*

$$z(s) + x(t+s) - x(s) \geq \theta t \quad \textit{for all } t \geq 0, \tag{5.1}$$

where

$$\theta = \alpha + P'\mu - \mu. \tag{5.2}$$

Then

$$y(t+s) - y(s) \leq (e - \rho)t \quad \textit{for all } t \geq 0 \tag{5.3}$$

and hence $\dot{y}(s) \leq (e - \rho)$ if $y(\cdot)$ is regular at s.

PROOF. Take $x^0(t) = \theta t$ for $t \geq 0$ and

$$y^0(t) = -R^{-1}x^0(t) = (e - \rho)t.$$

Then $y^0(0) = 0$ and $y^0(\cdot)$ is nondecreasing by (1.9). Let $(\tilde{z}^0(\cdot), \tilde{y}^0(\cdot))$ be the unique solution to the DCP$(x^0(\cdot), R)$. Because $\tilde{z}^0(t) \geq 0$ and the elements of R^{-1} are nonnegative, we have $\tilde{y}^0(t) \geq y^0(t)$ for $t \geq 0$. If follows from Propositions 1 and 2 of [45] that $\tilde{z}^0(t) \equiv 0$ and $\tilde{y}^0(t) \equiv y^0(t)$. Note that

$$z(t+s) = z(s) + x(t+s) - x(s) + R(y(t+s) - y(s)) \equiv \tilde{x}(t) + R\tilde{y}(t),$$

where $\tilde{x}(t) = z(s) + x(t+s) - x(s)$ and $\tilde{y}(t) = y(t+s) - y(s)$. Assume that (5.1) holds. Then, $\tilde{x}(t) \geq x^0(t)$ for $t \geq 0$. It follows again from Propositions 1 and 2 of [45] that

$$\tilde{y}(t) = y(t+s) - y(s) \leq y^0(t) = (e - \rho)t \quad \text{for } t \geq 0. \qquad \square$$

LEMMA 5.2. *Let f: $[0, \infty) \to [0, \infty)$ be a nonnegative function that is absolutely continuous and let $\kappa > 0$ be a constant. Suppose that for almost surely [with respect to Lebesgue measure on $[0, \infty)$] all regular points t, $\dot{f}(t) \leq -\kappa$ whenever $f(t) > 0$. Then f is nonincreasing and $f(t) \equiv 0$ for $t \geq f(0)/\kappa$.*

The elementary proof is omitted.

While this paper was under review, Chen [8] proved the following result.

LEMMA 5.3. *Suppose that there is $\delta > 0$ such that $\overline{Q}(\cdot + \delta) \equiv 0$ for any fluid solution $(\overline{Q}(\cdot), \overline{T}(\cdot))$ with $|\overline{Q}(0)| = 1$, $\overline{U} = 0$ and $\overline{V} = 0$. Then the fluid model is stable.*

PROOF OF THEOREM 5.1. Without loss of generality, we assume that $C_i = \{i\}$. By Theorem 4.2 and Lemma 5.3 it suffices to show that any solution $\overline{Q}(\cdot)$ to

$$\begin{aligned} \overline{Q}(t) &= \overline{Q}(0) + \alpha t - (I - P)' M^{-1} \overline{T}(t) \\ &= \overline{Q}(0) + \alpha t - (I - P)' \mu t + (I - P)' M^{-1} \bar{I}(t), \end{aligned} \tag{5.4}$$

$$\overline{Q}(t) \geq 0, \tag{5.5}$$

$$\overline{T}(0) = 0, \overline{T}(t) \text{ is nondecreasing}, \tag{5.6}$$

$$\bar{I}(t) = et - \overline{T}(t) \text{ is nondecreasing}, \tag{5.7}$$

$$\int_0^\infty \overline{Q}(t)\, d\bar{I}(t) = 0 \tag{5.8}$$

is identically zero on $[\delta, \infty)$ for some $\delta > 0$ that depends on α, μ and P only. At this point we could quote Chen and Mandelbaum [9] to finish the rest of the proof. Instead we provide a new approach to the proof. This approach was communicated to me by Meyn and Down. It follows that $(\overline{Q}(\cdot), \overline{T}(\cdot))$ is the solution to the DCP for $\overline{Q}(0) + \theta t$ and $R = (I - P')M^{-1}$. One can check that condition (5.1) is satisfied for each $s \geq 0$. Therefore by Lemma 5.1, $\dot{\bar{I}}(s) \leq (1 - \rho)$ at every regular point s of $\bar{I}(\cdot)$. Let

$$f(t) \equiv e'R^{-1}\overline{Q}(t) = e'M(I - P')^{-1}\overline{Q}(t).$$

It follows from (5.4) that

$$f(t) = f(0) + \sum_{i=1}^{d}((\rho_i - 1)t + \bar{I}_i(t)).$$

Assume that t is a regular point for $\overline{Q}(\cdot)$ and $\bar{I}(\cdot)$. If $f(t) > 0$, then $\overline{Q}_i(t) > 0$ for some i. Hence, by (5.8), $\dot{\bar{I}}_i(t) = 0$. Hence

$$\dot{f}(t) \leq -(1 - \rho_i).$$

It is clear that f is absolutely continuous and it follows from Lemma 4.2 that $f(t) \equiv 0$ for $t \geq f(0)/\kappa$, where

$$f(0) = e'M(I - P')^{-1}\overline{Q}(0) \leq m'(I - P)^{-1}e \quad \text{and} \quad \kappa = 1 - \max_{1 \leq j \leq d} \rho_j.$$

Put $\delta = m'(I - P')^{-1}e/(1 - \max_{1 \leq j \leq d} \rho_j)$. Then $\overline{Q}(t) \equiv 0$ for $t \geq \delta$. □

6. Multiclass station and feedforward network.

THEOREM 6.1. *Let $d = 1$. Any work-conserving queueing discipline discussed in Section* 2.1 *is stable under* (1.9).

PROOF. Assume that (1.9) holds. By Theorem 4.2 and Lemma 5.3, it is enough to show that the fluid model is stable when $\overline{U} = 0$ and $\overline{V} = 0$. Let

$$f(t) = e'M(I - P')^{-1}\overline{Q}(t) = f(0) + (\rho - 1)t + \bar{I}(t).$$

Let $t > 0$ be a regular point of $\overline{Q}(\cdot)$ and $\bar{I}(\cdot)$. If $f(t) > 0$, then $\sum_{k=1}^{K}\overline{Q}_k(t) > 0$. Therefore, it follows from (4.21) that $\dot{\bar{I}}(t) = 0$ and thus $\dot{f}(t) = -(1 - \rho) < 0$. Therefore, $\overline{Q}(t) \equiv 0$ for $t \geq f(0)/(1 - \rho) \leq m'(I - P')^{-1}e$. □

A network is *feedforward* if stations can be numbered in such a way that $P_{kl} = 0$ for any $k \in C_i$ and $l \in C_j$ with $i > j$.

COROLLARY 6.2. *In a feedforward network, any work-conserving queueing discipline is stable under* (1.9).

7. First-buffer–first-served in a reentrant line. Consider a reentrant line where each customer follows one deterministic route. Customers in the kth stage of service are called class k customers. The entries of the routing matrix take the form $P_{k,k+1} = 1$ for $k = 1, \dots, K-1$ and zero for all others. The deterministic version of this network was studied by Kumar [32]. Suppose that for any two customer classes k and l visiting the same station, class k customer priority is higher than class l's whenever $k < l$. Then the queueing discipline is called first-buffer–first-served (FBFS) discipline.

THEOREM 7.1. *For a reentrant line, the first-buffer–first-served preemptive resume discipline is stable under* (1.9).

REMARK. FBFS is one of the priority disciplines considered in [13], where the stability for a class of priority queueing disciplines is studied. We present this special case here mainly to illustrate the effectiveness of using Theorem 4.2 to establish stability for some multiclass networks with feedback. A deterministic version of this theorem was proved by Kumar [32].

PROOF. Assume that $\alpha_1 = 1$. Let H_k denote the set of indices for all classes served at station $s(k)$ that have priority greater than or equal to that of class k. That is, $H_k = \{l\colon\ 1 \le l \le k,\ s(l) = s(k)\}$. Let

$$T_k^{x,+}(t) = \sum_{\ell \in H_k} T_\ell^x(t)$$

$$I_k^{x,+}(t) = t - T_k^{x,+}(t),$$

$$Q_k^{x,+}(t) = \sum_{\ell \in H_k} Q_\ell^x(t).$$

Then $T_k^{x,+}(t)$ is the cumulative amount of service in $[0,t]$ dedicated to customers whose classes are included in H_k, and $I_k^{x,+}(t)$ is the total unused capacity that is available to serve customers whose class does not belong to H_k. Note that $I_i^x(t)$ is a station level quantity representing the total unused capacity in $[0,t]$ by server i, whereas $I_k^{x,+}(t)$ is a class level quantity. The priority service discipline requires that for every k, all the service capacity of station $s(k)$ is dedicated to classes in H_k, as long as the workload present in these buffers is positive. Because we assume the preemptive resume queueing discipline, we have that $I_k^{x,+}(\cdot)$ increases only when $Q_k^{x,+}(t) = 0$. Thus we may express additional condition (4.14) by the integral equation

$$\int_0^\infty Q_k^{x,+}(t)\, dI_k^{x,+}(t) = 0, \qquad 1 \le k \le K. \tag{7.1}$$

Condition (7.1) implies that for each $x \in \mathsf{X}^0$,

$$\int_0^\infty \frac{1}{|x|} Q_k^{x,+}(|x|\, t)\, d\left(\frac{1}{|x|} I_k^{x,+}(|x|\, t)\right) = 0, \qquad 1 \le k \le K.$$

Taking the limit as in Theorem 4.1 and using Lemma 4.4, we have

$$\int_0^\infty \overline{Q}_k^+(t)\,d\bar{I}_k^+(t) = 0, \qquad 1 \le k \le K, \tag{7.2}$$

where

$$\overline{T}_k^+(t) = \sum_{\ell \in H_k} \overline{T}_\ell(t),$$

$$\bar{I}_k^+(t) = t - \overline{T}_k^+(t),$$

$$\overline{Q}_k^+(t) = \sum_{\ell \in H_k} \overline{Q}_\ell(t).$$

Therefore, for the priority discipline, the additional condition (4.22) takes the form (7.2).

Now we show that the fluid model determined by (4.17)–(4.20) and (7.2) is stable under (1.9). In light of Lemma 5.3, we can assume that $\overline{U} = \overline{V} = 0$. We use induction to show that for each $k = 1, \ldots, K$, there exists t_k such that $\overline{Q}_\ell(t + t_k) \equiv 0$ for $\ell = 1, \ldots, k$. For $k = 1$, because

$$\overline{Q}_1(t) = \overline{Q}_1(0) + (1 - \mu_1)\,t + \mu_1 \bar{I}_1^+(t),$$

$$\overline{Q}_1(t) \ge 0,$$

$$\bar{I}_1(t) \text{ is nondecreasing,}$$

$$\int_0^\infty \overline{Q}_1(t)\,d\bar{I}_1^+(t) = 0,$$

we have $\overline{Q}_1(t) \equiv 0$ for $t \ge \overline{Q}_1(0)/(\mu_1 - 1)$. Because $|\overline{Q}(0)| = 1$, we can take $t_1 = 1/(\mu_1 - 1)$. Assume that at time t_{k-1} all the buffers $1, \ldots, k-1$ are empty and that they shall stay empty for $t > t_{k-1}$. Let the content of buffer k be $\overline{Q}_k(t_{k-1})$. This is bounded above by $1 + t_{k-1}$. Then, because for $t > t_{k-1}$ buffers $1, \ldots, k-1$ remain empty, it follows that if $\overline{Q}_k(t) > 0$ for $t \ge t_{k-1}$, then it is the first nonempty buffer, and so by Proposition 4.2 of Dai and Weiss [13] the arrival rate to buffer k is $a_k(t) = 1$ and the departure rate from buffer k is

$$d_k(t) = \frac{1 - \sum_{\ell \in H_k \setminus \{k\}} m_\ell}{m_k} > 1.$$

Therefore, $\dot{\overline{Q}}_k(t) = -\mu_k(1 - \sum_{\ell \in H_k} m_\ell) < 0$, and hence buffer k will be empty at time t_k, where $t_k - t_{k-1} = Q_k(t_{k-1}) m_k / (1 - \sum_{\ell \in H_k} m_\ell)$ and will stay empty at all times after t_k. Therefore, by the induction argument the fluid model will reach $\overline{Q}(t) = 0$ and remain zero thereafter no later than at

$$\delta = \sum_{k=1}^{K} \left(m_k \frac{\prod_{l=1}^{k-1} (1 - \sum_{j \in H_\ell \setminus \{\ell\}} m_j)}{\prod_{l=1}^{k} (1 - \sum_{j \in H_\ell} m_j)} \right).$$

(See [13] for discussions on fluid models under other priority disciplines.) □

Acknowledgments. My interest in this research was initiated by a conversation with Ruth Williams during the 21st Conference on Stochastic Processes and their Applications in Toronto in the summer of 1992. I am grateful to Ward Whitt for sending me a recent paper by Rybko and Stolyar, which provided further motivation for this research. Tom Kurtz helped to weaken condition (1.3) to the current form. Sean Meyn made numerous comments that have greatly improved the exposition of this paper. In particular, he and a referee corrected mistakes in an earlier formulation of Lemma 3.2. I also benefited from conversations with Hong Chen, Peter Glynn, Avi Mandelbaum and Gideon Weiss, and from comments by two anonymous referees.

REFERENCES

[1] BACCELLI, F. and FOSS, S. (1994). Stability of Jackson-type queueing networks, I. *Queueing Systems Theory Appl.* **17** 5–72.

[2] BASKETT, F., CHANDY, K. M., MUNTZ, R. R. and PALACIOS, F. G. (1975). Open, closed and mixed networks of queues with different classes of customers. *J. Assoc. Comput. Mach.* **22** 248–260.

[3] BERNARD, A. and EL KHARROUBI, A. (1991). Régulations déterministes et stochastiques dans le premier "orthant" de $\mathbb{R}^n$. *Stochastics Stochastics Rep.* **34** 149–167.

[4] BOROVKOV, A. A. (1986). Limit theorems for queueing networks. *Theory Probab. Appl.* **31** 413–427.

[5] BOTVITCH, D. D. and ZAMYATIN A. A. (1992). Ergodicity of conservative communication networks. Rapport de recherche 1772, INRIA.

[6] BRAMSON, M. (1994). Instability of FIFO queueing networks. *Ann. Appl. Probab.* **4** 414–431.

[7] CHANG, C. S., THOMAS, J. A. and KIANG, S. H. (1994). On the stability of open networks: a unified approach by stochastic dominance. *Queueing Systems Theory Appl.* **15** 239–260.

[8] CHEN, H. (1994). Fluid approximations and stability of multiclass queueing networks I: Work-conserving disciplines. Unpublished manuscript.

[9] CHEN, H. and MANDELBAUM, A. (1991). Discrete flow networks: Bottlenecks analysis and fluid approximations. *Math. Oper. Res.* **16** 408–446.

[10] CHEN, H. and MANDELBAUM, A. (1994). Open heterogeneous fluid networks, with applications to multi-type queues. Preprint.

[11] CHUNG, K. L. (1974). *A Course in Probability*. Wiley, New York.

[12] DAI, J. G. and MEYN, S. P. (1994). Stability and convergence of moments for multiclass queueing networks via fluid limit models. Unpublished manuscript.

[13] DAI, J. G. and WEISS, G. (1994). Stability and instability of fluid models for certain re-entrant lines. Unpublished manuscript.

[14] DAI, J. G. and WILLIAMS, R. J. (1995). Existence and uniqueness of semimartingale reflecting Brownian motions in convex polyhedrons. *Theory Probab. Appl.* **40**(1).

[15] DAVIS, M. H. A. (1984). Piecewise deterministic Markov processes: a general class of non-diffusion stochastic models. *J. Roy. Statist. Soc. Ser. B* **46** 353–388.

[16] DOWN, D. and MEYN, S. P. (1994). Piecewise linear test functions for stability of queueing networks. In *Proceedings of the 33rd Conference on Decision and Control*.

[17] DUPUIS, P. and WILLIAMS, R. J. (1994). Lyapunov functions for semimartingale reflecting Brownian motions. *Ann. Probab.* **22** 680–702.

[18] ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes: Characterization and Convergence*. Wiley, New York.

[19] FOSS, S. G. (1989). Some properties of open queueing networks. *Problems Inform. Transmission* **25** 241–246.

[20] FOSS, S. G. (1991). Ergodicity of queueing networks. *Siberian Math. J.* **32** 183–202.

[21] GETOOR, R. K. (1979). Transience and recurrence of Markov processes. *Séminaire de Probabilités XIV* **284** 397–409. Springer, New York.

[22] HARRISON, J. M. (1985). *Brownian Motion and Stochastic Flow Systems*. Wiley, New York.

[23] HARRISON, J. M. (1988). Brownian models of queueing networks with heterogeneous customer populations. In *Proceedings of the IMA Workshop on Stochastic Differential Systems*. Springer, Berlin.

[24] HARRISON, J. M. and NGUYEN, V. (1993). Brownian models of multiclass queueing networks: Current status and open problems. *Queueing Systems Theory Appl.* **13** 5–40.

[25] HARRISON, J. M. and REIMAN, M. I. (1981). Reflected Brownian motion on an orthant. *Ann. Probab.* **9** 302–308.

[26] JACKSON, J. R. (1957). Networks of waiting lines. *Oper. Res.* **5** 518–521.

[27] JOHNSON, D. P. (1983). Diffusion approximations for optimal filtering of jump processes and for queueing networks. Ph.D. thesis, Univ. Wisconsin.

[28] KASPI, H. and MANDELBAUM, A. (1992). Regenerative closed queueing networks. *Stochastics Stochastics Rep.* **39** 239–258.

[29] KASPI, H. and MANDELBAUM, A. (1994). On Harris recurrence in continuous time. *Math. Oper. Res.* **19** 211–222.

[30] KELLY, F. P. (1975). Networks of queues with customers of different types. *J. Appl. Probab.* **12** 542–554.

[31] KOZYAKIN, V. S. and MANDELBAUM, A. (1994). Absolute stability and dynamic complementarity. Unpublished manuscript.

[32] KUMAR, P. R. (1993). Re-entrant lines. *Queueing Systems Theory Appl.* **13** 87–110.

[33] KUMAR, P. R. and MEYN, S. P. (1995). Stability of queueing networks and scheduling policies. *IEEE Trans. Automat. Control.*

[34] KUMAR, P. R. and SEIDMAN, T. I. (1990). Dynamic instabilities and stabilization methods in distributed read-time scheduling of manufacturing systems. *IEEE Trans. Automat. Control* **35** 289–298.

[35] LU, S. H. and KUMAR, P. R. (1991). Distributed scheduling based on due dates and buffer priorities. *IEEE Trans. Automat. Control* **36** 1406–1416.

[36] MALYSHEV, V. A. (1993). Networks and dynamical systems. *Adv. in Appl. Probab.* **25** 140–175.

[37] MANDELBAUM, A. (1994). The dynamic complementarity problem. *Math. Oper. Res.* To appear.

[38] MEYN, S. P. and DOWN, D. (1994). Stability of generalized Jackson networks. *Ann. Appl. Probab.* **4** 124–148.

[39] MEYN, S. P. and TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability.* Springer, London.

[40] MEYN, S. P. and TWEEDIE, R. L. (1993). Generalized resolvents and Harris recurrence of Markov processes. *Contemp. Math.* **149** 227–250.

[41] MEYN, S. P. and TWEEDIE, R. L. (1993). Stability of Markovian processes II: Continuous time processes and sample chains. *Adv. in Appl. Probab.* **25** 487–517.

[42] MEYN, S. P. and TWEEDIE, R. L. (1993). Stability of Markovian processes III: Foster–Lyapunov criteria for continuous time processes. *Adv. in Appl. Probab.* **25** 518–548.

[43] MEYN, S. P. and TWEEDIE, R. L. (1994). State-dependent criteria for convergence of Markov chains. *Ann. Appl. Probab.* **4** 149–168.

[44] PERKINS, J. R. and KUMAR, P. R. (1989). Stable distributed real-time scheduling of flexible manufacturing/assembly/disassembly systems. *IEEE Trans. Automat. Control* **34** 139–148.

[45] REIMAN, M. I. (1984). Open queueing networks in heavy traffic. *Math. Oper. Res.* **9** 441–458.

[46] RESNICK, S. I. (1992). *Adventures in Stochastic Processes*. Birkhäuser, Boston.

[47] RYBKO, A. N. and STOLYAR, A. L. (1992). Ergodicity of stochastic processes describing the operations of open queueing networks. *Problems Inform. Transmission* **28** 199–220.

[48] SEIDMAN, T. I. (1994). 'First come, first serve' can be unstable! *IEEE Trans. Automatic Control.* **39** 2166–2171.

[49] SHANTHIKUMAR, J. G. and YAO, D. D. (1989). Stochastic monotonicity in general queueing networks. *J. Appl. Probab.* **26** 413–417.

[50] SIGMAN, K. (1990). The stability of open queueing networks. *Stochastic. Process. Appl.* **35** 11–25.

[51] WHITT, W. (1993). Large fluctuations in a deterministic multiclass network of queues. *Management Sci.* **39** 1020–1028.

SCHOOL OF INDUSTRIAL AND
SYSTEMS ENGINEERING AND
SCHOOL OF MATHEMATICS
GEORGIA INSTITUTE OF TECHNOLOGY
ATLANTA, GEORGIA 30332-0205