**On the Stability of Multiclass Queueing Networks:
A Relaxed Sufficient Condition via Limiting Fluid Processes**

Alexander L. Stolyar
Motorola, Inc., 1501 West Shure Drive, 1441, Arlington Heights, IL 60004;
stolyara@cig.mot.com

Abstract

We consider a multiclass queueing network, whose underlying stochastic process is a countable, continuous time Markov chain. Stability of the network is understood as ergodicity of this Markov chain. The message class determines a message route through the network and the mean message service time in each node on its route. Each node may have its own queueing discipline within a wide class, including FCFS, LCFS, Priority and Processor Sharing.

We will show that the sequence of scaled (in space and time) underlying stochastic processes converges to a fluid process with sample paths defined as fixed points of a special operator. This convergence together with continuity and similarity properties of the family of sample paths of the fluid process allows us to prove the following result.

If each sample path of the fluid process with non-zero initial state is such that the "amount of fluid" in the network falls below its initial value at least once, then the network is stable.

Key words and phrases: Multiclass queueing network, stability, fluid limit.

1. Introduction.

We consider an open queueing network with several customer classes. The class of a customer determines his route through the network and his service time in each node on the route. Each node of the network is a single-server system with its own queueing iscipline. By stability of a multiclass queueing network we will mean ergodicity of the underlying stochastic process.

It has been shown recently that the condition that *the load ρ_j of each network node j is less than* 1

$$\rho_j < 1 \tag{1.1}$$

is not sufficient for network stability. Examples of unstable deterministic networks were found by Kumar and Seidman [9], and Lu and Kumar [10]. A similar example of an unstable stochastic network was found by Rybko and Stolyar [14]. Bramson [1,2] and Seidman [15] derived examples of unstable stochastic networks with FCFS queueing disciplines in the nodes. The instability problem has also been addressed in Whitt [17].

The following approach has been proposed in [14]. A sufficient ergodicity condition for continuous time countable Markov chains is derived. Similar and more general ergodicity criteria for discrete time Markov chains were obtained by Malyshev and Menshikov [11]. This condition allows us to reduce the stability problem for the original stochastic network to checking property (1.6) (see below) for the deterministic (fluid) process, obtained from the original stochastic process by space and time scaling.

Not quite formally yet, let $q(t)$ denote the non-negative vector specifying the number of customers of each type in each node of the network at time t. Let $||q(t)||$ denote the sum of the components of the vector $q(t)$.

Consider the sequence of scaled processes

$$q^n(t) \equiv \frac{1}{n} q(nt), \; n = 1, 2, \ldots \tag{1.2}$$

$$q^n(0) \to q(0), \; ||q(0)|| = 1, \tag{1.3}$$

where $q(0)$ is a fixed vector. It has been shown in [14] that if the underlying stochastic process describing the network behaviour is a continuous time countable Markov chain, then the following condition is sufficient for network stability.

There exists a constant $T > 0$, such that for any sequence of scaled processes $q^n(t)$, $n \to \infty$, satisfying conditions (1.2) and (1.3),

$$E||q^n(t)|| \to 0, \; \forall t \geq T. \tag{1.4}$$

The sufficient condition (1.4) establishes a close connection between network stability and properties of the corresponding fluid process. Indeed, it is well known that a linear space-time scaling, i.e. $\frac{1}{n}q(nt)$-type scaling, "usually" leads to a deterministic (*fluid*) process in the limit. In other words, it is natural to expect that in some sense the convergence

$$q^n(t) \to q(t) \tag{1.5}$$

takes place, where $q(t)$ is some fluid process. This leads to the following property of the fluid process, which corresponds to property (1.4) of the sequence of scaled processes.

There exists constant $T > 0$ such that for any fluid processes $q(t)$ with $||q(0)|| =$ 1

$$||q(t)|| = 0, \; \forall t \geq T. \tag{1.6}$$

The paper [14] formally defines the fluid process for networks with FCFS and priority disciplines in the nodes. Property (1.6) is proven for a two node network with FCFS queueing discipline in each node. Subsequently using properties of the fluid process $q(t)$, condition (1.4) and therefore stability of the network are verified.

Remark 1.1 For many networks it is easy to see that for some constant $c < 0$

$$\frac{d}{dt}||q(t)|| < c < 0, \tag{1.7}$$

if $||q(t)|| > 0$. This obviously implies property (1.6). The two node FCFS network considered in [14], is an example of a network not satisfying condition (1.7) but nevertheless satisfying condition (1.6).

In this paper we consider multiclass networks, whose underlying stochastic process is a countable continuous time Markov chain. The queueing disciplines in the nodes may belong to a wide class including in particular FCFS, LCFS, Priority and Processor sharing. For any set of queueing disciplines in the nodes we will give a formal constructive definition of the family of sample paths of the corresponding fluid process. A general *similarity* property of the family of fluid process sample paths easily follows from this definition. Using this similarity property, it is shown in Theorem 6.1 that condition 1.6) is equivalent to the following relaxed condition.

For any fluid process $q(t)$ *with* $||q(0)|| = 1$,

$$\inf_{t \geq 0} ||q(t)|| < 1. \tag{1.8}$$

We will then show in Theorem 7.1 that the convergence (1.5) to a fluid process indeed takes place. This proves (Theorem 7.2) that property (1.6) (or equivalently property (1.8)) is sufficient for network stability. (Convergence of the type (1.5) for a multiclass queueing network was probably first derived by Chen and Mandelbaum [3]. See also [4].)

This paper contains the results presented in the talk [16]. Theorem 7.1 here is a little bit stronger: convergence is proved on the infinite time interval. The talk [16] was accepted for presentation, when Dai submitted paper [5] for publication. In his excellent work, the sufficiency of the stability condition (1.4) is generalised for multiclass queueing networks with arbitrary distributions of interarrival and service times, and "non-preemptive-within-each-customer-type" queueing disciplines in the nodes. Stability in [5] is understood as positive Harris recurrence of the underlying Markov process. The convergence (1.5) and therefore the sufficiency of condition (1.6) for network stability are also proven. Our convergence result Theorem 7.1 is less general in terms of the distributions of the interarrival and service times, but is somewhat more general with respect to the queueing disciplines allowed in the network nodes.

Remark 1.2 It is not hard to see that our definition of the family of fluid process sample paths and its properties (including Theorem 6.1) can be easily extended to the framework of [5]. Also, all results can be extended to cover a more general customer routing rule considered in [5].

We do not attempt to present a complete review of recent work on the stability of multiclass networks. Very good reviews can be found in [6], and a more recent one in [18].

The goal of our paper is

1) to prove that condition (1.6) is sufficient for network stability in a general framework *with respect to queueing disciplines* and
2) to relax condition (1.6) to the form (1.8), which makes it natural to expect that (1.6) is also necessary (or at least "almost" necessary) for network stability.

The rest of this paper is as follows. In section 2 we describe the network under consideration. In section 3 we introduce the notational conventions that are used throughout the paper. In section 4 the underlying stochastic process and the corresponding family of scaled processes are defined. In section 5 we formally define the family of sample paths of a fluid process. Section 6 discusses a similarity property of this family of fluid processes. This property allows us to prove the equivalence of conditions (1.6) and (1.8) in Theorem 6.1. In section 7 we prove convergence to a fluid process (Theorem 7.1), which implies sufficiency of condition (1.6), and consequently of the equivalent condition (1.8), for network stability (Theorem 7.2).

2. Model Description.

We consider a queueing network consisting of J nodes. The set of nodes we will also denote by $J = \{1, 2, \ldots, J\}$. There are I different customer classes. The set of customer classes will be denoted by $I = \{1, 2, \ldots, I\}$. Throughout the paper we assume that the arrival stream of class $i \in I$ customers to the network is a Poisson process with intensity λ_i, although this condition can be relaxed (see Remark 2.1 below). Each class i customer has his own route through the network,

$$\hat{\jmath}(i,1), \ldots, \hat{\jmath}(i,k), \ldots, \hat{\jmath}(i,K(i)),$$

where $K(i)$ is the length of the class i customer route and $\hat{\jmath}(i,k) \in J$ is the node in which a class i customer will be served in the k-th stage of his route. After completing service in the last stage of his route, a customer leaves the network. A class i customer in k-th stage of his route will be called a *type* (i,k) *customer*, or just an (i,k)-*customer*.

Service times for all customers of all types are independent. The service time for type (i,k)-customers is exponential with mean $v_{ik} > 0$. (This condition can be relaxed as well, see Remark 2.1 below.)

Each network node is a single-server queue with its own queueing discipline. The queueing discipline in each node is work-conserving, and satisfies the following condition.

(QD1). *The service rate assigned to each customer that is present in the node at time t, may depend only on the number of customers of the different types in the node and the time order in which they arrived at the node. (The sum of the service rates is 1 if the node is non-empty and 0 otherwise.)*

We will denote by

$$G = \{(i,k) \mid k = 1, 2, \ldots, K(i); \ \ i \in I\}$$

the set of all customer types, and by

$$G_j = \{(i,k) \in G \mid \hat{\jmath}(i,k) = j\}$$

the set of customer types that visit node $j \in J$.

We will assume that the load of each network node is less than 1, i.e.

$$\rho_j \equiv \sum_{(i,k) \in G_j} \lambda_i v_{ik} < 1, \qquad \forall j \in J. \tag{2.1}$$

It is clear that the underlying stochastic process describing the behaviour of the network is a continuous time countable Markov chain. We will say that the network is *stable* if the underlying Markov chain is ergodic.

Remark 2.1 All results of the paper hold as well if each customer class $i \in I$ has a more general interarrival time distribution. It is sufficient that it is a phase type distribution (i.e. has a rational Laplace transform) and all residual times are uniformly bounded by a distribution with finite mean. We can also allow such distributions for the service times, if the queueing disciplines in the nodes satisfy condition (QD2) below in addition to condition (QD1).

(QD2) NON-PREEMPTIVENESS WITHIN EACH CUSTOMER TYPE. *Once serving an (i,k)-customer has been started, no other (i,k)-customer can be served before the service of the customer in service has been completed. In other words, service of any customer can not be preempted by another customer of the same type.*

In the nodes with queueing disciplines not satisfying condition (QD2) (examples of such disciplines are preemptive-resume LCFS and Processor Sharing), it is essential that the the service time is exponentially distributed.

3. Notational Conventions.

Throughout the paper we will use the following notational conventions. The set of real numbers is denoted by R. The norm of a function $h = h(\xi)$ with the domain of definition $\Xi = \{\xi\}$ is defined by

$$||h|| = \sup_{\xi \in \Xi} |h(\xi)|.$$

For a finite set of functions, the norm $||\cdot||$ means the sum of the norms of the components.

If it is not stated otherwise, convergence of functions (or sets of functions) $h \to h^*$ (or $\lim h = h^*$) means point-wise convergence everywhere in their domain of definition.

The scaling operator. The scaling operator $\sigma_\alpha,\ \alpha > 0$, maps a function $h = h(\xi)$ with domain of definition $\Xi = \{\xi\}$ to the function

$$(\sigma_\alpha h)(\xi) = \frac{1}{\alpha} h(\alpha \xi)$$

with domain of definition

$$\sigma_\alpha \Xi = \{\frac{1}{\alpha}\xi \mid \xi \in \Xi\}.$$

(The case of a vector argument ξ is clearly included.)

For any set of functions $H = \{h_\gamma = h_\gamma(\xi), \gamma \in \Gamma\}$, in particular vector functions,

$$\sigma_\alpha H = \{\sigma_\alpha h_\gamma, \gamma \in \Gamma\}.$$

Two functions (or sets of functions) h^* and h are called *similar* if $h^* = \sigma_\alpha h$ for some $\alpha > 0$.

The truncation operator. The truncation operator $\pi_T,\ T \in R$, maps a function $h = h(\xi)$ with domain of definition $\Xi = \{\xi\}$ to the same function $(\pi_T h)(\xi) = h(\xi)$ with the truncated domain of definition

$$\pi_T \Xi = \Xi \bigcap \{\xi \leq T\}.$$

In the case of a vector argument ξ this truncation is applied to the domain of definition of each component function.

The set of functions $H = \{h_\gamma, \gamma \in \Gamma\}$ is mapped to the set of functions

$$\pi_T H = \{\pi_T h_\gamma \mid \gamma \in \Gamma\}.$$

4. The Underlying Stochastic Process. Family of Scaled Processes.

Consider the process

$$q(t) = \{q_{ik}(t),\ \ (i,k) \in G\}, \quad t \geq 0,$$

where $q_{ik}(t)$ is the number of (i,k)-customers in node $j = \hat{\jmath}(i,k)$ at time t. Except for some special cases, $q(t)$ is not a Markov process.

Consider the sequence of scaled processes

$$q^n(t) = \frac{1}{n} q(nt),\ t \geq 0, \quad n = 1, 2, \dots$$

or, using the definition of the scaling operator,

$$q^n = \sigma_n q.$$

The following result has been proven in [14].

LEMMA 4.1. *Suppose there exists a constant $T > 0$ such that*

$$\lim_{n \to \infty} E||q^n(T)|| = 0 \tag{4.1}$$

for an arbitrary sequence of scaled processes $q^n(t), ||q^n(0)|| = 1, \quad n = 1, 2, \dots$. Then the queueing network under consideration is stable. The sequence $||q^n(T)||$ is uniformly integrable, and therefore condition (4.1) is equivalent to convergence in probability:

$$||q^n(T)|| \to 0. \tag{4.2}$$

Following [14], we will consider the process describing the network behaviour in more detail.

Let $f_{ik}(t), t \geq 0$, be the total number of (i,k)-customers that arrived at node $j = \hat{\jmath}(i,k)$ at or before time t, including the customers present in the node at initial time 0. In order to describe the initial tate of the network properly, we will extend the functions $f_{ik}(t)$ to the domain $t < 0$. If $n = ||q(0)||$ is the initial number of customers in the network, we will assume that these customers subsequently arrived at the negative time instants

$$-\infty < t_n < t_{n-1} < \ldots < t_1 < 0.$$

The order of arrival of customers of different types corresponds to the order in which they are present in the queues at initial time 0. Now each function $f_{ik}(t)$ is well defined for all $t \in R$. Note that for some fixed $T_0 < 0$,

$$f_{ik}(t) = 0, \quad \forall t \leq T_0.$$

Let $w_{ik}(t), t \in R$, denote the sum of service times of all (i,k)-customers that arrived at node $j = \hat{\jmath}(i,k)$ before or at time t.

Let $\hat{f}_{ik}(t_1, t_2), \quad t_1, t_2 \in R$, denote the number of those (i,k)-customers that arrived at node $j = \hat{\jmath}(i,k)$ before or at time t_1 and departed before or at time t_2. Obvious properties of the functions $\hat{f}_{ik}(t_1, t_2)$ are

$$\hat{f}_{ik}(t_1, t_2) = \hat{f}_{ik}(t_2, t_2), \quad t_1 \geq t_2 \tag{4.3}$$

$$\hat{f}_{ik}(t_1, t_2) = 0, \quad t_2 \leq 0. \tag{4.4}$$

Let $\hat{w}_{ik}(t_1, t_2), \ \ t_1, t_2 \in R$, denote the total time till time t_2 spent at node $j = \hat{\jmath}(i,k)$ on serving the (i,k)-customers that arrived before or at time t_1. The functions $\hat{w}_{ik}(t_1, t_2)$ satisfy properties analogous to (4.3) and (4.4). Finally, let $u_{ik}(y), y \geq 0$, be the sum of the service times of the first $[y]$ (i,k)-customers that arrived at node $j = \hat{\jmath}(i,k)$. Here $[y]$ denotes the integer part of y.

The underlying stochastic process then is

$$s = (f, \hat{f}, w, \hat{w}, u),$$

where

$$f = \{f_{ik}(t), t \in R, \quad (i,k) \in G\}$$

and all other components are defined similarly. We will also use the notation $f = \{f_j,\ \ j \in J\}$, $f_j = \{f_{ik}, \quad (i,k) \in G_j\}$, and so on; and we write $x = (f, \hat{f})$.

The sample paths of the process s, which we also will denote by s, belong to the space

$$S = F \times \hat{F} \times W \times \hat{W} \times U,$$

where

$$F = \times_{j \in J} F_j, \quad F_j = \times_{(i,k) \in G_j} F_{ik}$$

$$F_{ik} = \{f_{ik}(t), t \in R\}$$

$$\hat{F} = \times_{j \in J} \hat{F}_j, \quad \hat{F}_j = \times_{(i,k) \in G_j} \hat{F}_{ik}$$

$$\hat{F}_{ik} = \{\hat{f}_{ik}(t_1, t_2), \ \ t_1, t_2 \in R\}$$

$$U = \times_{j \in J} U_j, \quad U_j = \times_{(i,k) \in G_j} U_{ik}$$

$$U_{ik} = \{u_{ik}(t), t \geq 0\}.$$

The spaces W and $\hat{W}$ are defined analogously to the spaces F and $\hat{F}$ respectively.

Each component space F_{ik} consists of functions such that for some $T_0 < 0$

$$f_{ik}(t) = 0, \quad t \leq T_0,$$

and $f_{ik}(t), t \in [T_0, \infty)$, is an element of the Skorohod space $D[T_0, \infty)$.

The component spaces W_{ik} and U_{ik} are defined similarly to F_{ik}.

Each space $\hat{F}_{ik}$ consists of the functions $\hat{f}_{ik}(t_1, t_2)$ that satisfy the following conditions for some $T_0 < 0$.

1) $\hat{f}_{ik}(t_1, t_2) = 0$ if $t_1 \leq T_0$ or $t_2 \leq T_0$.
2) For any fixed $t_2 \geq T_0$, the function $\hat{f}_{ik}(t_1, t_2), t_1 \geq T_0$, belongs to the Skorohod space $D[T_0, \infty)$.
3) The function $((\hat{f}_{ik}(t_1, t_2), t_1 \geq T_0), t_2 \geq T_0)$ as a function of t_2 is an element of the $D[T_0, \infty)$-valued $D[T_0, \infty)$ Skorohod space. Each component space $\hat{W}_{ik}$ is defined similarly to $\hat{F}_{ik}$. The space X is defined as a product of F and $\hat{F}$ only, $X = F \times \hat{F}$; its elements will be denoted by $x = (f, \hat{f})$.

The sample paths of the process s satisfy the following additional conditions.

a) The number of events (customer arrivals and departures) that occur on any finite time interval is finite. Clearly, the sample paths satisfy this condition with probability 1, but we can assume that the condition is true for all sample paths.
b) For each $j \in J$

$$(w_j, \hat{f}_j, \hat{w}_j) = B_j(f_j, u_j), \tag{4.5}$$

where B_j is a *deterministic* operator, specified by the queueing discipline in node j.
c) For each $i \in I$ and $k = 2, 3, \ldots, K(i)$,

$$f_{ik}(t) = f_{ik}(0) + \hat{f}_{i,k-1}(t,t), \ \ t \geq 0. \tag{4.6}$$

For any constant $c > 0$ we define a corresponding scaled process

$$s^c = (f^c, \hat{f}^c, w^c, \hat{w}^c, u^c) \equiv \sigma_c s.$$

Note that $s^1 \equiv s$.

Obviously, the scaled process $q^c = \sigma_c q$ is a projection of the process s^c. In the sequel, we will use the superscript c in our notation for the scaled processes s^c, including the original process s^1, and for projections of the process s^c.

The sample paths of any process s^c, $c > 0$, also belong to the space S; their properties are induced by the properties of the sample paths of the process s, and in particular they satisfy conditions (4.5) and (4.6):

$$(w_j^c, \hat{f}_j^c, \hat{w}_j^c) = B_j(f_j^c, u_j^c), \;\; j \in J, \tag{4.7}$$

where by definition

$$B_j(f_j^c, u_j^c) \equiv \sigma_c B_j(f_j, u_j)$$

and

$$f_{ik}^c(t) = f_{ik}^c(0) + \hat{f}_{i,k-1}^c(t,t), \; t \geq 0, \;\; i \in I, \;\; k = 2, 3, \ldots, K(i). \tag{4.8}$$

Let $S^c \subseteq S$, $\;c > 0$, be the set of all possible sample paths of the process s^c. Similarly, for example, $F^c \subseteq F$, $F_j^c \subseteq F_j$ and $F_{ik}^c \subseteq F_{ik}$ are the sets of possible sample paths of f^c, f_j^c and f_{ik}^c respectively.

The set of functions

$$f^{c,(0)} \equiv \pi_0 f^c = \{f_{ik}^c(t), t \leq 0\}$$

will be called the initial state of the process s^c.

Let

$$L \equiv \sum_I \lambda_i + \sum_G v_{ik}^{-1}.$$

Informally, L is an upper bound for the mean intensity at which events (customer arrivals and departures from any node) can occur in the network.

Consider the set of non-negative, non-decreasing, continuous functions

$$f^{(0)} = \{f_{ik}(t), t \leq 0\} \tag{4.9}$$

satisfying the Lipschitz condition with constant L as well as the condition

$$f_{ik}(aT_0) = 0, \quad \forall (i,k) \in G, \tag{4.10}$$

with

$$||f^{(0)}|| = a \geq 0 \tag{4.11}$$

and $T_0 \equiv -L^{-1} < 0$.

The following Lemma is a corollary of Lemma 4.1.

Lemma 4.2. *Suppose there exists a constant $T > 0$ such that for any constant $a > 0$, any set of functions $f^{(0)}$ specified above by (4.9)-(4.11) and any sequence of scaled processes*

$$s^c, \quad c = c_n \to \infty, \ \ n \to \infty,$$

with

$$f^{c,(0)} \to f^{(0)},$$

such that the following condition holds

$$E||q^c(aT)|| \to 0.$$

(In this case it is equivalent to convergence of $||q^c(aT)||$ to 0 in probability). Then the stochastic network under consideration is stable.

Proof: Consider a sequence of scaled processes $s^n, n \to \infty$, with $||q^n(0)|| = 1$. We assign arrival times to the customers present in the system at time 0 in the following way. If n is the number of customers at time 0 in the original (non-scaled) process, then we assume that their arrival times were

$$nT_0, (n-1)T_0, \ldots, T_0.$$

The sequence s^n has a subsequence $s^l, l \to \infty, \ \ \{l\} \subseteq \{n\}$ satisfying the conditions of the Lemma with $a = 1$. Application of Lemma 4.1 completes the proof.

It is well known, that the sequence of scaled processes described in Lemma 4.2 "should" generally speaking converge to a deterministic (fluid) process. But in the case of multiclass queueing networks we cannot expect that the limiting process will always be "truly deterministic", i.e. concentrated on a unique sample path, see for example the network with priority queueing discipline described in [14]. In Theorem 7.1 we will show that the sequence of scaled processes $s^c, c \to \infty$, converges weakly to a (fluid) process with sample paths defined as fixed points of a special operator A. Informally speaking, this operator A is obtained by the limiting transition in equations (4.7) and (4.8), as $c \to \infty$.

The following simple Lemma explains the definition of the operator A and the fluid process sample paths given in the next section. It is also an important part of the proof of the convergence result.

Lemma 4.3 (Asymptotic Lipschitz properties). *The sequence of scaled processes $s^c, c \to \infty$, defined in Lemma 4.2 satisfies the following conditions.*

For any $(i,k) \in G, \quad t, \ t_1 < t_2$, and any $\epsilon > 0$

$$\lim_{c\to\infty} P\{f^c_{ik}(t_2) - f^c_{ik}(t_1) \le L(t_2 - t_1) + \epsilon\} = 1 \tag{4.12}$$

$$\lim_{c\to\infty} P\{|v_{ik} f^c_{ik}(t) - w^c_{ik}(t)| \le \epsilon\} = 1 \tag{4.13}$$

$$\lim_{c\to\infty} P\{\hat{f}^c_{ik}(t_2, t_2) - \hat{f}^c_{ik}(t_1, t_1) \le L(t_2 - t_1) + \epsilon\} = 1 \tag{4.14}$$

$$\lim_{c\to\infty} P\{|v_{ik} \hat{f}^c_{ik}(t_1, t_2) - \hat{w}^c_{ik}(t_1, t_2)| \le \epsilon\} = 1. \tag{4.15}$$

In addition, for any $i \in I$, any $t \ge 0$ and any $\epsilon > 0$

$$\lim_{c\to\infty} P\{|f^c_{i1}(t) - f^c_{i1}(0) - \lambda_i t| \le \epsilon\} = 1. \tag{4.16}$$

The proof of this Lemma follows directly from the law of large numbers. Note that conditions (4.12), (4.13), and (4.16) hold for arbitrary service and interarrival time distributions such that all residual time distributions are bounded by a distribution with finite mean value. Conditions (4.14) and (4.15) hold under these more general assumptions, if the queueing disciplines in the corresponding nodes $\hat{\jmath}(i,k)$ satisfy condition (QD2) (see section 2). Only if the queueing disciplines in some nodes do not satisfy condition (QD2), the exponential distribution of the service times in those nodes is essential.

5. Sample paths of a fluid process.

For each $(i,k) \in G$ denote by $F_{ik,(L)} \subseteq F_{ik}$ the subset of F_{ik} consisting of the functions $f_{ik}(t)$ satisfying the condition that $f_{ik}(t)$ is non-negative, non-decreasing, and Lipschitz continuous with constant L. Furthermore, for each $(i,k) \in G$ denote by $\hat{F}_{ik,(L)} \subseteq \hat{F}_{ik}$ the subset of $\hat{F}_{ik}$ consisting of the functions $\hat{f}_{ik}(t_1,t_2)$ satisfying the following conditions:

1) $\hat{f}_{ik}(t_1,t_2)$ is a non-negative function, which is non-decreasing, continuous and Lipschitz with constant L with respect to each argument;
2) $\hat{f}_{ik}(t,t)$ as a function of t is Lipschitz with constant L;
3) $\hat{f}_{ik}(0,0) = 0$;
4) $\hat{f}_{ik}(t_1,t_2) = \hat{f}_{ik}(t_2,t_2)$ if $t_1 \geq t_2$.

The subsets $W_{ik,(L)} \subseteq W_{ik}$ and $\hat{W}_{ik,(L)} \subseteq \hat{W}_{ik}$ are defined similarly to $F_{ik,(L)}$ and $\hat{F}_{ik,(L)}$ respectively except for the Lipschitz constant, which should be chosen equal to $v_{ik}L$ instead of L.

Finally, for each $(i,k) \in G$ the subset $U_{ik,(L)} \subseteq U_{ik}$ by definition contains exactly one function

$$\bar{u}_{ik}(y) = v_{ik}y, \quad y \geq 0.$$

We will also use the notation

$$F_{j,(L)} = \times_{G_j} F_{ik,(L)}, \quad j \in J$$

$$F_{(L)} = \times_J F_{j,(L)}$$

and so on. Define the subsets

$$S_{(L)} = F_{(L)} \times W_{(L)} \times \hat{F}_{(L)} \times \hat{W}_{(L)} \times U_{(L)} \subseteq S$$

and

$$X_{(L)} = F_{(L)} \times \hat{F}_{(L)} \subseteq X.$$

Define the operator A_j mapping elements of $F_{j,(L)}$ into *subsets* f $\hat{F}_{j,(L)}$ in the following way. Fix an element (sample path) $f_j \in F_{j,(L)}$. Then $\hat{f}_j \in \hat{F}_{j,(L)}$ is an element of $A_j(f_j)$ if and only if there exists a sequence $c = c_n \to \infty$ and sequences (of sample paths)

$$f_j^c \in F_j^c, \;\; u_j^c \in U_j^c, \quad c \to \infty,$$

satisfying the conditions

$$f_j^c \to f_j \tag{5.1}$$

$$u_j^c \to \bar{u}_j \equiv \times_{G_j} \bar{u}_{ik}, \tag{5.2}$$

such that

$$\hat{f}_j^c \to \hat{f}_j \tag{5.3}$$

$$\hat{w}_j^c \to \{v_{ik}\hat{f}_{ik},\ \ (i,k) \in G_j\}, \tag{5.4}$$

where

$$(w_j^c, \hat{f}_j^c, \hat{w}_j^c) = B_j(f_j^c, u_j^c),\ \ j \in J. \tag{5.5}$$

$B_j(\cdot)$ is a deterministic operator (see (4.7)) and convergence is understood as convergence everywhere. In this case convergence everywhere is equivalent to uniform convergence on compact sets, because all functions are monotonic and the limiting functions are continuous.

Remark 5.1 A wide class of queueing disciplines (including FCFS, Priority disciplines, Head-of-the-Line Processor Sharing) has the following property.

(QD3) UNIQUENESS PROPERTY. *For any $f_j \in F_{j,(L)}$ there is a unique $\hat{f}_j = A_j(f_j)$ such that convergences (5.3) and (5.4) take place for any sequences f_j^c and u_j^c, $c \to \infty$, satisfying (5.1) and (5.2).*

Remark 5.2 The definition of the operator A_j implies its explicit expression in many cases, when the queueing discipline in node j satisfies the uniqueness condition (QD3). (For example, explicit expressions for the operator A_j for the FCFS and Priority disciplines were considered and used in [14].) In other cases, it still allows us to get properties of the elements of the set $A_j(f_j)$.

LEMMA 5.1 (CONTINUITY). *The operator A_j is continuous in the following sense. If*

$$g_l \to g,\ \ \hat{g}_l \to \hat{g}, \quad l = 1, 2, \ldots,$$

where $g, g_l \in F_{j,(L)}$, $\hat{g}_l \in A_j(g_l)$, then

$$\hat{g} \in A_j(g).$$

The operator A mapping elements of $X_{(L)}$ into *subsets* of $X_{(L)}$ is defined as follows. Fix an element (sample path) $x = (f, \hat{f}) \in X_{(L)}$. Then $x^* = (f^*, \hat{f}^*) \in A(x)$ if and only if it satisfies the following conditions:

$$\hat{f}_j^* \in A_j(f_j),\ \ \forall j \in J$$

$$f_{ik}^*(t) = \begin{cases} f_{ik}(0) + \lambda_i t,\ t \ge 0 & \text{if } k = 1 \\ f_{ik}(0) + \hat{f}_{i,k-1}^*(t,t),\ t \ge 0 & \text{if } k > 1. \end{cases}$$

Note that the operator A does not change the initial state $f^{(0)} = \pi_0 f = \pi_0 f^*$.

The continuity of the operators A_j obtained in Lemma 5.1 implies continuity of the operator A.

LEMMA 5.2 (CONTINUITY). *The operator A is continuous in the following sense. If*

$$x_l \to x,\ \ y_l \to y, \quad l = 1, 2, \ldots,$$

where $x, y, x_l \in X_{(L)}$, $y_l \in Ax_l \subseteq X_{(L)}$, then

$$y \in Ax.$$

Let $\bar{X} \subseteq X_{(L)}$, be the set of fixed points of the operator A, i.e.

$$\bar{X} = \{x \in X_{(L)} \mid x \in Ax\}.$$

Each element x of the set $\bar{X}$ will be called a *sample path of the fluid process* or just *a fluid sample path*. Denote by

$$\bar{X}(f^{(0)}) = \{x \in \bar{X} \mid \pi_0 f = f^{(0)}\}$$

the set of fluid sample paths with fixed initial state $f^{(0)}$.

Remark 5.3 It should be noted that even if the queueing discipline in each node j satisfies the uniqueness property (QD3) (i.e. each operator A_j is *not* multivalued), uniqueness of a fluid sample path is *not* guaranteed for all initial states. Thus, the dynamics of a fluid system is quite complicated even in the case of "well-behaved" queueing disciplines in nodes and a relatively simple ("output-input" type) interaction between the nodes. For a treatment of the network dynamics in the case of a more general interaction between the nodes see [12].

The q-projection of a fluid sample path x,

$$q(t) = \{q_{ik}(t) \equiv f_{ik}(t) - \hat{f}_{ik}(t,t), \quad (i,k) \in G\}$$

describes the "amount of fluid" of each type in all network nodes at any time $t \geq 0$. Therefore, the norm $||q(t)||$ can be interpreted as the total amount of fluid in the network at time t.

The continuity of the operator A (Lemma 5.2) and the fact that $\bar{X} \subseteq X_{(L)}$ imply the following properties of the family of fluid sample paths.

LEMMA 5.3.
1) *The set $\bar{X}$ is a closed set in $X_{(L)}$.*
2) *For any fixed initial state $f^{(0)}$ the set $\bar{X}(f^{(0)})$ is compact in $X_{(L)}$.*
3) *For any constant $a > 0$, the set*

$$\{x \in \bar{X} \mid ||f^{(0)}|| = a, \ f_{ik}(aT_0) = 0, (i,k) \in G\}$$

is compact in $X_{(L)}$.

6. Similarity.

Similarity is an inherent property of the family of fluid sample paths. It is implied by the definition of a fluid sample path as the limit of a sequence of sample paths of scaled processes.

LEMMA 6.1 (SIMILARITY). *The family $\bar{X}$ of the fluid sample paths is invariant with respect to the scaling operator σ_α. Namely, if $x \in \bar{X}$ then*

$$\sigma_\alpha x \in \bar{X}, \quad \forall \alpha > 0. \tag{6.1}$$

In particular, for any initial state $f^{(0)}$

$$\bar{X}(\sigma_\alpha f^{(0)}) = \sigma_\alpha \bar{X}(f^{(0)}) \quad \forall \alpha > 0.$$

PROOF: For any $j \in J$, any $f_j \in F_{j,(L)}$, and any $\alpha > 0$,

$$A_j(\sigma_\alpha f_j) = \sigma_\alpha A_j(f_j). \tag{6.2}$$

Indeed, if conditions (5.1)-(5.5) from the definition of the operator A_j hold, then

$$\sigma_\alpha f_j^c \to \sigma_\alpha f_j$$
$$\sigma_\alpha u_j^c \to \sigma_\alpha \bar{u}_j \equiv \bar{u}_j$$
$$\sigma_\alpha \hat{f}_j^c \to \sigma_\alpha \hat{f}_j$$
$$\sigma_\alpha \hat{w}_j^c \to \{v_{ik}\sigma_\alpha \hat{f}_{ik}, \quad (i,k) \in G_j\}$$
$$(\sigma_\alpha w_j^c, \sigma_\alpha \hat{f}_j^c, \sigma_\alpha \hat{w}_j^c) = B_j(\sigma_\alpha f_j^c, \sigma_\alpha u_j^c).$$

But $\sigma_\alpha f_j^c \in F_j^{\alpha c}, \sigma_\alpha \hat{f}_j^c \in \hat{F}_j^{\alpha c}$, etc. This proves equation (6.2).

Equation (6.2) implies that for any $x \in \bar{X}$, and any $\alpha > 0$,

$$A(\sigma_\alpha x) = \sigma_\alpha A x.$$

This proves the Lemma.

Obviously, a fluid sample path x contains information on the contents of the queues in all network nodes at any time t. We need to define the state of a fluid sample path at time t as a construction describing the contents of the queues at time t only, and so it will not carry extra information on the past and future of the system. This will enable us to use both the similarity and the "memoryless" properties (see Lemma 6.2) of fluid sample paths. The state $\phi^{(t)}$ defined below is such a construction. It still carries more information than just the content of the queues, but it does have the properties we need.

Consider the following projection ϕ of fluid sample paths $x \in \bar{X}$. The sample path

$$\phi = (\phi^{(t)}, t \geq 0),$$

which we will also call a fluid sample path, is defined as follows

$$\phi^{(t)} = \{\phi_{ik}^{(t)},\ \ (i,k) \in G\},$$

where

$$\phi_{ik}^{(t)} = \{\phi_{ik}^{(t)}(\xi) \equiv f_{ik}(t+\xi) - \hat{f}_{ik}(t+\xi, t),\ \ \xi \leq 0\}.$$

We will call the construction $\phi^{(t)}$ the *state* of (a deterministic process) ϕ at time t. Obviously,

$$\phi^{(0)} \equiv f^{(0)},$$

$$||\phi^{(t)}|| \equiv ||q(t)|| \equiv \sum_G q_{ik}(t),\ \ t \geq 0.$$

Denote by

$$\bar{\Phi} = \{\phi = \phi(x) \mid x \in \bar{X}\}$$

the set of all possible fluid sample paths ϕ and by $\bar{\Phi}(\phi^{(0)})$ the subset of all ϕ with fixed initial state $\phi^{(0)}$.

Similarity and continuity properties of the family of fluid sample paths x imply the corresponding properties of its ϕ-projections:

1) for any $\phi^{(0)}$*,*

$$\bar{\Phi}(\sigma_\alpha \phi^{(0)}) = \sigma_\alpha \bar{\Phi}(\phi^{(0)}), \quad \forall \alpha > 0;$$

2) for any constant $a > 0$*, the set*

$$\{x \in \bar{\Phi} \mid ||\phi^{(0)}|| = a,\ \phi_{ik}(aT_0) = 0, (i,k) \in G\}$$

is compact.

The following "memoryless" property along with similarity plays a central role in the proof of the main result of this section, Theorem 6.1.

Lemma 6.2. *Let* $\phi = (\phi^{(t)}, t \geq 0) \in \bar{\Phi}$*. Then for any* $t^* \geq 0$*,*

$$\varphi = (\varphi^{(t)} \equiv \phi^{(t^*+t)}, t \geq 0) \in \bar{\Phi}(\phi^{(t^*)}).$$

Proof: Similarly to the proof of Lemma 6.1, it is easy to verify that the sequences of the sample paths of the scaled processes defining the fluid sample path ϕ, can be used to construct the sequences defining the fluid sample path φ. We omit details.

Remark 6.1 If the queueing disciplines in all nodes satisfy the Uniqueness property (QD3) (see section 5), then the following stronger version of Lemma 6.2 is valid.

For any $\phi \in \bar{\Phi}$ *and any* $t^* \geq 0$*, the element*

$$\varphi = \{\phi^{(t)}, 0 \leq t \leq t^*,\ \ \varphi_1^{(t-t^*)}, t \geq t^*\} \in \bar{\Phi}$$

if and only if

$$\varphi_1 = (\varphi_1^{(t)}, t \geq 0) \in \bar{\Phi}(\phi^{(t^*)}).$$

We further need some simple technical observations for proving Theorem 6.1.

We will say that two states of a fluid sample path, ψ_1 and ψ_2,

$$\psi_m = \{(\psi_m)_{ik}(\xi), \xi \leq 0, \ \ (i,k) \in G\}, \quad m = 1,2$$

are *equivalent*, and we denote this by $\psi_1 \cong \psi_2$, if they are equal up to a time ξ change; more precisely, if there exists a non-decreasing function $\theta = \theta(\xi), \xi \leq 0$,

$$\theta(-\infty) = -\infty, \quad \theta(0) = 0,$$

such that

$$(\psi_1)_{ik}(\xi) = (\psi_2)_{ik}(\theta(\xi)), \xi \leq 0, \quad \forall (i,k) \in G.$$

We will call two fluid sample paths $\phi_1, \phi_2 \in \bar{\Phi}$ equivalent, if $\phi_1^{(t)} \cong \phi_2^{(t)}, \forall t \geq 0$.

LEMMA 6.3. *The initial state $\phi^{(0)}$ of any fluid sample path ϕ is equivalent to some initial state $f^{(0)}$ satisfying conditions (4.10) and (4.11).*

LEMMA 6.4. *If $\phi_1^{(0)} \cong \phi_2^{(0)}$, then the sets $\bar{\Phi}(\phi_1^{(0)})$ and $\bar{\Phi}(\phi_2^{(0)})$ coincide up to equivalence of their elements, i.e. $\phi_1 \in \bar{\Phi}(\phi_1^{(0)})$ if and only if there exists $\phi_2 \in \bar{\Phi}(\phi_2^{(0)})$ such that $\phi_1 \cong \phi_2$.*

THEOREM 6.1. *The following two properties (FL1) and (FL2) of the family $\bar{\Phi}$ of fluid sample paths are equivalent.*

(FL1) There exists a constant $T > 0$ such that for any fluid sample path $\phi \in \bar{\Phi}$, with $||\phi^{(0)}|| > 0$, the following conditions are satisfied:

$$t_* = \min\{t > 0 \mid ||\phi^{(t)}|| = 0\} \leq ||\phi^{(0)}||T$$

and

$$||\phi^{(t)}|| = 0, \quad \forall t \geq t_*.$$

(FL2) For any $\phi \in \bar{\Phi}$, with $||\phi^{(0)}|| = 1$,

$$\inf_{t \geq 0} ||\phi^{(t)}|| < ||\phi^{(0)}|| = 1,$$

or, in other words, there is no $\phi \in \bar{\Phi}$ such that $||\phi^{(t)}|| \geq ||\phi^{(0)}||, \forall t \geq 0$.

PROOF OF THEOREM 6.1: Only the implication (FL2) $\Rightarrow$ (FL1) needs to be proven. Due to the similarity property, it is sufficient to prove (FL1) for the case that $||\phi^{(0)}|| = 1$; throughout the proof we will always assume that this holds. The proof consists of the following sequence of statements (FL3) - (FL7).

(FL3). *There exists a constant $0 \leq r < 1$ such that for any $\phi \in \bar{\Phi}$ with $||\phi^{(0)}|| = 1$,*

$$\inf_{t \geq 0} ||\phi^{(t)}|| < r.$$

Suppose (FL3) is not true. Then there exist sequences $r_n \uparrow 1$, and $\phi_n \in \bar{\Phi}, ||\phi_n^{(0)}|| = 1$, $n = 1, 2, \ldots$, such that

$$(\phi_n)_{ik}(T_0) = 0, \quad \forall (i,k) \in G,$$

$$\inf_{t \geq 0} ||\phi_n^{(t)}|| \geq r_n.$$

Therefore, there exists a subsequence $\phi_l, \ \ l \to \infty, \ \ \{l\} \subseteq \{n\}$, such that

$$\phi_l \to \phi \in \bar{\Phi}, \ \ ||\phi^{(0)}|| = 1$$

and

$$\inf_{t \geq 0} ||\phi^{(t)}|| = 1.$$

This contradicts (FL2), thus proving property (FL3).

(FL4). *For any* $\phi \in \bar{\Phi}$ *with* $||\phi^{(0)}|| = 1$,

$$\inf_{t \geq 0} ||\phi^{(t)}|| = 0.$$

By virtue of (FL3) we have for arbitrary ϕ that

$$t_1 = \min\{t > 0 \mid ||\phi^{(t)}|| = r\} < \infty, \tag{6.3}$$

where $r < 1$ is the constant from the statement of (FL3). But $\phi^{(t_1)}$ is similar to the initial state $\sigma_{1/r}\phi^{(t_1)}$ with norm 1. Therefore, the similarity property implies that for any $m = 1, 2, \ldots$,

$$t_m = \min\{t > 0 \mid ||\phi^{(t)}|| = r^m\} < \infty.$$

Property (FL4) has been proven.

(FL5). *Let* $0 < r < 1$ *be fixed. Then there exists a constant* $T_1 < \infty$ *such that*

$$\sup_{||\phi^{(0)}||=1} \min\{t > 0 \mid ||\phi^{(t)}|| = r\} \leq T_1.$$

If (FL5) would not hold, then similarly to the proof of (FL3) we would obtain ϕ with $||\phi^{(0)}|| = 1$, such that

$$\inf_{t \geq 0} ||\phi^{(t)}|| \geq r.$$

This contradicts (FL4) and so property (FL5) is proved.

(FL6). *The following property holds:*

$$\sup_{||\phi^{(0)}||=1} \min\{t > 0 \mid ||\phi^{(t)}|| = 0\} < \infty.$$

Let r be the constant introduced in the statement of (FL5). Consider any ϕ with $||\phi^{(0)}|| = 1$ and the sequence $t_1 < t_2 < \ldots < t_m < \ldots$ defined by (6.3). Property (FL5) and the similarity property imply that $t_m - t_{m-1} \leq T_1 r^{m-1}$. Therefore,

$$\lim_{m \to \infty} t_m = t^* \leq T \equiv T_1/(1-r).$$

Continuity of the norm $||\phi^{(t)}||$ as a function of $t \geq 0$, implies $||\phi^{(t_*)}|| = 0$. This proves property (FL6).

(FL7). *If* $||\phi^{(t_*)}|| = 0$, *then* $||\phi^{(t)}|| = 0, \ \forall t \geq t_*$.

The set $\{t \geq t_* \mid ||\phi^{(t)}|| > 0\}$ is open. So, there exists an interval $[t_1, t_2]$ such that $||\phi^{(t_1)}|| = 0$ and $||\phi^{(t)}|| > 0, \ t_1 < t < t_2$. For any ϵ with $0 < \epsilon < \sup_{t_1 \leq t \leq t_2} ||\phi^{(t)}||$ let $t_\epsilon = \min\{t \geq t_1 \mid \ ||\phi^{(t)}|| = \epsilon\}$. Property (FL6) and similarity imply that

$$\bar{t}_\epsilon = \min\{t \geq t_\epsilon \mid ||\phi^{(t)}|| = 0\} \leq t_\epsilon + \epsilon T.$$

If $\epsilon \downarrow 0$, then $t_\epsilon \downarrow t_1$ and $\bar{t}_\epsilon \downarrow t_1$. This contradicts the choice of the interval $[t_1, t_2]$. This completes the proof of property (FL7) and consequently of Theorem 6.1.

7. Convergence to a Fluid Process. Sufficient stability condition.

Consider the sequence of scaled processes

$$s^c, \quad c \to \infty \tag{7.1}$$

introduced in Lemma 4.2. So we have

$$f^{c,(0)} \to f^{(0)}$$

with $f^{(0)}$ satisfying conditions (4.9)-(4.11). Condition (4.10) allows us to assume that the space S is restricted to the functions with time arguments $t,\ t_1,\ t_2\ \in [aT_0, \infty]$. Therefore, on each component space $F_{ik}, \hat{F}_{ik}, W_{ik}, \hat{W}_{ik}, U_{ik},\ (i,k) \in G$, we will consider the Skorohod topology and the Borel σ-algebra induced by this topology.

The σ-algebra $\mathcal{B}(S)$ on S is defined as the direct product of the σ-algebras on the component spaces.

Theorem 7.1. *Consider the sequence of scaled processes defined by (7.1). Let the measure Q^c on $(S, \mathcal{B}(S))$ be the distribution of s^c. Then*

1) *the sequence of measures Q^c, $c \to \infty$ is relatively compact;*
2) *any limiting point (in the sense of weak convergence) Q of the sequence Q^c is such that*

$$Q(\{x \in \bar{X}(f^{(0)})\}) = 1. \tag{7.2}$$

Proof:

1) Consider the sequence of projections of measures Q^c on a component space, say F_{ik}. The asymptotic Lipschitz properties from Lemma 4.3 imply density of the sequence of projections of the measures Q^c on the component space (cf. Theorem 3.7.2 in [8]). This in turn implies density and therefore relative compactness of the sequence of measures Q^c on the entire space S.
2) Let Q be a limiting point (in the sense of weak convergence) of the sequence of measures Q^c, $c \to \infty$. To simplify notation we will assume that the sequence Q^c itself converges to Q.
   Let

$$S_1 \equiv S_{(L)} \bigcap \{w_{ik} = v_{ik} f_{ik}, (i,k) \in G\} \bigcap$$
$$\{\hat{w}_{ik} = v_{ik} \hat{f}_{ik}, (i,k) \in G\} \bigcap$$
$$\{f_{i1}(t) - f_{i1}(0) = \lambda_i t, t \geq 0], i \in I\} \bigcap \{\pi_0 f = f^{(0)}\}.$$

Asymptotic Lipschitz conditions (Lemma 4.3) imply that

$$Q(S_1) = 1.$$

To verify this, we can construct a countable set of events, each of which has Q-measure 0, such that their union contains any element of the set $S \setminus S_1$. To do this, we use the following argument based on the standard properties of convergence of probability measures in a Skorohod space. Consider, for example, the component space $\hat{F}_{ik}$. ay, There exist sets C_1 and C_2 that are countably dense in $[T_0, \infty)$, such that for any $(t_1, t_2) \in C_1 \times C_2$ the projection mapping

$$s \mapsto \hat{f}_{ik}(t_1, t_2)$$

is continuous with Q-probability 1. Therefore, the distribution of any finite dimensional projection $\{\hat{f}^c_{ik}(t_1, t_2), (t_1, t_2) \in C\}$ (C is a finite subset of $C_1 \times C_2$) converges weakly to the distribution induced by the limiting measure Q. We omit further details.

For any c,

$$Q^c([S^c]) = 1,$$

where S^c is the set of all possible sample paths of the process s^c, and $[\cdot]$ denotes the closure of a set.

Then

$$Q([S_2]) = 1,$$

where $S_2 \equiv \bigcup_c S^c$.

Therefore, $Q(S_1 \bigcap [S_2]) = 1$. The set $S_1 \bigcap [S_2]$ consists of the elements of S_1 that are limiting points of the set S_2. By virtue of the definition of the operator A, we get

$$Q(\{x \in \bar{X}(f^{(0)})\}) = 1. \tag{7.3}$$

This completes the proof.

THEOREM 7.2. *Property (FL2) of the family of fluid sample paths with* $a = ||f^{(0)}|| = 1$, *is sufficient for stability of the class of stochastic networks described in section 2.*

(FL2) For any fluid sample path $x \in \bar{X}$ *and any initial state* $f^{(0)}$ *with* $||f^{(0)}|| = 1$,

$$\inf_{t \geq 0} ||q(t)|| < 1.$$

PROOF: Theorem 6.1 implies that property (FL2) is equivalent to (FL1). Therefore, there exists $T > 0$ such that for any fluid sample path $x \in \bar{X}$ with any initial state $f^{(0)}$, $||f^{(0)}|| = 1$,

$$||q(t)|| = 0, \quad \forall t \geq T. \tag{7.4}$$

Consider the sequence of processes s^c introduced in Theorem 7.1 and the sequence Q^c of their distributions on S. Then Theorem 7.1 and (7.4) imply that

$$||q^c(T)|| \to 0$$

in probability. Hence, the conditions of Lemma 4.2 with $a = 1$ (and therefore any $a > 0$ due to similarity) are satisfied. This completes the proof.

8. Conclusions.

The necessity of condition (1.6) for stability of stochastic networks is an important open problem. The relaxed form (1.8) of condition (1.6) makes it very plausible that *some form* of necessity indeed holds. Some results in this direction were recently obtained in [7] and [13].

Acknowledgement The author would like to thank Phil Fleming who reviewed the paper and made suggestions for its improvement.

References.

[1] M. Bramson, "Instability of FIFO Queueing Networks", *Annals of Applied Probability*, Vol. 4, (1994), pp. 414-431.

[2] M. Bramson, "Instability of FIFO Queueing Networks with Quick Service Times", *Annals of Applied Probability*, Vol. 4, (1994), pp. 693-718.

[3] H. Chen and A. Mandelbaum, "Open Heterogeneous Fluid Networks, with Applications to Multi-type Queues". Preprint. 1993.

[4] H. Chen, "Fluid Approximations and Stability of Multiclass Queueing Networks: Work-conserving Disciplines", *Annals of Applied Probability*, Vol. 5, (1995), pp. 637-665.

[5] J.G. Dai, "On the Positive Harris Recurrence for Open Multiclass Queueing Networks: A Unified Approach Via Fluid Limit odels", *Annals of Applied Probability*, Vol. 5, (1995), pp. 49-77.

[6] J.G.Dai, "Stability of Open Multiclass Queueing Networks via Fluid Models". In *Stochastic Networks*, IMA Volumes in Mathematics and its Applications, F.P.Kelly and R.J.Williams (eds.), 71, (1995), Springer-Verlag, New York, pp. 71-90.

[7] J.G. Dai, "Instability of Multiclass Queueing Networks", Submitted.

[8] S.N.Ethier, T.G.Kurtz, *Markov Processes: Characterization and Convergence*, Wiley, 1986.

[9] P.R. Kumar and T.I. Seidman, "Dynamic Instabilities and Stabilization Methods in Distributed Real-Time Scheduling of Manufacturing Systems", *IEEE Transactions on Automatic Control*, Vol. AC-35, (1990), pp. 289-298.

[10] S.H. Lu and P.R. Kumar, "Distributed Scheduling Based on Due Dates and Buffer Priorities", *IEEE Transactions on Automatic Control*, Vol. AC-36, (1991), pp. 1406-1416.

[11] V.A. Malyshev and M.V. Menshikov, "Ergodicity, Continuity, and Analyticity of Countable Markov Chains", *Transactions of Moscow Mathematical Society*, Vol. 39, (1979), pp. 3-48.

[12] V.A. Malyshev, "Networks and Dynamical Systems", *Advances in Applied Probability*, Vol. 25, (1993), pp. 140-175.

[13] S.P.Meyn, "Transience of Multiclass Queueing Networks via Fluid Limit Models", *Annals of Applied Probability.* Submitted.

[14] A.N. Rybko and A.L. Stolyar, "Ergodicity of Stochastic Processes Describing the Operation of Open Queueing Networks", *Problems of Information Transmission*, Vol. 28, (1992), pp. 199-220.

[15] T.I. Seidman, "First Come First Served can be Unstable!", *IEEE Transactions on Automatic Control*, Vol. 39, (1994), pp. 2166-2171.

[16] A.Stolyar, "On the Stability of Multiclass Queueing Networks", Proceedings of 2nd Conference on Telecommunication Systems - Modeling and Analysis, Nashville, March 24-27, 1994, pp. 23-36.

[17] W. Whitt, "Large Fluctuations in a Deterministic Multiclass Network of Queues", *Management Science*, Vol. 39, (1993), pp. 1020-1028.

[18] R.J. Williams, "On the Approximation of Queueing Networks in Heavy Traffic". To appear in *Stochastic Networks: Theory and Applications*, F P Kelly, S Zachary and I Ziedins (eds.), Oxford University Press Oxford, 1996.