

# STABILIZATION OF AN OVERLOADED QUEUEING NETWORK USING MEASUREMENT-BASED ADMISSION CONTROL

LASSE LESKELÄ,\* *Helsinki University of Technology*

## Abstract

Admission control can be employed to avoid congestion in queueing networks subject to overload. In distributed networks, the admission decisions are often based on imperfect measurements on the network state. In this paper, we study how the lack of complete state information affects the system performance, by considering a simple network model for distributed admission control. The stability region of the network is characterized and it is shown how feedback signaling makes the system very sensitive to its parameters.

*Keywords:* Queueing network; admission control; stability; overload; perturbed Markov process

2000 Mathematics Subject Classification: Primary 60K25

Secondary 68M20; 90B15; 90B22

## 1. Introduction

Consider an overloaded queueing network in which the incoming traffic exceeds the service capacity over a long time period. In such a case, it is often necessary to employ admission control to avoid the network becoming fully congested. Many networks of practical interest are composed of subnetworks, not all of which are administrated over by a single party. In such a network, the admission controller seldom has complete, up-to-date system information available. Instead, the admission decisions must be based on partial measurements on the network state.

In this paper, we study the effect of imperfect information on the performance of the admission control scheme. Typical performance measures in this kind of setting include the average amount of rejected traffic per unit time and the mean proportion of time during which the network load is undesirably high. However, assuming that the network under study is subjected to long-term overload, there is another performance criterion that must first be analyzed by answering the following question: *if the network is subjected to a stationary load exceeding the service capacity, how strict must the admission control rules be set in order to stabilize the system?*

To deal with the question mathematically, it is assumed that the network can be modeled using the simplest nontrivial model for a distributed network, the two-node tandem network with independent, exponential service times and unlimited buffers. The network state is denoted by  $X = (X_1, X_2)$ , where  $X_i$  is the number of jobs in node  $i$ . It is assumed that the admission control can be modeled so that the input to the system is a Poisson process with a stochastic time-varying intensity  $\lambda \equiv \lambda(X)$ , a function of the network state.

The lack of complete state information is reflected in the model by assuming that the input rate  $\lambda$  is a function of only one of the  $X_i$ . If  $\lambda(X) \equiv \lambda(X_1)$  then the analysis of the system can

Received 1 April 2004; revision received 24 March 2005.

\* Postal address: Institute of Mathematics, PO Box 1100, FI-02015 TKK, Finland. Email address: lasse.leskela@iki.fi

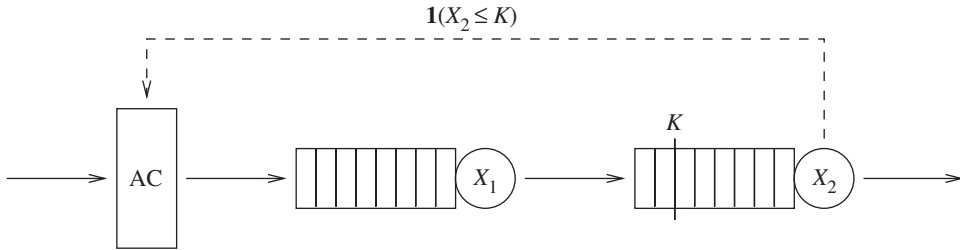


FIGURE 1: Admission control based on the single threshold level  $K$ .

be reduced to the study of birth–death processes, which are well understood. This is why in the following it is always assumed that  $\lambda(X) \equiv \lambda(X_2)$ ; so that the admission control introduces a feedback signaling loop to the system. For example, we can model a network in which arriving traffic is rejected when the size of the second buffer exceeds a threshold level  $K$  by setting  $\lambda(X) = \mathbf{1}(X_2 \leq K)$ , an indicator function (see Figure 1). In order to also cover more complex admission policies, with multiple thresholds and thinning of input traffic, the shape of  $\lambda(X_2)$  will not be restricted in any way.

More precisely,  $X$  is defined as a continuous-time stochastic process, as follows. Let  $\lambda$  be a nonnegative function on  $\mathbb{Z}_+$ , and let  $\mu_1, \mu_2 > 0$ . Define the transition rates  $q(x, y)$ , for  $x \neq y, x, y \in \mathbb{Z}_+^2$ , by

$$q(x, y) = \begin{cases} \lambda(x_2), & y = x + e_1, \\ \mu_1, & y = x - e_1 + e_2, \\ \mu_2, & y = x - e_2, \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where  $e_i$  denotes the  $i$ th unit vector of  $\mathbb{Z}_+^2$  and  $x_i$  denotes the  $i$ th component of  $x$ . As usual, set  $q(x, x) \equiv -q(x)$ , where the transition rate out of state  $x$  is defined by

$$q(x) = \sum_{y \neq x} q(x, y).$$

It is clear that  $q(x) < \infty$  for all  $x$ , so, using the minimal construction [3, pp. 39–45], the rates  $q(x, y)$  define a unique Markov process  $X$  on  $\mathbb{Z}_+^2 \cup \{\kappa\}$ . Here  $\kappa$  denotes an additional state not in  $\mathbb{Z}_+^2$ , with

$$T_\kappa := \inf\{t > 0: X(t) = \kappa\} \leq \infty$$

being the time of explosion of  $X$ . The notation  $S(\lambda, \mu_1, \mu_2)$  will be used for the set of transition rates corresponding to the triple  $(\lambda, \mu_1, \mu_2)$ , and the system  $S(\lambda, \mu_1, \mu_2)$  will be said to be *stable* if the corresponding Markov process is ergodic, that is, irreducible and positive recurrent.

In its most general form, the stability problem may now be stated as follows.

**Problem 1.** Characterize the set of all  $(\lambda, \mu_1, \mu_2) \in \mathbb{R}_+^{\mathbb{Z}_+} \times \mathbb{R}_+ \times \mathbb{R}_+$  for which  $S(\lambda, \mu_1, \mu_2)$  is stable.

Specializing to networks with threshold-based admission control, the offered traffic is assumed to arrive at unit rate, without loss of generality. With the admission threshold denoted by  $K$ , Problem 1 now takes the following form.

**Problem 2.** For each  $(\mu_1, \mu_2) \in \mathbb{R}_+^2$ , determine the values of  $K \in \mathbb{Z}_+ \cup \{\infty\}$ , if any, for which the system  $S(\mathbf{1}(\cdot \leq K), \mu_1, \mu_2)$  is stable.

Note that the system corresponding to  $K = \infty$  in Problem 2 is the ordinary tandem queue, for which it is well known that the condition  $\min(\mu_1, \mu_2) > 1$  is sufficient and necessary for stability. However, assuming overload, verifying the existence of a threshold level that can stabilize the system is not as straightforward.

The queueing systems literature includes a vast amount of work on various admission control mechanisms. However, most earlier studies on tandem networks require at least one of the buffers to be finite, so that the two-dimensional nature of the problem can partly be made one-dimensional by applying matrix-geometric methods [8]. For networks with unlimited buffers and state-dependent service times, Bambos and Walrand [4] provided stability results extending to non-Markovian systems; this, however, rules out networks with the type of feedback signaling loop considered here. Concerning the network  $S(\lambda, \mu_1, \mu_2)$  defined above, the compensation approach introduced by Adan *et al.* [1] can be used to compute the invariant measure in the special case where  $\lambda$  is constant on  $\{n \in \mathbb{Z}_+, n \geq 1\}$ . For more general input rates, Leskelä and Resing [6] have described a numerical method for calculating stationary performance characteristics of the system. Altman *et al.* [2] have recently introduced perturbation techniques that seem appropriate for asymptotically analyzing the behavior of  $S(\lambda, \mu_1, \mu_2)$  under suitable parameter scaling.

This paper partially answers Problem 1 by deriving conditions sufficient and necessary for stability. Furthermore, by showing that the sufficient and necessary conditions coincide in the special case of threshold-based admission control, a complete solution of Problem 2 is given. In addition, the sensitivity of the system is analyzed with respect to changes in the service rates, and it is shown how acceleration of one of the servers may, rather paradoxically, destabilize the system.

### 2. A sufficient condition for stability

Let  $S$  be a countable set. For a function  $V : S \rightarrow \mathbb{R}$ , let

$$\lim_{x \rightarrow \infty} V(x) = \infty$$

if the set  $\{x : V(x) \leq M\}$  is finite for all  $M \in \mathbb{R}$ . Furthermore, we denote the mean drift of  $V$ , with respect to transition rates  $q(x, y)$ , by

$$\Delta V(x) = \sum_{y \neq x} (V(y) - V(x))q(x, y), \tag{2}$$

assuming that the sum on the right-hand side converges.

**Definition 1.** A map  $V : S \rightarrow \mathbb{R}$  is called a *Lyapunov function* for  $q$  if it satisfies the following conditions, called *Foster’s criteria*:

- (F1)  $\sum_{y \neq x} |V(y) - V(x)|q(x, y) < \infty$  for all  $x$  (so that the right-hand side of (2) makes sense).
- (F2)  $\lim_{x \rightarrow \infty} V(x) = \infty$ .
- (F3) There is a finite set  $S_0 \subset S$  such that  $\sup_{x \in S \setminus S_0} \Delta V(x) < 0$ .

The following continuous-time analogue of Foster’s classic theorem [5] provides a condition sufficient for stability.

**Theorem 1.** (Tweedie [9].) *Let  $X$  be an irreducible Markov process on a countable state space  $S$ , generated by transition rates  $q(x, y)$  such that  $q(x) < \infty$  for all  $x$ . The existence of a Lyapunov function for  $q$  is then sufficient for  $X$  to be ergodic.*

Consider the system  $S(\lambda, \mu_1, \mu_2)$ , and let  $q(x, y)$  be as defined in (1). Assume  $V$  to be a function on  $\mathbb{Z}_+^2$  of the form  $V(x) = x_1 + v(x_2)$ , for some  $v: \mathbb{Z}_+ \rightarrow \mathbb{R}$  with  $v(0) = 0$ . To search for a Lyapunov function of this type, let us fix a number  $r > 0$  and require that the mean drift of  $V$ , with respect to  $q$ , satisfies

$$\Delta V(x) = -r \quad \text{for all } x \text{ with } x_1 > 0. \tag{3}$$

It is straightforward to verify that (3) is equivalent to

$$v(1) = 1 - \frac{\lambda(0) + r}{\mu_1},$$

$$v(n + 1) = 1 - \frac{\lambda(n) + r}{\mu_1} + \left(1 + \frac{\mu_2}{\mu_1}\right)v(n) - \frac{\mu_2}{\mu_1}v(n - 1), \quad n \geq 1.$$

With  $\alpha(n) = 1 - (\lambda(n) + r)/\mu_1$  and  $w(n) = v(n + 1) - v(n)$ , the above difference equation can be written as  $w(n) = \alpha(n) + (\mu_2/\mu_1)w(n - 1)$  for  $n \geq 1$ , with  $w(0) = \alpha(0)$ . Thus,  $w(n) = \sum_{k=0}^n \alpha(k)(\mu_1/\mu_2)^{k-n}$ , implying that

$$v(n) = \sum_{j=0}^{n-1} w(j) = \sum_{j=0}^{n-1} \sum_{k=0}^j \alpha(k) \left(\frac{\mu_1}{\mu_2}\right)^{k-j},$$

and we conclude that, for each  $r > 0$ , (3) defines the function

$$V_r(x) = x_1 + \sum_{j=0}^{x_2-1} \sum_{k=0}^j \left(1 - \frac{\lambda(k) + r}{\mu_1}\right) \left(\frac{\mu_1}{\mu_2}\right)^{k-j}.$$

Thus, we have constructed a family of functions  $\mathcal{V} = \{V_r, r > 0\}$  whose elements satisfy  $\sup_{\{x: x_1 > 0\}} \Delta V_r(x) < 0$ , so we might hope that  $V_r$  satisfies (F3) for a suitably chosen finite subset of  $\mathbb{Z}_+^2$ . In order to investigate whether this is the case, let us study the mean drift of  $V_r$  for  $x = (0, n)$ ,  $n \geq 1$ :

$$\Delta V_r(0, n) = \lambda(n) - \mu_2(v_r(n) - v_r(n - 1)). \tag{4}$$

**Definition 2.** For a  $z \geq 0$ , let  $Z_n \sim \text{geom}_n(z)$  if  $Z_n$  is a random variable on  $\mathbb{Z} \cap [0, n]$  with  $P(Z_n = j) = cz^j$ . For  $0 \leq z \leq 1$ , let  $Z \sim \text{geom}(z)$  if the random variable  $Z$ , on  $\mathbb{Z}_+$ , satisfies  $P(Z = j) = (1 - z)z^j$ .

In this paper,  $Z_n$  and  $Z$  will always represent generic random variables with respective distributions  $\text{geom}_n(\mu_1/\mu_2)$  and  $\text{geom}(\mu_1/\mu_2)$ . Using this notation, we may verify that (4) can be alternatively written as

$$\Delta V_r(0, n) = \frac{E(\lambda(Z_n)) - \mu_2(1 - r/\mu_1)P(Z_n > 0)}{P(Z_n = n)}, \quad Z_n \sim \text{geom}_n\left(\frac{\mu_1}{\mu_2}\right). \tag{5}$$

We will also use ‘ $\overline{\lim}$ ’ and ‘ $\underline{\lim}$ ’ as shorthand for ‘lim sup’ and ‘lim inf’, respectively.

**Theorem 2.** *The family  $\mathcal{V} = \{V_r, r > 0\}$  contains a Lyapunov function for  $S(\lambda, \mu_1, \mu_2)$  if and only if*

$$\overline{\lim} E(\lambda(Z_n)) < \min(\mu_1, \mu_2), \quad Z_n \sim \text{geom}_n(\mu_1/\mu_2). \tag{6}$$

*In particular, if  $\lambda(0) > 0$  then (6) is sufficient for the stability of  $S(\lambda, \mu_1, \mu_2)$ .*

The proof of the theorem will use the following two lemmas.

**Lemma 1.** *Condition (6) is equivalent to  $\overline{\lim} \Delta V_r(0, n) < 0$ , for some  $r > 0$ .*

*Proof.* Let  $Z_n \sim \text{geom}_n(\mu_1/\mu_2)$  for  $n \geq 0$ . First observe that, since  $\lim P(Z_n > 0) = \min(1, \mu_1/\mu_2)$ ,

$$\overline{\lim} E(\lambda(Z_n)) - \min(\mu_1, \mu_2) = \overline{\lim} \{E(\lambda(Z_n)) - \mu_2 P(Z_n > 0)\}. \tag{7}$$

Now assume that (6) holds. We can then choose an  $r > 0$  such that

$$E(\lambda(Z_n)) - \mu_2 P(Z_n > 0) \leq -r$$

for  $n$  sufficiently large. It follows that  $\overline{\lim} \Delta V_r(0, n) < 0$ , since, using (5), we see that, for large  $n$ ,

$$\Delta V_r(0, n) \leq \frac{-r + r(\mu_2/\mu_1) P(Z_n > 0)}{P(Z_n = n)} = -r.$$

To prove the implication in the opposite direction, assume that  $\overline{\lim} \Delta V_r(0, n) < 0$  for some  $r > 0$ . Then there exists an  $s \in (0, r)$  such that  $\Delta V_r(0, n) \leq -s$  for  $n$  sufficiently large. Applying (5) then gives

$$\frac{E(\lambda(Z_n)) - \mu_2(1 - s/\mu_1) P(Z_n > 0)}{P(Z_n = n)} \leq \Delta V_r(0, n) \leq -s.$$

This shows that

$$E(\lambda(Z_n)) - \mu_2 P(Z_n > 0) \leq -s \left( P(Z_n = n) + \frac{\mu_2}{\mu_1} P(Z_n > 0) \right) = -s$$

for all sufficiently large  $n$ , and, in light of (7), it follows that  $\overline{\lim} E(\lambda(Z_n)) < \min(\mu_1, \mu_2)$ .

**Lemma 2.** *Let  $f$  be a function of the form  $f(x) = u(x_1) + v(x_2)$ , for some functions  $u, v: \mathbb{Z}_+ \rightarrow \mathbb{R}$ . Then  $\lim_{x \rightarrow \infty} f(x) = \infty$  if and only if*

$$\lim_{x_1 \rightarrow \infty} u(x_1) = \infty \quad \text{and} \quad \lim_{x_2 \rightarrow \infty} v(x_2) = \infty.$$

*Proof.* Assume that  $\lim u(x_1) = \lim v(x_2) = \infty$ , and fix an  $M \in \mathbb{R}$ . Since  $u_0 := \inf u(x_1)$  and  $v_0 := \inf v(x_2)$  are finite, we can choose an  $m_1$  and an  $m_2$  such that  $u(x_1) > M - v_0$  for all  $x_1 > m_1$  and  $v(x_2) > M - u_0$  for all  $x_2 > m_2$ . Hence,  $f(x) > M$  if either  $x_1 > m_1$  or  $x_2 > m_2$ , implying that the set  $\{x: f(x) \leq M\} \subset [0, m_1] \times [0, m_2]$  is finite. Since  $M$  is arbitrary, it follows that  $\lim_{x \rightarrow \infty} f(x) = \infty$ .

Next suppose that  $\lim_{x \rightarrow \infty} f(x) = \infty$ . Then if  $\underline{\lim} u(x_1) < \infty$ , there exists a  $c \in \mathbb{R}$  such that the set  $S = \{x_1: u(x_1) \leq c\}$  is infinite. This implies that the set  $\{x: f(x) \leq c + v(0)\} \supset S \times \{0\}$  is infinite, in contradiction to the assumption that  $\lim_{x \rightarrow \infty} f(x) = \infty$ . Thus,  $\underline{\lim} u(x_1) = \infty$ . We can prove that  $\underline{\lim} v(x_2) = \infty$  similarly.

*Proof of Theorem 2.* Let  $r > 0$ , and assume that  $V_r \in \mathcal{V}$  is a Lyapunov function for  $q$ . Let  $S_0$  be a finite set such that (F3) holds. Then  $\{0\} \times (n_0, \infty) \subset S_0^c$  for some  $n_0$ , which implies that

$$\overline{\lim} \Delta V_r(0, n) \leq \sup_{n > n_0} \Delta V_r(0, n) \leq \sup_{x \in S_0^c} \Delta V_r(x) < 0.$$

By Lemma 1, this implies (6).

To prove the implication in the opposite direction, assume that (6) holds. Applying Lemma 1, we can pick an  $r > 0$  such that  $\overline{\lim} \Delta V_r(0, n) < 0$ . Hence, there exist an  $n_0$  and an  $\varepsilon > 0$  such that  $\Delta V_r(0, n) \leq -\varepsilon$  for all  $n > n_0$ . With  $S_0 = \{0\} \times [0, n_0]$ , it follows that

$$\sup_{x \in S_0^c} \Delta V_r(x) = \max \left( \sup_{n > n_0} \Delta V_r(0, n), \sup_{\{x : x_1 > 0\}} \Delta V_r(x) \right) \leq \max(-\varepsilon, -r) < 0,$$

since, by the construction of  $V_r$ ,  $\Delta V_r(x) = -r$  for all  $x$  with  $x_1 > 0$ . Thus,  $V_r$  satisfies (F3). Next observe that, using (4), we have

$$\lambda(n) - \mu_2(v_r(n) - v_r(n - 1)) = \Delta V_r(0, n) \leq -\varepsilon \quad \text{for } n > n_0.$$

This shows that  $v_r(n) - v_r(n - 1) \geq \varepsilon/\mu_2$  for sufficiently large  $n$ , implying that  $\lim_{n \rightarrow \infty} v_r(n) = \infty$ . By Lemma 2, we conclude that  $V_r$  satisfies (F2). Furthermore, (F1) holds trivially, since the set  $\{x : q(x, y) > 0\}$  is finite for all  $x$ . Thus,  $V_r$  is a Lyapunov function for  $q$ . Finally, note that  $X$  is irreducible when  $\lambda(0) > 0$ . An application of Theorem 1 thus completes the proof.

### 3. Necessary conditions for stability

Assume that  $\lambda(0) > 0$ , so that the system  $S(\lambda, \mu_1, \mu_2)$  is irreducible. In the previous section, we saw that the condition

$$\overline{\lim} E(\lambda(Z_n)) < \min(\mu_1, \mu_2), \quad Z_n \sim \text{geom}_n(\mu_1/\mu_2),$$

is sufficient for the stability of  $S(\lambda, \mu_1, \mu_2)$ . This section is devoted to studying whether the above condition is also *necessary* for stability.

#### 3.1. Small perturbations of Markov processes

In this section, we study how ergodicity is preserved under small perturbations of generators of Markov processes. Let  $q(x, y)$  and  $q'(x, y)$  be the generators of Markov processes on a countable state space  $S$ , and let

$$D(q, q') = \{x : q(x, y) \neq q'(x, y), \text{ for some } y\}$$

and

$$\bar{D}(q, q') = D(q, q') \cup \{y : q(x, y) > 0 \text{ or } q'(x, y) > 0, \text{ for some } x \in D(q, q')\}.$$

Furthermore, for a set  $F \subset S$ , let

$$T_F = \inf\{t > 0 : X(t-) \neq X(t), X(t) \in F\},$$

with the convention  $\inf \emptyset = \infty$ , and write  $T_x := T_{\{x\}}$  for  $x \in S$ .

**Lemma 3.** *Let  $X$  and  $X'$  be irreducible Markov processes on a countable state space  $S$ , respectively generated by  $q(x, y)$  and  $q'(x, y)$ , with  $q(x), q'(x) < \infty$  for all  $x$ . Assume that  $\bar{D}(q, q')$  is finite. Then  $X$  is ergodic if and only if  $X'$  is ergodic.*

*Proof.* By symmetry, it is sufficient to show that the ergodicity of  $X'$  implies that of  $X$ . Therefore, assume  $X'$  to be ergodic, and let  $x$  be a state in  $D = D(q, q')$ . Denote the first jump time of  $X$  by  $\tau = \inf\{t > 0: X(t-) \neq X(t)\}$ . By irreducibility  $E_x(\tau) < \infty$ ; thus, by the strong Markov property,

$$E_x(T_D) = E_x(\tau) + E_x(E_{X(\tau)}(T_D); X(\tau) \notin D),$$

where  $E(Z; F)$  is used as shorthand for  $E(Z \mathbf{1}_F)$ . Since  $q(x, y)$  and  $q'(x, y)$  coincide outside  $D$ , and  $P_x(X(\tau) \in \bar{D}) = 1$ , we have

$$E_x(E_{X(\tau)}(T_D); X(\tau) \notin D) = E_x(E_{X(\tau)}(T'_D); X(\tau) \in \bar{D} \setminus D) \leq \sup_{y \in \bar{D}} E_y(T'_D).$$

Since  $X'$  is ergodic and  $\bar{D}$  is finite, the right-hand side of the above inequality is also finite, and we conclude that  $E_x(T_D) < \infty$ . Because  $X$  is irreducible, this property implies that  $X$  is positive recurrent [7, Theorem 4.3(ii) and Theorem 4.4].

**3.2. Bottleneck at node 1**

Assume that  $\mu_1 < \mu_2$ . Intuition suggests that in this case the stability of the system depends on whether or not the buffer content at node 1 grows to infinity. Observe that, during the periods of time in which node 1 remains busy, the input to node 2 is a Poisson process with rate  $\mu_1$ . The approach here is to compare the original process to a saturated system in which node 2 also receives input at rate  $\mu_1$  during the time periods in which node 1 is empty, and show that the stability regions for the two systems are close to each other. With this goal in mind, let us introduce another model family, denoted by  $S^N(\lambda, \mu_1, \mu_2)$ . Fix a nonnegative integer  $N$  and, for  $x \neq y$ , define

$$q^N(x, y) = q(x, y) + \mu_1 \mathbf{1}(x_1 = 0, x_2 < N, y = x + e_2).$$

It is clear that, when  $\lambda(0) > 0$ , the transition rates  $q^N(x, y)$  define, using the minimal construction, an irreducible Markov process  $X^N$  on  $\mathbb{Z}_+^2 \cup \{\kappa\}$ . By Lemma 3, we know that the stability of  $S^N(\lambda, \mu_1, \mu_2)$  is equivalent to that of  $S(\lambda, \mu_1, \mu_2)$ . Furthermore, as we let  $N$  approach infinity,  $S^N(\lambda, \mu_1, \mu_2)$  comes to resemble a network in which node 2 receives stationary input at rate  $\mu_1$ .

**Lemma 4.** *Assume that  $S^N(\lambda, \mu_1, \mu_2)$  is stable. Then the stationary distribution of  $X^N$  satisfies*

$$E(\lambda(X_2^N)) = \mu_2 P(X_2^N > 0) - \mu_1 P(X_1^N = 0, X_2^N < N), \tag{8}$$

$$P(X_2^N = n) = \left(\frac{\mu_1}{\mu_2}\right)^n P(X_2^N = 0) - \mathbf{1}(n > N) \sum_{j=N}^{n-1} \left(\frac{\mu_1}{\mu_2}\right)^{n-j} P(X_1^N = 0, X_2^N = j), \tag{9}$$

and, for all real-valued functions  $f$  on  $\mathbb{Z}_+$ ,

$$E(f(X_2^N); X_2^N \leq N) = E(f(Z_N)) P(X_2^N \leq N), \quad Z_N \sim \text{geom}_N(\mu_1/\mu_2). \tag{10}$$

*Proof.* Starting from the balance equations for  $X^N$ , it is not hard to check that  $E(\lambda(X_2)) = \mu_1 P(X_1^N > 0)$  and

$$\mu_1 P(X_1^N > 0) + \mu_1 P(X_1^N = 0, X_2^N < N) = \mu_2 P(X_2^N > 0),$$

from which it follows that (8) is true. Furthermore, it is straightforward to verify that, for all  $n$ ,

$$P(X_2^N = n + 1) = \frac{\mu_1}{\mu_2} [P(X_2^N = n) - \mathbf{1}(n \geq N) P(X_1^N = 0, X_2^N = n)],$$

from which (9) and (10) follow.

**Theorem 3.** Assume that  $\mu_1 < \mu_2$ , and let  $Z \sim \text{geom}(\mu_1/\mu_2)$ . Then

$$\begin{aligned} E(\lambda(Z)) < \mu_1 &\Rightarrow S(\lambda, \mu_1, \mu_2) \text{ is stable,} \\ E(\lambda(Z)) > \mu_1 &\Rightarrow S(\lambda, \mu_1, \mu_2) \text{ is unstable.} \end{aligned}$$

*Proof.* Let  $Z_N \sim \text{geom}_N(\mu_1/\mu_2)$ . Because  $\mu_1 < \mu_2$ , it follows that  $\lim E(\lambda(Z_N)) = E(\lambda(Z))$ , and the first statement follows from Theorem 2. To prove the second claim, assume, on the contrary, that  $S(\lambda, \mu_1, \mu_2)$  is stable. Then, by Lemma 3, so is  $S^N(\lambda, \mu_1, \mu_2)$ , for each  $N$ . By applying (8) and (10), we see that

$$E(\lambda(Z_N)) P(X_2^N \leq N) = E(\lambda(X_2^N); X_2^N \leq N) \leq \mu_2 P(X_2^N > 0). \tag{11}$$

Next, (9) implies that

$$P(X_2^N > N) \leq \sum_{n>N} \left(\frac{\mu_1}{\mu_2}\right)^n \text{ for all } N,$$

meaning that  $\lim P(X_2^N \leq N) = 1$ . This observation, combined with (9), implies that

$$P(X_2^N = 0) = \frac{P(X_2^N \leq N)}{\sum_{n=0}^N (\mu_1/\mu_2)^n} \rightarrow 1 - \frac{\mu_1}{\mu_2} \text{ as } N \rightarrow \infty.$$

Hence,  $\lim P(X_2^N > 0) = \mu_1/\mu_2$ . Letting  $N \rightarrow \infty$  on both sides of (11) now shows that  $E(\lambda(Z)) \leq \mu_1$ , and the second claim follows by contraposition.

**3.3. Bottleneck at node 2**

In studying necessary stability conditions for the system when  $\mu_1 \geq \mu_2$ , the following asymptotic property of truncated geometric random variables will be useful.

**Lemma 5.** Let  $Z_n \sim \text{geom}_n(z)$ , with  $z \geq 1$ . Then, for all nonnegative functions  $f$  on  $\mathbb{Z}_+$ ,

$$\underline{\lim} f(n) \leq \underline{\lim} E(f(Z_n)) \leq \overline{\lim} E(f(Z_n)) \leq \overline{\lim} f(n).$$

*Proof.* Without loss of generality, assume that  $\overline{\lim} f(n) < \infty$ . Choose a number  $r$  such that  $\overline{\lim} f(n) < r$ . Then there exists an  $n_0$  such that  $f(n) \leq r$  for all  $n > n_0$ , and, thus,

$$E(f(Z_n)) \leq r + \frac{\sum_{j=0}^{n_0} (f(n) - r)z^j}{\sum_{j=0}^n z^j} \text{ for } n > n_0.$$

This implies that  $\overline{\lim} E(f(Z_n)) \leq r$ , so by letting  $r \downarrow \overline{\lim} f(n)$  it follows that  $\overline{\lim} E(f(Z_n)) \leq \overline{\lim} f(n)$ . The proof is completed by applying this inequality to  $-f$ .



**Theorem 4.** Assume that  $\mu_1 \geq \mu_2$ , and let  $Z_n \sim \text{geom}_n(\mu_1/\mu_2)$ . Then

$$\begin{aligned} \overline{\lim} E(\lambda(Z_n)) < \mu_2 &\Rightarrow S(\lambda, \mu_1, \mu_2) \text{ is stable,} \\ \underline{\lim} \lambda(n) > \mu_2 &\Rightarrow S(\lambda, \mu_1, \mu_2) \text{ is unstable.} \end{aligned}$$

In particular, if  $\lim \lambda(n)$  exists then

$$\begin{aligned} \lim \lambda(n) < \mu_2 &\Rightarrow S(\lambda, \mu_1, \mu_2) \text{ is stable,} \\ \lim \lambda(n) > \mu_2 &\Rightarrow S(\lambda, \mu_1, \mu_2) \text{ is unstable.} \end{aligned}$$

*Proof.* The first statement follows from Theorem 2. To prove the second claim, assume, on the contrary, that  $S(\lambda, \mu_1, \mu_2)$  is stable. Then, by Lemma 3, so is  $S^N(\lambda, \mu_1, \mu_2)$ , for each  $N$ . Choose an  $r \in \mathbb{R}$  such that  $r < \underline{\lim} \lambda(n)$ . It follows from Lemma 5 that  $\underline{\lim} \lambda(n) \leq \underline{\lim} E(\lambda(Z_N))$ . Thus,  $\lambda(N) \geq r$  and  $E(\lambda(Z_N)) \geq r$  for all sufficiently large  $N$ , and, for all such  $N$ ,

$$\begin{aligned} E(\lambda(X_2^N)) &= E(\lambda(X_2^N); X_2^N > N) + E(\lambda(Z_N)) P(X_2^N \leq N) \\ &\geq r P(X_2^N > N) + r P(X_2^N \leq N) \\ &= r, \end{aligned}$$

implying that  $\underline{\lim} E(\lambda(X_2^N)) \geq r$ . By letting  $r$  approach  $\underline{\lim} \lambda(n)$ , we see that  $\underline{\lim} E(\lambda(X_2^N)) \geq \underline{\lim} \lambda(n)$ . Next,  $\lim P(X_2^N > 0) = 1$ , because  $P(X_2^N = 0) \leq (\sum_{j=0}^N (\mu_1/\mu_2)^j)^{-1}$  by (9). Moreover, (8) shows that  $E(\lambda(X_2^N)) \leq \mu_2 P(X_2^N > 0)$  for all  $N$ , implying that

$$\underline{\lim} \lambda(n) \leq \underline{\lim} E(\lambda(X_2^N)) \leq \underline{\lim} \mu_2 P(X_2^N > 0) = \mu_2,$$

which proves the second claim, by contraposition. In the special case in which  $\lambda(n)$  has a limit when  $n$  tends to infinity, Lemma 5 shows that

$$\underline{\lim} \lambda(n) = \lim \lambda(n) = \overline{\lim} E(\lambda(Z_n)),$$

meaning that the last two implications of the theorem now follow from the first two.

There may be a substantial ‘gap’ between the necessary and sufficient stability conditions of Theorem 4 if  $\lambda(n)$  is divergent. To gain some insight into why characterization of the stability of the system is difficult for such  $\lambda$ , let us consider the behavior of  $S(\lambda, \mu_1, \mu_2)$  as  $\mu_1$  tends to infinity. Intuition suggests that in this case the system should resemble the single-server queue with service rate  $\mu_2$  and state-dependent input rate  $\lambda(n)$ , for which it is known [3, Corollary III.2.5, p. 74] that stability is equivalent to having

$$\sum_{n=0}^{\infty} \frac{\lambda(0) \cdots \lambda(n)}{\mu_2^{n+1}} < \infty. \tag{12}$$

Consider, for example, the input rates  $\lambda(n) = a$  for even  $n$  and  $\lambda(n) = b$  for odd  $n$ , where  $0 < a < b$ . Then (12) reduces to  $\sqrt{ab} < \mu_2$ , while, with  $Z_n \sim \text{geom}_n(\mu_1/\mu_2)$ ,

$$\underline{\lim} \lambda(n) = a < \frac{\mu_1 b + \mu_2 a}{\mu_1 + \mu_2} = \overline{\lim} E(\lambda(Z_n)).$$

Hence, the gap between the necessary and sufficient stability conditions in Theorem 4 grows according to

$$\left[ a, \frac{\mu_1 b + \mu_2 a}{\mu_1 + \mu_2} \right] \rightarrow [a, b], \quad \mu_1 \rightarrow \infty.$$

However, condition (12) may not in general be the correct asymptotic stability characteristic of  $S(\lambda, \mu_1, \mu_2)$  as  $\mu_1 \rightarrow \infty$ , due to a fundamental difference between the single-server queue and the tandem network. Namely, if  $\lambda(n) = 0$  for some  $n$ , then the single-server queue is stable because the queue size cannot exceed  $n$ . Obviously, this property is not true for  $S(\lambda, \mu_1, \mu_2)$ , and this is why the necessary and sufficient stability conditions for  $S(\lambda, \mu_1, \mu_2)$  must be more complex than (12).

### 3.4. Eventually vanishing input rate function

In most applications, it is natural to assume that  $\lambda(n)$  eventually becomes 0 for large  $n$ , and that the admission controller thus strictly blocks all incoming traffic when the number of jobs in node 2 becomes too large. In this case  $\lim \lambda(n) = 0$ , so Theorem 4 shows that, for  $\mu_1 \geq \mu_2$ ,  $S(\lambda, \mu_1, \mu_2)$  is stable regardless of the shape of the function  $\lambda$ . However, if node 1 is the bottleneck then Theorem 3 determines the stability of the system, except in the critical case when  $E(\lambda(Z)) = \mu_1$ . Intuition about birth–death processes suggests that the system is unstable also in this special case. The validity of this intuition will be proved next. The key to the proof is the following lemma, which shows that the stability of  $S(\lambda, \mu_1, \mu_2)$  implies the stability of the saturated system  $S^*(\lambda, \mu_1, \mu_2)$ , in which node 2 behaves as if node 1 were never empty.

**Lemma 6.** *Assume that  $\mu_1 < \mu_2$ , and that  $\lambda(n) = 0$  for sufficiently large  $n$ . If  $S(\lambda, \mu_1, \mu_2)$  is stable then so is the system  $S^*(\lambda, \mu_1, \mu_2)$  generated by the transition rates*

$$q^*(x, y) = q(x, y) + \mu_1 \mathbf{1}(x_1 = 0, y = x + e_2), \quad x \neq y.$$

*Proof.* Fix a  $K \in \mathbb{Z}_+$  such that  $\lambda(n) = 0$  for all  $n > K$ , and define the transition rates  $q'$  by

$$q'(x, y) = q(x, y) + \mu_1 \mathbf{1}(x_1 = 0, x_2 > K, y = x + e_2), \quad x \neq y.$$

Because  $q'(x) < \infty$  for all  $x$ , the rates  $q'(x, y)$  define an irreducible Markov process  $X'$  on  $\mathbb{Z}_+^2 \cup \{\kappa\}$ . The first step is to show that  $X'$  is ergodic. Note that the set of states over which  $q$  and  $q'$  differ is now given by  $D(q, q') = \{0\} \times [K + 1, \infty)$ . The key to the proof is to observe that the behavior of  $X'$  inside  $D \equiv D(q, q')$  is similar to a birth–death process with birth rate  $\mu_1$  and death rate  $\mu_2$ . Let  $x = (0, K + 1)$ . Then, since  $\mu_1 < \mu_2$ , it follows that, for all  $y \in D \setminus \{x\}$ ,

$$E_y(T'_x) = \frac{y_2 - x_2}{\mu_2 - \mu_1}. \tag{13}$$

The ergodicity of  $X$  implies that  $E_{x-e_2}(T'_D) = E_{x-e_2}(T_D) < \infty$ . Next, since

$$P_{x-e_2}(T'_D \leq T'_x) = 1,$$

we obtain

$$\begin{aligned} E_{x-e_2}(T'_x) &= E_{x-e_2}(T'_D) + E_{x-e_2}(E_{X'(T'_D)}(T'_x); X'(T'_D) \neq x) \\ &= E_{x-e_2}(T'_D) + E_{x-e_2}\left(\frac{X'_2(T'_D) - x_2}{\mu_2 - \mu_1}\right) \\ &= E_{x-e_2}(T_D) + E_{x-e_2}\left(\frac{X_2(T_D) - x_2}{\mu_2 - \mu_1}\right), \end{aligned} \tag{14}$$

using the strong Markov property and (13). Since  $E_y(T_x) = (y_2 - x_2)/\mu_2$  for all  $y \in D \setminus \{x\}$ , we similarly find that

$$E_{x-e_2}(T_x) = E_{x-e_2}(T_D) + E_{x-e_2}\left(\frac{X_2(T_D) - x_2}{\mu_2}\right). \tag{15}$$

Since  $X$  is ergodic, comparison of (14) and (15) shows that  $E_{x-e_2}(T'_x) < \infty$ . Conditioning on the first transition of  $X'$  now yields

$$\begin{aligned} E_x(T'_x) &= \frac{1}{\mu_1 + \mu_2} + \frac{\mu_1}{\mu_1 + \mu_2} E_{x+e_2}(T'_x) + \frac{\mu_2}{\mu_1 + \mu_2} E_{x-e_2}(T'_x) \\ &= \frac{1}{\mu_1 + \mu_2} + \frac{\mu_1}{\mu_1 + \mu_2} \frac{1}{\mu_2 - \mu_1} + \frac{\mu_2}{\mu_1 + \mu_2} E_{x-e_2}(T'_x), \end{aligned}$$

showing that  $E_x(T'_x) < \infty$ . By irreducibility, it now follows that  $X'$  is ergodic.

Finally, note that the set  $\bar{D}(q', q^*) \subset [0, 1] \times [0, K + 1]$  is finite. Thus, in light of Lemma 3, we may now conclude that the Markov process  $X^*$  generated by  $q^*(x, y)$  is ergodic.

**Theorem 5.** Assume that  $\lambda(n) = 0$  for sufficiently large  $n$ .

(i) If  $\mu_1 < \mu_2$  then  $S(\lambda, \mu_1, \mu_2)$  is stable if and only if  $E(\lambda(Z)) < \mu_1$ , with

$$Z \sim \text{geom}(\mu_1/\mu_2).$$

(ii) If  $\mu_1 \geq \mu_2$  then  $S(\lambda, \mu_1, \mu_2)$  is always stable.

*Proof.* In light of Theorems 3 and 4, all we need show is that the stability of  $S(\lambda, \mu_1, \mu_2)$  implies that  $E(\lambda(Z)) < \mu_1$  when  $\mu_1 < \mu_2$ . Therefore, assume that  $\mu_1 < \mu_2$  and that  $S(\lambda, \mu_1, \mu_2)$  is stable. By Lemma 6, so then is  $S^*(\lambda, \mu_1, \mu_2)$ . From the balance equations for  $X^*$ , it is easy to see that

$$X_2^* \sim \text{geom}(\mu_1/\mu_2).$$

Thus, the stationary mean rate of jobs arriving at node 1 equals  $E(\lambda(Z))$ , while the corresponding rate out is equal to  $\mu_1 P(X_1^* > 0)$ . Because these two quantities must be equal in a stable system, we conclude that

$$E(\lambda(Z)) = \mu_1 P(X_1^* > 0) < \mu_1,$$

where the last inequality is strict because  $P(X_1^* = 0) > 0$ , by the ergodicity of  $X^*$ .

#### 4. Sensitivity analysis of the stability region

In this section, we focus on the stability of the system subject to fluctuations in the system parameters. The treatment here is restricted to the case of input rates that eventually vanish, for which Theorem 5 completely characterizes the stable parameter region.

##### 4.1. Sensitivity with respect to varying service rates

The next proposition shows that, with nonincreasing input rates, the stability of the system is preserved under the acceleration of node 1.

**Proposition 1.** Assume that  $\lambda$  is nonincreasing, and that  $\lambda(n) = 0$  for sufficiently large  $n$ . Then, for all  $\mu'_1 \geq \mu_1$ ,

$$S(\lambda, \mu_1, \mu_2) \text{ is stable} \quad \Rightarrow \quad S(\lambda, \mu'_1, \mu_2) \text{ is stable.}$$

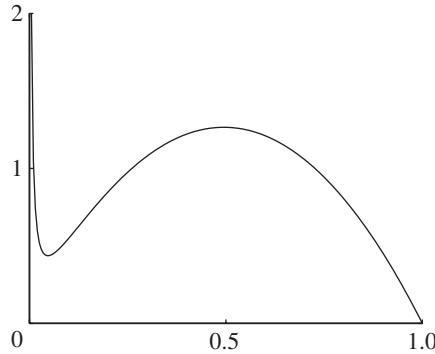


FIGURE 2: The left-hand side of (16), as a function of  $\mu_1$ .

*Proof.* Assume that  $S(\lambda, \mu_1, \mu_2)$  is stable and let  $\mu'_1 \geq \mu_1$ . If  $\mu'_1 \geq \mu_2$  then  $S(\lambda, \mu'_1, \mu_2)$  is stable, by Theorem 5. On the other hand, if  $\mu'_1 < \mu_2$  then  $\mu_1 < \mu_2$ , and the necessary condition of Theorem 5 shows that  $f(\mu_1/\mu_2) < \mu_1$ , where  $f(x) = (1 - x) \sum_{n=0}^{\infty} \lambda(n)x^n$ . Because the sequence  $\lambda(n)$  is bounded and nonnegative,  $f$  is differentiable in  $(0, 1)$ , with

$$f'(x) = \sum_{n=0}^{\infty} (n + 1)(\lambda(n + 1) - \lambda(n))x^n \leq 0,$$

implying that  $f(\mu'_1/\mu_2) \leq f(\mu_1/\mu_2)$ . It follows that  $f(\mu'_1/\mu_2) < \mu'_1$ , which guarantees the stability of  $S(\lambda, \mu'_1, \mu_2)$ , by Theorem 5.

To see why it is necessary to require  $\lambda$  to be nonincreasing, consider the following example.

**Example 1.** Let  $\mu_2 = 1$  and assume that  $\lambda(n) = 0$  for  $n \geq 3$ . Then  $S(\lambda, \mu_1, \mu_2)$  is stable for all  $\mu_1 \geq 1$  and, for  $\mu_1 \in (0, 1)$ , the stability of  $S(\lambda, \mu_1, \mu_2)$  is equivalent to having

$$\mu_1^{-1}(1 - \mu_1)(\lambda(0) + \lambda(1)\mu_1 + \lambda(2)\mu_1^2) < 1. \tag{16}$$

Figure 2 shows the left-hand side of (16) as a function of  $\mu_1$ , where  $\lambda(0) = \lambda(1) = \frac{1}{100}$  and  $\lambda(2) = 5$ . The plot shows that increasing the service rate  $\mu_1$  from  $\frac{1}{5}$  to  $\frac{1}{2}$  destabilizes the system.

Alternatively, we may fix  $\mu_1$  and see what happens when  $\mu_2$  varies. The following proposition contains a rather surprising result: even with a nonincreasing  $\lambda$ , acceleration of one of the servers may indeed destabilize the system. The physical intuition behind Proposition 2 is that, when  $\mu_2$  is very large, the admission controller finds node 2 empty most of the time. This means that the input rate of the system is close to  $\lambda(0)$ .

**Proposition 2.** Assume that  $\lambda$  is nonincreasing and that  $\lambda(n) = 0$  for sufficiently large  $n$ , and fix  $\mu_1 > 0$ .

- (i) For  $\lambda(0) \leq \mu_1$ ,  $S(\lambda, \mu_1, \mu_2)$  is stable for all  $\mu_2 > 0$ .
- (ii) For  $\lambda(0) > \mu_1$ ,  $S(\lambda, \mu_1, \mu_2)$  becomes unstable for sufficiently large  $\mu_2$ .

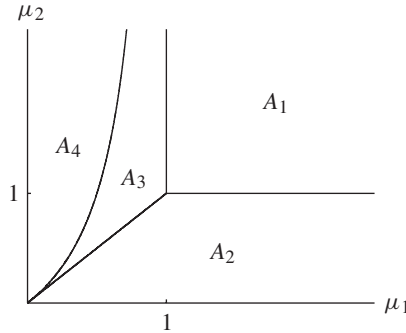


FIGURE 3: Phase diagram of the system with threshold-based admission control.

*Proof.* First observe that, by Theorem 5,  $S(\lambda, \mu_1, \mu_2)$  is stable for all small  $\mu_2$ ,  $\mu_2 \leq \mu_1$ . To study the case with  $\mu_2 > \mu_1$ , fix a number  $n_0$  such that  $\lambda(n) = 0$  for all  $n > n_0$ . Then, with  $Z \sim \text{geom}(\mu_1/\mu_2)$ ,

$$E(\lambda(Z)) = \left(1 - \frac{\mu_1}{\mu_2}\right) \sum_{n=0}^{n_0} \lambda(n) \left(\frac{\mu_1}{\mu_2}\right)^n. \tag{17}$$

If  $\lambda(0) \leq \mu_1$  then (17) implies that, for all  $\mu_2 > \mu_1$ ,

$$E(\lambda(Z)) \leq \lambda(0)(1 - (\mu_1/\mu_2)^{n_0+1}) < \mu_1,$$

which, by Theorem 5, is sufficient for stability. Moreover, the right-hand side of (17) converges to  $\lambda(0)$  as  $\mu_2 \rightarrow \infty$ . From this we can conclude that if  $\lambda(0) > \mu_1$ , then  $E(\lambda(Z)) > \mu_1$  for sufficiently large values of  $\mu_2$ . By Theorem 5,  $S(\lambda, \mu_1, \mu_2)$  is unstable for such values.

**4.2. Phase partition for threshold-based admission control**

Consider the network with threshold-based admission control, and assume, without loss of generality, that jobs arrive at the network at unit rate. Denoting the threshold level by  $K$ , this system can be modeled as  $S(\lambda, \mu_1, \mu_2)$  with  $\lambda(n) = \mathbf{1}(n \leq K)$ . Theorem 5 now implies that, for each  $K \in \mathbb{Z}_+ \cup \{\infty\}$ , the set of pairs  $(\mu_1, \mu_2)$  for which the system is stable is

$$R_K = \{(\mu_1, \mu_2) : 1 - (\mu_1/\mu_2)^{K+1} < \min(\mu_1, \mu_2)\}.$$

Since  $R_K \supset R_{K+1}$  for all  $K$ , the stabilizable region is given by  $\bigcup_{K \leq \infty} R_K = R_0$ , where  $R_\infty = \{(\mu_1, \mu_2) : \min(\mu_1, \mu_2) > 1\}$  represents the system with no overload. The positive orthant of  $\mathbb{R}^2$  can now be partitioned into four phases, as follows:

- $A_1 = R_\infty$  represents the region in which the uncontrolled system is stable;
- $A_2 = \bigcap_{K < \infty} R_K$  is the region in which any control stabilizes the overloaded system;
- $A_3 = R_0 \setminus \bigcap_{K < \infty} R_K$  is the region in which the overloaded system is stabilizable using sufficiently strict admission control;
- $A_4 = R_0^c$  is the region in which the system cannot be stabilized.

This partition is depicted in Figure 3. The phase diagram clearly illustrates the results of Propositions 1 and 2, showing that the acceleration of server 1 drives the system towards more stable regions, while the acceleration of server 2 may destabilize the network.

## 5. Conclusion

In this paper, we have considered the problem of characterizing the stability region of a two-node queueing network with feedback admission control. For eventually vanishing input rates, the characterization was shown to be complete. It was also illustrated how the presence of feedback signaling removes some typical monotonicity properties of queueing networks, by showing that increasing service rates may destabilize the network.

For a diverging input rate function and a bottleneck at node 2, the exact characterization of the stability region remains an open problem. Other possible directions for future research include generalizing the results for nonexponential service and interarrival times, and considering queueing networks with more than two nodes.

## Acknowledgements

This work was funded by the Academy of Finland Teletronics II / FIT project and the Finnish Graduate School in Stochastics. The author would like to thank Ilkka Norros for his valuable help and advice during the project, Jarmo Malinen for many inspiring discussions, and the anonymous referee for helpful comments on improving the presentation of the results.

## References

- [1] ADAN, I. J. B. F., WESSELS, J. AND ZIJM, W. H. M. (1993). Compensation approach for two-dimensional Markov processes. *Adv. Appl. Prob.* **25**, 783–817.
- [2] ALTMAN, E., AVRACHENKOV, K. E. AND NÚÑEZ-QUEJIA, R. (2004). Perturbation analysis for denumerable Markov chains with application to queueing models. *Adv. Appl. Prob.* **36**, 839–853.
- [3] ASMUSSEN, S. (2003). *Applied Probability and Queues*, 2nd edn. Springer, New York.
- [4] BAMBOS, N. AND WALRAND, J. (1989). On stability of state-dependent queues and acyclic queueing networks. *Adv. Appl. Prob.* **21**, 681–701.
- [5] FOSTER, F. G. (1953). On the stochastic matrices associated with certain queueing processes. *Ann. Math. Statist.* **24**, 355–360.
- [6] LESKELÄ, L. AND RESING, J. A. C. (2004). A tandem queueing network with feedback admission control. Res. Rep. 09, 2004/2005, Institut Mittag-Leffler.
- [7] MEYN, S. P. AND TWEEDIE, R. L. (1993). Stability of Markovian processes. II. Continuous-time processes and sampled chains. *Adv. Appl. Prob.* **25**, 487–517.
- [8] NEUTS, M. F. (1981). *Matrix-Geometric Solutions in Stochastic Models*. Johns Hopkins University Press, Baltimore, MD.
- [9] TWEEDIE, R. L. (1975). Sufficient conditions for regularity, recurrence and ergodicity of Markov processes. *Math. Proc. Camb. Philos. Soc.* **78**, 125–136.