Stability of Generalized Jackson Networks with Permanent Customers

D. Down
ddown@isye.gatech.edu
School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332

T. Bonald
tbonald@sophia.inria.fr
INRIA
2004 route des Lucioles, BP 93
06902 Sophia-Antipolis Cedex, France

Abstract

Jackson networks are typically *open* or *closed*: Either all customers join the network and eventually leave it, or no customers ever enter or exit. Here we focus on *mixed* Jackson networks, with both types of customers, general arrival streams and general service time distributions. We examine the stability of the model in terms of the positive Harris recurrence or transience of a Markov process which describes the state of the system. We show that this stability study reduces to that of an associated macroscopic deterministic model called the *fluid model*, obtained by an appropriate time-space scaling. This fluid model is shown to coincide with that associated with the equivalent open network, obtained by removing the closed component. As a result, the stability condition for the mixed Jackson network is the same as that for the equivalent open Jackson network.

1 Introduction

Jackson networks are an efficient tool for modeling complex system behavior, such as appears in packet-switched networks. As a modeling tool, one typically considers the case where the network is *open* or *closed*: Either all customers are *exogenous* (i.e. they join the network and eventually leave it), or all customers are *permanent* (i.e. they never enter or exit). Here we focus on *mixed* Jackson networks, with both types of customers. The goal of this paper is to show that the stability condition for such a mixed Jackson network is the same as that of the equivalent *open* Jackson network, obtained when the number of permanent customers is reduced to zero.

To motivate the problem, consider the case of two sources, one controlled and one uncontrolled, which send packets to the same destination (see Figure 1). The uncontrolled source sends its packets as soon as they are generated, while the controlled source cannot send more than a given number of packets without having received an acknowledgment from the destination[1]. Provided that the flow control is always active, there will always be a fixed number of packets/acknowledgments of the controlled connection in the network, so that the system is an example of a mixed Jackson network, where the exogenous customers represent the packets of the uncontrolled connection and the permanent customers those of the controlled connection. Note that the sending rate of the controlled source depends on the load of the network: When the network becomes congested, the time elapsed between the sending of a packet and the reception of the corresponding acknowledgment increases, resulting in a reduced throughput. We will show that the controlled connection cannot be responsible for congestion in the network.

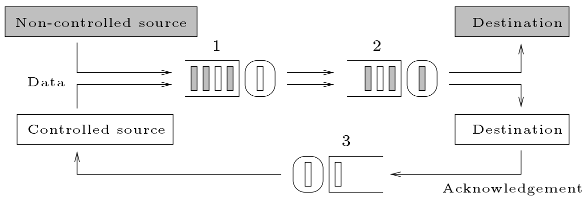

Figure 1: Example of a mixed Jackson network

Whereas the stability of open or closed Jackson networks has been studied in many papers (for example Jackson [9], Gordon and Newell [8], Borovkov [3, 4], Sigman [14, 15], Kaspi and Mandelbaum [10], and Meyn and Down [12], the stability of such mixed networks has never been considered, except for the single queue case which was studied by Boxma and Cohen [5] for a Poisson arrival process and exponential service times. Here we make general i.i.d. assumptions on the interarrival times of the exogenous customers and the service times of both types of customers at each queue, and we study the stability properties of the model in terms of the positive Harris recurrence or transience of a Markov process which describes the state of the system. Under mild additional assumptions, we show by using results from Dai [6] and Meyn [11], that this stability study reduces to that of an associated *fluid*

[1] Such a mechanism is known as window flow control.

model, obtained by an appropriate time-space scaling. This fluid model is shown to coincide with that associated with the equivalent open network, obtained by removing the permanent customers. As a result, the stability condition for the mixed Jackson network is the same as that for the equivalent open Jackson network.

2 The Model

2.1 Network structure

We consider a network composed of N queues. Each of the queues is of infinite capacity and served at unit rate by a single server under the FIFO (First In First Out) discipline. The network is populated by two types of customers, exogenous and permanent.

The exogenous customers arrive in N separate arrival streams, one for each queue in the network. If no exogenous customers arrive to a queue, the corresponding arrival stream is null, and we denote by $\mathcal{J}$ the set of non-null arrival streams. Without loss of generality, we can assume that $\mathcal{J} = \{1, \ldots, J\}$, where $J \leq N$. After the completion of its service at queue i, an exogenous customer is routed to queue j with probability R_{ij} and leaves the network with probability $1 - \sum_{j=1}^{N} R_{ij}$. The exogenous customers eventually leave the network, in other words the routing matrix $R = (R_{ij})_{1 \leq i,j \leq N}$ is taken to be transient, i.e. $I + R + R^2 + \ldots < \infty$.

The permanent customers have their own routing rules given by a routing matrix $S = (S_{ij})_{1 \leq i,j \leq N}$. The matrix S is stochastic, so that these customers never leave the network. We denote by $\mathcal{C}_1, \ldots, \mathcal{C}_K$ the K communicating classes of S, defined as the classes of the relationship

$$i \sim j \iff \exists m, n \in \mathbb{N}, \quad S_{ij}(m) > 0 \quad \text{and} \quad S_{ji}(n) > 0,$$

where for all $n \in \mathbb{N}$, we denote by $S_{ij}(n)$ the general term of the matrix S^n. We assume that these classes are irreducible, i.e. for $k = 1, \ldots, K$,

$$\forall i \in \mathcal{C}_k, \ \sum_{j \in \mathcal{C}_k} S_{ij} = 1.$$

We denote by M the total number of permanent customers in the system.

2.2 Stochastic assumptions

For any $i \in \mathcal{J}$, the interarrival times of the exogenous customers at queue i are given by $\{\tau_i^n\}_{n \geq 1}$. The service times of the exogenous customers and the service times of the permanent customers at queue i are given by $\{\sigma_i^n\}_{n \geq 1}$ and $\{\theta_i^n\}_{n \geq 1}$, respectively. These sequences of random variables on the probability space $(\Omega, \mathbb{P})$ are assumed to be i.i.d. and mutually independent. We denote by τ_i^0, σ_i^0 and θ_i^0 a generic element of each of these sequences, and by $\tilde{\tau}_i$, $\tilde{\sigma}_i$ and $\tilde{\theta}_i$ their respective distribution. We assume that each of these generic elements has a finite mean and also make the following additional assumptions:

Assumption 1 *For all $i = 1, \ldots, J$, the interarrival time τ_i^0 is unbounded and spread out, i.e. there exists a function $\bar{\tau} : \mathbb{R}_+^J \to \mathbb{R}_+^J$, non-singular with respect to Lebesgue measure, and integers $\ell_1, \ldots, \ell_J$, such that the joint distribution $\tilde{\tau}^* = (\tilde{\tau}_i^{*\ell_i})_{1 \leq i \leq J}$ (where $*\ell$ denotes ℓ-fold convolution) satisfies component-wise*

$$\forall A \in \mathcal{B}(\mathbb{R}_+^J), \quad \tilde{\tau}^*(A) \geq \int_A \bar{\tau}(u) du.$$

Assumption 2 *For all $k = 1, \ldots, K$, there exists $i_k \in \mathcal{C}_k$ such that the service time $\theta_{i_k}^0$ of the permanent customers at queue i_k is unbounded and strongly spread out, i.e. there exists a function $\bar{\theta} : \mathbb{R}_+^K \to \mathbb{R}_+^K$, strictly positive outside of a compact set, and integers $m_1, \ldots, m_K$, such that the joint distribution $\tilde{\theta}^* = (\tilde{\theta}_{i_k}^{*m_k})_{1 \leq k \leq K}$ satisfies component-wise*

$$\forall A \in \mathcal{B}(\mathbb{R}_+^K), \quad \tilde{\theta}^*(A) \geq \int_A \bar{\theta}(u) du.$$

2.3 Traffic intensities

The external arrival rate of exogenous customers at queue i is defined by $\gamma_i = 1/\mathbb{E}(\tau_i^0)$ if $i \in \mathcal{J}$, and 0 otherwise. We denote by $\sigma_i = \mathbb{E}(\sigma_i^0)$ the mean service time for an exogenous customer at queue i, and define the matrices $\Gamma = \operatorname{diag}(\gamma_1, \ldots, \gamma_N)$ and $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_N)$. Let $\gamma = (\gamma_1, \ldots, \gamma_N)'$. The vector $\lambda = (\lambda_1, \ldots, \lambda_N)'$ of *effective* arrival rates of the exogenous customers satisfies the traffic equation

$$\lambda = R'\lambda + \gamma.$$

Note that the actual arrival rates in the network would equal the effective rates if the system were stable. As the matrix R is transient, the inverse $(I - R')^{-1}$ exists, so that $\lambda = (I - R')^{-1}\gamma$. The *traffic intensity* of the exogenous flow at queue i is defined by $\rho_i = \lambda_i \sigma_i$. Denoting by ρ the vector $(\rho_1, \ldots, \rho_N)'$,

$$\rho = \Sigma(I - R')^{-1}\gamma. \tag{1}$$

3 Markovian Structure

In this section, we define a Markov process $\mathbf{X}$ (with state space E) which describes the state of the system. Then we show that compact sets of E are petite. We refer the reader to Appendix A for the main definitions and results used in the paper concerning Markov processes on general state spaces.

3.1 Markov process

At any time t, let $Q_i(t)$ be the number of exogenous customers present in queue i, including the customer in service, and define $Q(t) = (Q_1(t), \dots, Q_N(t))'$. In the same manner, we define the vector $Q^P(t)$ which gives the number of permanent customers in each queue at time t. It remains to describe the positions of the permanent customers in the queues, which can be done by means of a vector $P(t)$ of dimension M: for $m = 1, \dots, M$, $P_m(t)$ is the position in its queue of the mth permanent customer encountered when the queues $1, 2, \dots, N$ are screened in FIFO order. By convention, a customer in service is in position 0.

Example. For the network of Figure 1, we obtain

$$Q(t) = (3,3,0),\ Q^P(t) = (2,1,2),\ P(t) = (0,2,2,0,1).$$

Denote by $\alpha_i(t)$ the remaining time before the next (exogenous) customer will arrive from outside to queue i, and let $\alpha(t) = (\alpha_1(t), \dots, \alpha_N(t))'$. In the same manner, we define the vector $\beta(t) = (\beta_1(t), \dots, \beta_N(t))'$ of the remaining service times at each queue at time t. The state of the system is then described by the Markov process $\mathbf{X} = \{X(t), t \in \mathbb{R}_+\}$ defined by $X(t) = (Q(t), Q^P(t), P(t), \alpha(t), \beta(t))$. We consider the version of $\mathbf{X}$ with right-continuous sample paths and left-hand limits. Similar to Kaspi and Mandelbaum [10], it can be shown that $\mathbf{X}$ is a PDM (piecewise deterministic Markov process), and from Davis [7], $\mathbf{X}$ has the strong Markov property. The state space E of the process $\mathbf{X}$ satisfies $E \subset \mathbb{N}^N \times \mathbb{N}^N \times \mathbb{N}^M \times \mathbb{R}_+^J \times \mathbb{R}_+^N$, so that it can be endowed with the induced topology.

3.2 Structural properties

Define the element $x = 0$ of E as the state where there are no exogenous customers in the network, the permanent customers are in the queues $i_1, \dots, i_K$, and the remaining interarrival and service times are null. The proof of the following lemma is similar to that of Meyn and Down [12, Lemma 3.7]. We choose to omit it here, but it can be found in Bonald and Down [2]. It is useful to note that this is the only place where Assumptions 1 and 2 are used.

Lemma 1 *Every compact set of E is petite.*

4 Stability

The stability study of the model is approached using tools given in the work of Dai [6], which provides sufficient conditions for stability of a queueing network in terms of an associated deterministic fluid model. Thus our stability study consists of two parts: construction, followed by an analysis of, the associated deterministic fluid model for the mixed network.

4.1 Fluid model construction

Define a pseudo-norm on E by: for all $x \in E$, $\|x\| = |q| + |\alpha| + |\beta|$, where $|\cdot|$ denotes the L_1 norm. From Lemma 1, the compact sets $\{x \in E,\ \|x\| \le \ell\}_{\ell > 0}$ are petite. It follows then from Dai [6, Theorem 3.1] that the existence of some $h > 0$ such that

$$\lim_{\|x\| \to \infty} \frac{1}{\|x\|} \mathbb{E}_x \|X(\|x\| h)\| = 0,$$

is sufficient for the positive Harris recurrence of the Markov process $\mathbf{X}$. The following lemma, whose proof is similar to that of Dai [6, Lemma 4.3/4.5], shows that this condition can be reduced to the existence of some $h \ge 1$ such that

$$\frac{1}{\|x_n\|} Q(\|x_n\| h) \xrightarrow{a.s.} 0. \qquad (2)$$

The fluid model will then provide a simple criterion for (2) to hold.

Lemma 2 *For all $h > 0$, the sequences of random variables $\{\frac{1}{\|x_n\|} Q(\|x_n\| h)\}_{n \in \mathbb{N}}$, $\{\frac{1}{\|x_n\|} \alpha(\|x_n\| h)\}_{n \in \mathbb{N}}$ and $\{\frac{1}{\|x_n\|} \beta(\|x_n\| h)\}_{n \in \mathbb{N}}$ are uniformly integrable. Furthermore, whenever $h \ge 1$, we have uniformly on compact sets,*

$$\frac{1}{\|x_n\|} \alpha(\|x_n\| h) \xrightarrow{a.s.} 0 \quad \text{and} \quad \frac{1}{\|x_n\|} \beta(\|x_n\| h) \xrightarrow{a.s.} 0.$$

Let $\{x_n\}_{n \in \mathbb{N}}$ be any sequence of elements of E such that $\|x_n\|$ tends to infinity. By considering a subsequence of $\{x_n\}_{n \in \mathbb{N}}$, we can assume without loss of generality that

$$\frac{q_n}{\|x_n\|} \longrightarrow \bar{q},\ \frac{\alpha_n}{\|x_n\|} \longrightarrow \bar{\alpha} \quad \text{and} \quad \frac{\beta_n}{\|x_n\|} \longrightarrow \bar{\beta},$$

where $x_n = (q_n, q_n^P, p_n, \alpha_n, \beta_n)$. We can also assume that the customer initially in service at queue i (if there is one) is always of the same type (exogenous or permanent). Starting from the initial state x_n, the number of external arrivals of exogenous customers to queue i by time t is given by $A_i^{x_n}(t) = \max\{k \ge 0, [\alpha_n]_i + \tau_i^1 + \cdots + \tau_i^{k-1} \le t\}$, if $i \in \mathcal{J}$ and 0 otherwise. In the same manner, the *maximal* number of departures of exogenous customers from queue i by time t is given by $D_i^{x_n}(t) = \max\{k \ge 0, [\beta_n^\sigma]_i + \sigma_i^1 + \cdots + \sigma_i^{k-1} \le t\}$, where $[\beta_n^\sigma]_i = [\beta_n]_i$ if an exogenous customer is in service at time 0 at queue i, and 0 otherwise. Denote by $\Phi_{ij}(k)$ the kth routing decision for the exogenous customers, that is $\Phi_{ij}(k) = 1$ if the kth exogenous customer having completed its service at server i is routed to queue j, and 0 otherwise. Starting from the initial state x_n, the number of exogenous customers in queue i satisfies the equation:

$$\begin{aligned} Q_i^{x_n}(t) &= Q_i^{x_n}(0) + A_i^{x_n}(t) \\ &\quad + \sum_{j=1}^N \Phi_{ji}(D_j^{x_n}(T_j^{x_n}(t))) - D_i^{x_n}(T_i^{x_n}(t)), \end{aligned}$$

where $T_i^{x_n}(t)$ is the total time up to time t spent by server i on exogenous customers. The proofs of the following results are similar to those of Dai [6, Lemma 4.3 and Theorem 4.1]. We define the vectors $T^{x_n}(t) = (T_1^{x_n}(t), \ldots, T_N^{x_n}(t))'$ and $e = (1, \ldots, 1)'$.

Lemma 3 *When $n \to \infty$, we have uniformly on compact sets,*

$$\begin{aligned} \frac{1}{\|x_n\|}\Phi(\lfloor \|x_n\| t \rfloor) &\xrightarrow{a.s.} Rt, \\ \frac{1}{\|x_n\|}A^{x_n}(\|x_n\| t) &\xrightarrow{a.s.} \Gamma(et - \bar{\alpha})^+, \\ \frac{1}{\|x_n\|}D^{x_n}(\|x_n\| t) &\xrightarrow{a.s.} \Sigma^{-1}(et - \bar{\sigma})^+, \end{aligned}$$

where $\lfloor t \rfloor$ is the integer part of t, $(\cdot)^+$ denotes $\max(\cdot, 0)$ component-wise, and $\bar{\sigma}$ is the vector whose ith component is equal to $\bar{\beta}_i$ if an exogenous customer is initially in service at queue i and 0 otherwise.

Theorem 1 *For almost all sample paths $\omega \in \Omega$, there is a subsequence $\{x_{n_k}\}_{k \in \mathbb{N}}$ with $\|x_{n_k}\| \to \infty$ such that uniformly on compact sets,*

$$\frac{1}{\|x_{n_k}\|}(Q^{x_{n_k}}(\|x_{n_k}\| t), T^{x_{n_k}}(\|x_{n_k}\| t)) \to (\bar{Q}(t), \bar{T}(t)).$$

$(\bar{Q}, \bar{T})$ is a fluid limit *of the model and satisfies the equations*

$$\bar{Q}(t) = \bar{Q}(0) + \Gamma(et - \bar{\alpha})^+ - (I - R')\Sigma^{-1}(\bar{T}(t) - \bar{\sigma})^+, \quad (3)$$

$$0 \leq \bar{T}(s) - \bar{T}(t) \leq e(s - t), \quad (4)$$

where (4) holds component-wise for all $s \geq t$.

Definition 1 *The collection of all fluid limits is called the* fluid model*. The fluid model is said to be stable if there exists a constant $h > 0$, such that for any fluid limit $(\bar{Q}, \bar{T})$,*

$$\bar{Q}(\cdot + h) \equiv 0.$$

The following theorem, due to Dai [6, Theorem 4.2], connects the stability of the fluid model to that of the original network.

Theorem 2 *If the fluid model is stable, the limit (2) holds, so that $\mathbf{X}$ is positive Harris recurrent. In particular, there exists a unique invariant probability π.*

To provide a partial converse to Theorem 2 we first introduce the notion of *instability* of the fluid model.

Definition 2 *The fluid model is said to be unstable if there exist constants $a > 0$ and $b \geq 0$, such that for every fluid limit $(\bar{Q}, \bar{T})$,*

$$|\bar{Q}(t)| \geq at - b.$$

The following theorem, due to Meyn [11, Theorem 3.2], connects the instability of the fluid model to that of the original network.

Theorem 3 *If the fluid model is unstable, then the Markov process $\mathbf{X}$ is transient.*

4.2 Stability of the fluid model

Note that the residual interarrival and service times $\bar{\alpha}$ and $\bar{\beta}$ simply introduce delays in the evolution of the fluid model, and in this way do not affect the stability analysis. Since for any fluid limit, $|\bar{q}| + |\bar{\alpha}| + |\bar{\beta}| = 1$, we consider in the following the evolution of the system from time $t = 1$. Defining the *workload* of the network as

$$\bar{W}(t) = \Sigma(I - R')^{-1}\bar{Q}(t), \quad (5)$$

it follows from (1) and (3) that for all $t \geq 1$,

$$\bar{W}(t) = \bar{W}(1) + \rho(t - 1) - (\bar{T}(t) - \bar{T}(1)). \quad (6)$$

Theorem 4 *If $\rho_i > 1$ for some i then the fluid model is unstable and thus the Markov process $\mathbf{X}$ is transient.*

Proof: Let i be such that $\rho_i > 1$. It follows from (4) and (6) that for all $t \geq 1$, $\bar{W}_i(t) \geq \bar{W}_i(1) + (\rho_i - 1)(t - 1)$. From the definition of the workload, $\bar{Q}(t) = (I - R')\Sigma^{-1}\bar{W}(t)$. Since the quantity of fluid $\bar{Q}$ and the workload $\bar{W}$ are nonnegative functions,

$$\forall j = 1, \ldots, N, \quad \bar{W}_j(t) \geq \sigma_j \frac{R_{ij}}{\sigma_i}\bar{W}_i(t).$$

Denoting by $\mathcal{I}_k$ the set of kth successors of i, i.e. $\mathcal{I}_k = \{j = 1, \ldots, N,\ R_{ij}(k) > 0\}$, where $R_{ij}(k)$ denotes the general term of the matrix R^k, there exists $\varepsilon_1 > 0$ such that for all $j \in \mathcal{I}$, $\bar{W}_j(t) \geq \varepsilon_1(t - 1)$. Define the set of all possible successors of i, $\mathcal{I}$, as the union over $k \in \mathbb{N}$ of $\mathcal{I}_k$.

$$\mathcal{I} = \bigcup_{k \in \mathbb{N}} \mathcal{I}_k = \bigcup_{k=1}^{N} \mathcal{I}_k.$$

By induction, there exists $\varepsilon > 0$ such that for all $j \in \mathcal{I}$, $\bar{W}_j(t) \geq \varepsilon(t - 1)$. Also, there exists some $j \in \mathcal{I}$ such that $\sum_{j'=1}^{N} R_{jj'} < 1$, so that

$$|\bar{Q}(t)| \geq (1 - \sum_{j'=1}^{N} R_{jj'})\varepsilon(t - 1).$$

The result follows from Theorem 3. ∎

Given the above theorem, for the rest of the paper we assume that $\rho_i \leq 1$ for all i. From (4), the function $\bar{T}$ is Lipschitz continuous so that its derivative exists at almost every point. In the following, we denote by $\dot{\bar{T}}(t)$ the derivative of $\bar{T}$ at t, where t is implicitly assumed to be a regular point of $\bar{T}$. From (5) and (6), the functions $\bar{Q}$ and $\bar{W}$ are also Lipschitz continuous and regular points of $\bar{T}$ are regular points of $\bar{Q}$ and $\bar{W}$.

The following lemma, which states that when there is some exogenous fluid at a queue, this fluid is served at unit rate, is the key to the demonstration of stability. Indeed, a direct consequence of this lemma is that the fluid model associated with the mixed Jackson network is the same as that associated with the equivalent open Jackson network.

Lemma 4 *For all $t > 1$, we have*

$$\bar{Q}(t).(e - \dot{\bar{T}}(t)) = 0.$$

Proof: Assume that for some i and some $t > 1$, we have $\bar{Q}_i(t) > 0$. As $\bar{Q}$ is Lipschitz continuous, there exists $1 < u < t$ such that $t - u < \sigma_i \bar{Q}_i(u)$. Let $u_n = u\|x_n\|$ and denote by v_n the time at which all customers (including the permanent ones) present in queue i at time u_n are served. As the service discipline is FIFO, defining $\delta_i(u_n)$ (resp. $\delta_i^P(u_n)$) equal to 1 if at time u_n an exogenous customer (resp. a permanent customer) is in service at queue i,

$$\begin{aligned} v_n &= u_n + \beta_i(u_n) + \sum_{k=1}^{Q_i(u_n)-\delta_i(u_n)} \sigma_i^{k_n^\sigma + k} \\ &+ \sum_{k=1}^{Q_i^P(u_n)-\delta_i^P(u_n)} \theta_i^{k_n^\theta + k}, \end{aligned}$$

where k_n^σ and k_n^θ are stopping times of the sequences $\{\sigma_i^k\}$ and $\{\theta_i^k\}$. As the number of permanent customers is bounded by M in any queue, it follows from Lemma 2 and the strong law of large numbers that $v_n/\|x_n\| \to v = u + \sigma_i \bar{Q}_i(u)$, and we have $\bar{T}_i(v) - \bar{T}_i(u) = v - u$. From (4), we obtain for all $u \le s \le v$, $\bar{T}_i(s) - \bar{T}_i(u) = s - u$. As $u < t < v$, we finally get $\dot{\bar{T}}_i(t) = 1$. ∎

At this point, we could quote Dai, who gives in [6] a sufficient condition for the stability of the fluid model associated with an open Jackson network. Instead, we provide a new approach to the proof, which uses the following conservation property.

Lemma 5 *For all $t > 1$, we have component-wise*

$$\dot{\bar{T}}(t) \ge \rho.$$

Proof: From (5) and (6) we get for all $t > 1$,

$$\dot{\bar{Q}}(t) = (I - R')\Sigma^{-1}(\rho - \dot{\bar{T}}(t)). \qquad (7)$$

Let $t > 1$ be any regular point of $\bar{T}$, and denote by $\mathcal{L}$ the set of indices i such that $\bar{Q}_i(t) = 0$. The function $\bar{Q}_i$ is nonnegative, so $\dot{\bar{Q}}_i(t) = 0$ for all $i \in \mathcal{L}$. We can assume without loss of generality that $\mathcal{L} = \{1, \ldots, L\}$ and denote by V_1 (resp. V_2) the vector which consists of the first L (resp. last $N - L$) components of $V = \rho - \dot{\bar{T}}(t)$. In the same way, we define $I = \mathrm{diag}(I_1, I_2)$, $\Sigma = \mathrm{diag}(\Sigma_1, \Sigma_2)$ and

$$R = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}$$

where I_1, R_{11} and Σ_1 are matrices of dimension $L \times L$. From (7), $(I_1 - R'_{11})\Sigma_1^{-1}V_1 - R'_{12}\Sigma_2^{-1}V_2 = 0$. Define $U_1 = (I_1 - R'_{11})\Sigma_1^{-1}V_1$. As for all $i \notin \mathcal{L}$, we have $\bar{Q}_i(t) > 0$, it follows from Lemma 4 that for all $i = L+1, \ldots, N$, $\dot{\bar{T}}_i(t) = 1$. In particular, $V_2 \le 0$ so that $U_1 \le 0$. As the matrix R_{11} is transient and positive, $V_1 \le 0$. Therefore, $V \le 0$ and $\dot{\bar{T}}(t) \ge \rho$. ∎

Theorem 5 *If $\rho_i < 1$ for all i then the fluid model is stable and thus the Markov process $\mathbf{X}$ is positive Harris recurrent.*

Proof: From (6), $\dot{\bar{W}}(t) = \rho - \dot{\bar{T}}(t)$. The result is a simple consequence of Lemma 5 and Theorem 2. ∎

Note that in the particular case of a closed network (J=0), the fluid model trivially empties since $\bar{Q}(t) = 0$ for all t. Hence we get the following result, similar to that obtained in [10, Theorem 2.1].

Corollary 1 *A closed queueing network satisfying Assumption 2 is always stable. In particular, the associated Markov process $\mathbf{X}$ is positive Harris recurrent.*

5 Conclusion

We have shown under mild assumptions on the interarrival times of the exogenous customers and the service times of the permanent customers, that the stability condition for a mixed generalized Jackson network is the same as that for the equivalent open Jackson network. The approach consists of reducing the stability analysis to that of an associated deterministic fluid model, and showing that this fluid model coincides with that for the equivalent open network.

In this respect, the key result is that of Lemma 4, which ensures that in some sense the effect of the permanent customers becomes negligible under the fluid scaling. It is worth noting that this property holds under many work-conserving service disciplines. Hence our results which were proved under the FIFO discipline for convenience, remain valid for instance under the processor sharing discipline, under any priority discipline as soon as the exogenous customers have priority over the permanent customers, or under any combination of these service disciplines.

Finally, the results of this paper may be easily extended to other queueing networks, provided that their stability analysis may be reduced to that of a fluid model, and that under the service discipline considered, the fluid model for the mixed network coincides with that for the equivalent open network. Bonald studied in [1] the case of certain polling models, and we are currently working on the case of multiclass networks.

Appendix A gives the main definitions and results on Markov processes used in the paper. We refer the reader to Meyn and Tweedie [13] for a more complete presentation on the subject.

A Markov processes on a general state space

Let $\mathbf{X} = \{X(t), t \in \mathbb{R}_+\}$ be a Markov process on a probability space $(\Omega, \mathbb{P})$. Assume that $\mathbf{X}$ satisfies the strong Markov property and has right-continuous sample paths. The state space of $\mathbf{X}$ is a topological space $(E, \mathcal{B}(E))$. For all $t \geq 0$, denote by P^t the transition semigroup for $\mathbf{X}$ defined for all $x \in E$, $A \in \mathcal{B}(E)$, by $P^t(x, A) = \mathbb{P}_x(X(t) \in A)$, where $\mathbb{P}_x$ is the probability measure for the process with initial state x. For any set $A \in \mathcal{B}(E)$, the occupation time is defined by

$$\eta_A = \int_0^\infty \mathbb{I}\{X(t) \in A\} dt.$$

For any distribution ν of $\mathbb{R}_+$, we denote by K_ν the Markov transition function for the chain sampled by ν as

$$\forall x \in E,\ K_\nu(x, \cdot) = \int_0^\infty P^t(x, \cdot)\nu(dt).$$

A set $A \in \mathcal{B}(E)$ is said to be *petite* if for some sampling distribution ν and a nontrivial measure μ on E, we have for all $x \in A$, $K_\nu(x, \cdot) \geq \mu(\cdot)$. A set $A \in \mathcal{B}(E)$ is said to be *Harris recurrent* if for all $x \in E$, $\mathbb{P}_x(\eta_A = \infty) = 1$. The Markov process $\mathbf{X}$ is Harris recurrent if for some σ-finite measure μ, any set $A \in \mathcal{B}(E)$ such that $\mu(A) > 0$ is Harris recurrent. There then exists a measure π called the *invariant measure* of $\mathbf{X}$, unique up to scalar multiples, which satisfies for all $t \geq 0$,

$$\pi(A) = \int_E P^t(x, A)\pi(dx), \qquad A \in \mathcal{B}(E).$$

A Harris recurrent process $\mathbf{X}$ is *positive Harris recurrent* if its invariant measure is finite, and thus may be normalized to a probability measure, the unique *stationary distribution* π of $\mathbf{X}$.

Acknowledgement. This work was initiated while the second author was at INRIA Sophia-Antipolis with the support of a postdoctoral fellowship from the European Research Consortium for Informatics and Mathematics. The second author was supported in part by France Telecom/CNET.

References

[1] T. Bonald, “Stabilité d'un système de polling mixte,” submitted to *Revue d'Automatique, Informatique et Recherche Opérationnelle,* 1997.

[2] T. Bonald and D. Down, “Stability of mixed generalized Jackson networks,” in preparation, 1997.

[3] A. Borovkov, “Limit theorems for queueing networks I,” *Theory Prob. Appl.,* vol. 31, pp. 413-427, 1986.

[4] A. Borovkov, “Limit theorems for queueing networks II,” *Theory Prob. Appl.,* vol. 32, pp. 257-272, 1988.

[5] O.J. Boxma, and J.W. Cohen, “The M/G/1 queue with permanent customers,” *IEEE Journal on Selected Areas on Communications,* vol. 9, pp. 179-184, 1991.

[6] J.G. Dai, “On positive Harris recurrence of multiclass queueing networks : A unified approach via fluid limit models,” *Annals of Applied Probability,* vol. 5, pp. 49-77, 1995.

[7] M.H.A. Davis, “Piecewise deterministic Markov processes : a general class of diffusion stochastic models,” Journal of the Royal Statistical Society series B, vol. 46, pp. 353-388, 1984.

[8] W. Gordon, and G. Newell, “Closed queueing systems with exponential servers,” *Oper. Res.,* vol. 15, pp. 254-265, 1967.

[9] J.R. Jackson, “Jobshop-like queueing systems”, *Manag. Science,* vol. 10, pp. 131-142, 1963.

[10] H. Kaspi and A. Mandelbaum, “Regenerative closed queueing networks,” *Stoch. and Stoch. Reports,* vol. 39, pp. 239-258, 1992.

[11] S. Meyn, “Transience of multiclass queueing networks via fluid limit models,” *Ann. Appl. Prob.,* vol. 5, pp. 946-957, 1995.

[12] S. Meyn and D. Down, “Stability of generalized Jackson networks,” *Ann. Appl. Prob.,* vol. 4, pp. 124-149, 1994.

[13] S.P. Meyn and R.L. Tweedie, “Stability of Markovian processes II: Continuous time processes and sample chains,” *Advances in Applied Probability,* vol. 25, pp. 487-517, 1993.

[14] K. Sigman, “Notes on the stability of closed queueing networks,” *J. Appl. Prob.,* vol. 26, pp. 678-682, 1989.

[15] K. Sigman, “The stability of open queueing networks,” *Stoch. Proc. and Appl.,* vol. 35, pp. 11-25, 1990.