A470

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

Lasse Leskelä



TEKNILLINEN KORKEAKOULU TEKNISKA HÖGSKOLAN HELSINKI UNIVERSITY OF TECHNOLOGY TECHNISCHE UNIVERSITÄT HELSINKI UNIVERSITE DE TECHNOLOGIE D'HELSINKI

A470

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

Lasse Leskelä

Helsinki University of Technology Department of Engineering Physics and Mathematics Institute of Mathematics Lasse Leskelä: Stabilization of an overloaded queueing network using measurementbased admission control; Helsinki University of Technology Institute of Mathematics Research Reports A470 (2004).

Abstract: Admission control is often employed to avoid congestion in queueing networks subject to overload. In distributed networks the admission decisions must usually be based on imperfect measurements on the network state. We will study how this lack of complete state information affects the system performance by considering a simple network model for distributed admission control. In this paper we will characterize the stability region of the network and show how the presence of feedback makes the stability of the system very sensitive to its parameters.

AMS subject classifications: 60K25, 68M20, 90B15, 90B22

Keywords: queueing network; admission control; stability; overload; perturbed Markov process

lasse.leskela@iki.fi

ISBN 951-22-7039-0 ISSN 0784-3143 Espoo, 31 March 2004

Helsinki University of Technology Department of Engineering Physics and Mathematics Institute of Mathematics P.O. Box 1100, 02015 HUT, Finland email:math@hut.fi http://www.math.hut.fi/

1 Introduction

Consider an overloaded queueing network where the mean offered traffic exceeds the service capacity over a long time period. In this case, it is often necessary to employ admission control to avoid the network to become fully congested. Many networks of practical interest are composed of subnetworks, not all of which are administered 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 will study the effect of imperfect information to the performance of the admission control scheme. Typical performance measures for well-dimensioned networks in this kind of setting include the average amount of rejected traffic per unit time, and the mean proportion of time the network load is undesirably high. However, assuming the network under study is subjected to long-term overload, there is another performance criterion that must first be analyzed, namely: *If the network is subjected to a stationary load exceeding the service capacity, how strict admission control rules should we set in order to stabilize the system*?

To deal with the question mathematically, we will assume that the network can be modeled using the simplest non-trivial model for a distributed network, the two-node tandem network with independent and exponential service times and unlimited buffers. We will denote the network state as $X = (X_1, X_2)$, where X_i is the number of jobs in node *i*. We assume that the admission control can be modeled so that the input to the system is a Poisson process with a stochastic time-varying intensity, the intensity $\lambda = \lambda(X)$ being a function of the network state. The lack of complete state information will be reflected in the model by assuming that the input rate λ is a function of only one of the X_i . If we assume $\lambda(X) = \lambda(X_1)$, then the analysis of the system can be reduced to the study of birth-death processes, which is well understood. This is why in the following we will always assume $\lambda(X) = \lambda(X_2)$, so that the admission control introduces a feedback loop to the system. For example, with $\lambda(X) = 1(X_2 \leq K)$ for some $K \in \mathbb{Z}_+$, we can model a network where traffic arriving at unit rate is rejected when the size of the second buffer exceeds a threshold level K, see Figure 1. However, in order to also cover more complex admission policies with multiple thresholds and thinning of input traffic, in the following we are not going to restrict the shape of $\lambda(X_2)$ in any way.

More precisely, X will be defined as a continuous-time stochastic process as follows. Let λ be a non-negative function on \mathbb{Z}_+ and $\mu_1, \mu_2 > 0$. We define the the transition rates q(x, y) for $x, y \in \mathbb{Z}_+^2$ by

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

PSfrag replacements



Figure 1: A simple network with threshold-based admission control.

where e_i denotes the *i*:th unit vector of \mathbb{Z}^2_+ . We adopt the convention that q(x, x) = 0 always, and denote the transition rate out of x by

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

It is clear that $q(x) < \infty$ for all x, so using the minimal construction (cf. eg. [3, 5]) the rates q(x, y) define a unique Markov process X on $\mathbb{Z}^2_+ \cup \{\kappa\}$. Here κ denotes an additional state not in \mathbb{Z}^2_+ , with $T_{\kappa} = \inf\{t > 0 : X(t) = \kappa\} \in [0, \infty]$ being the time of explosion of X. We will use the notation $S(\lambda, \mu_1, \mu_2)$ for the set of transition rates corresponding to the the triple (λ, μ_1, μ_2) , and say that the system $S(\lambda, \mu_1, \mu_2)$ is *stable* if the corresponding Markov process is ergodic, i.e., irreducible and positive recurrent.

In its most general form, our stability problem may now be stated as

(P1) 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, we can without loss of generality assume that the offered traffic arrives at unit rate. Denoting the admission threshold by K, (P1) now takes the form

(P2) For each $(\mu_1, \mu_2) \in \mathbb{R}^2_+$, determine for which values of $K \in \mathbb{Z}_+ \cup \infty$, if any, the system $S(1(\cdot \leq K), \mu_1, \mu_2)$ is stable.

Note that the system corresponding to $K = \infty$ in (P2) is the ordinary tandem queue, for which it is well-known that $\min(\mu_1, \mu_2) > 1$ is sufficient and necessary for stability. On the other hand, assuming overload, answering the question on 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 include the requirement that at least one of the queues be finite. In this case the two-dimensional nature of the problem can partly be reduced back to one-dimensional by applying matrix-geometric methods [7]. For networks with unlimited buffers and state-dependent service times, [4] provides stability results extending to non-Markovian systems, however ruling out networks with the type of feedback loop present here. Concerning the network $S(\lambda, \mu_1, \mu_2)$ defined above, the compensation approach introduced in [1] can be used for computing the invariant measure in the special case where λ is constant on $\{n \in \mathbb{Z}_+ : n \geq 1\}$. Further, there is a recent article [2], introducing perturbation techniques that seem appropriate for asymptotically analyzing the behavior of $S(\lambda, \mu_1, \mu_2)$ under suitable parameter scaling.

In this paper we will partially answer (P1) by deriving sufficient and necessary conditions for stability. Furthermore, by showing that in the special case of threshold-based admission control the sufficient and necessary conditions coincide, we will provide a complete solution of (P2). In addition, we will analyze the sensitivity of the system with respect to changes in the service rates and show how acceleration of one of the servers may, rather paradoxically, unstabilize the system.

2 A sufficient condition for stability

Let E be a countable set. For a function $V: E \to \mathbb{R}$, we will denote

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

if the set $\{x : V(x) \leq M\}$ is finite for all $M \in \mathbb{R}$. Further, if $q(x, y) \geq 0$ for $x, y \in E$, then the mean drift of V with respect to q is denoted by

$$\Delta V(x) = \sum_{y} [V(y) - V(x)] q(x, y),$$

assuming the sum on the right-hand side is well-defined.

Definition 1. Let $q(x, y) \ge 0$ for $x, y \in E$. A map $V : E \to \mathbb{R}$ is called a *Lyapunov function* for q if it satisfies the following conditions called *Foster's criteria*:

- (F1) $\lim_{x\to\infty} V(x) = \infty$.
- (F2) $\Delta V(x) < \infty$ for all x.
- (F3) There is a finite set $E_0 \subset E$ such that $\sup_{x \in E \setminus E_0} \Delta V(x) < 0$.

The following continuous-time analogue of Foster's classical theorem [6], providing a sufficient condition for stability, is found in [8].

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

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

$$\Delta V(x) = -r \quad \text{for all } x \text{ with } x_1 > 0. \tag{2}$$

It is straightforward to verify that (2) is equivalent to

$$v(1) = 1 - (\lambda(0) + r)/\mu_1,$$

$$v(n+1) = 1 - (\lambda(n) + r)/\mu_1 + (1 + \mu_2/\mu_1)v(n) - \mu_2/\mu_1v(n-1), \quad n \ge 1.$$

Denoting $\alpha(n) = 1 - (\lambda(n) + r)/\mu_1$ 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 $n \ge 1$, with $w(0) = \alpha(0)$. Thus, $w(n) = \sum_{k=0}^{n} \alpha(k) (\mu_1/\mu_2)^{k-n}$, so 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 (2) defines for each r > 0 the function

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

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 there are hopes that V_r might satisfy (F3) for a suitably chosen finite subset of \mathbb{Z}^2_+ . In order to investigate whether this is the case, we will study the mean drift of V_r for x = (0, n)with $n \ge 1$,

$$\Delta V_r(0,n) = \lambda(n) - \mu_2(v_r(n) - v_r(n-1)).$$
(3)

For $z \ge 0$, we will denote $Z_n \sim \text{geom}_n(z)$ if Z_n is a random variable on $\mathbb{Z} \cap [0, n]$ with $\mathbb{P}(Z_n = j) = cz^j$ for $0 \le j \le n$. Using this notation, one may verify that the above expression can be alternatively written as

$$\Delta V_r(0,n) = \frac{\mathbb{E}\lambda(Z_n) - \mu_2(1 - r/\mu_1) \mathbb{P}(Z_n > 0)}{\mathbb{P}(Z_n = n)}, \quad Z_n \sim \text{geom}_n(\mu_1/\mu_2).$$
(4)

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

$$\overline{\lim} \mathbb{E}\lambda(Z_n) < \min(\mu_1, \mu_2), \quad Z_n \sim \operatorname{geom}_n(\mu_1/\mu_2).$$
(5)

In particular, if $\lambda(0) > 0$, then (5) is sufficient for the stability of $S(\lambda, \mu_1, \mu_2)$.

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

Lemma 1. Condition (5) is equivalent to $\overline{\lim} \Delta V_r(0, n) < 0$ for some r > 0. *Proof.* Let $Z_n \sim \operatorname{geom}_n(\mu_1/\mu_2)$ for $n \ge 0$. Observe first that since $\lim \mathbb{P}(Z_n > 0) = \min(1, \mu_1/\mu_2)$,

$$\overline{\lim} \mathbb{E}\lambda(Z_n) - \min(\mu_1, \mu_2) = \overline{\lim} \{\mathbb{E}\lambda(Z_n) - \mu_2 \mathbb{P}(Z_n > 0)\}.$$
 (6)

Assume now that (5) holds. Then we can choose an r > 0 so that $\mathbb{E}\lambda(Z_n) - \mu_2 \mathbb{P}(Z_n > 0) \leq -r$ for n large enough. It follows that $\overline{\lim} \Delta V_r(0, n) < 0$, since using (4) we see that eventually for large n,

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

For the other direction, assume $\overline{\lim} \Delta V_r(0,n) < 0$ for some r > 0. Then there is an $s \in (0,r)$ so that for n large enough, $\Delta V_r(0,n) \leq -s$, and applying (4),

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

This shows that

 $\mathbb{E}\lambda(Z_n) - \mu_2 \mathbb{P}(Z_n > 0) \le -s(\mathbb{P}(Z_n = n) + \mu_2/\mu_1 \mathbb{P}(Z_n > 0)) = -s$

for all *n* large enough, and in light of (6) it follows that $\overline{\lim} \mathbb{E}\lambda(Z_n) < \min(\mu_1, \mu_2)$.

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

Proof. Assume $\lim u(x_1) = \lim v(x_2) = \infty$, 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 m_1 and 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 $x_2 > m_2$, so that the set $\{x : f(x) \le M\} \subset [0, m_1] \times [0, m_2]$ is finite. Since M was arbitrary, $\lim_{x \to \infty} f(x) = \infty$.

Suppose next that $\lim_{x\to\infty} f(x) = \infty$. Then if $\underline{\lim} u(x_1) < \infty$, there is a $c \in \mathbb{R}$ so that $S = \{x_1 : u(x_1) \leq c\}$ is infinite. This implies that $\{x : f(x) \leq c+v(0)\} \supset S \times \{0\}$ is infinite, contrary to the assumption $\lim_{x\to\infty} f(x) = \infty$. Thus, $\underline{\lim} u(x_1) = \infty$. Similarly, one proves that $\underline{\lim} v(x_2) = \infty$.

Proof of Theorem 2. Let r > 0 and assume $V_r \in \mathcal{V}$ is a Lyapunov function for q. Let E_0 be a finite set so that (F3) holds. Then $\{0\} \times (n_0, \infty) \subset E_0^c$ for some n_0 , which implies

$$\overline{\lim}\,\Delta V_r(0,n) \le \sup_{n > n_0} \Delta V_r(0,n) \le \sup_{x \in E_0^c} \Delta V_r(x) < 0.$$

By Lemma 1, this implies (5).

For the other direction, assume (5) holds with $Z_n \sim \text{geom}_n(\mu_1/\mu_2)$. Applying Lemma 1, we can pick an r > 0 so that $\overline{\lim} \Delta V_r(0, n) < 0$. Hence, there is an n_0 and an $\epsilon > 0$ so that $\Delta V_r(0, n) \leq -\epsilon$ for all $n > n_0$. Denoting $E_0 = \{0\} \times [0, n_0]$, it follows that

$$\sup_{x \in E_0^c} \Delta V_r(x) = \max\{\sup_{n > n_0} \Delta V_r(0, n), \sup_{x: x_1 > 0} \Delta V_r(x)\} \le \max\{-\epsilon, -r\} < 0,$$

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 (3),

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

This shows that $v_r(n) - v_r(n-1) \ge \epsilon/\mu_2$ eventually for large n, so that $\lim_{n\to\infty} v_r(n) = \infty$. By Lemma 2, we conclude that V_r satisfies (F1). Further, (F2) 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 if $\lambda(0) > 0$. Hence, application of Theorem 1 now completes the proof.

3 Necessary conditions for stability

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

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

is sufficient for the stability of $S(\lambda, \mu_1, \mu_2)$. We would next like to study whether this condition is also necessary.

3.1 Small perturbations of Markov processes

In this section we will study how ergodicity is preserved under small perturbations of generators of Markov processes. If q(x, y) and q'(x, y) are generators of Markov processes on a countable state space E, we will denote

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

and

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

Further, for $F \subset E$ we will denote

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

with the convention $\inf \emptyset = \infty$; and $T_x = T_{\{x\}}$ for $x \in E$.

Proposition 1. Let X and X' be irreducible Markov processes on a countable state space E generated by q(x, y) and q'(x, y), respectively, with q(x), $q'(x) < \infty$ for all x. Assume that $\overline{D}(q, q')$ is finite. Then X is ergodic iff X' is ergodic.

The proof will utilize the following lemma, which is a continuous-time analogue of Lemma I.3.9 in [3].

Lemma 3. Let X be as in Proposition 1. Assume there is a finite $F \subset E$ such that $\mathbb{E}_x T_F < \infty$ for all $x \in F$. Then X is ergodic.

Proof. Fix $x \in F$ and define the stopping times T_F^n by

$$T_F^0 = 0, \quad T_F^{n+1} = \inf\{t > T_F^n : X(t-) \neq X(t), \ X(t) \in F\} \text{ for } n \ge 0.$$

Note that if $\mathbb{E}_x T_F^n < \infty$ for some *n*, then by the strong Markov property,

$$\mathbb{E}_x T_F^{n+1} = \mathbb{E}_x T_F^n + \mathbb{E}_x \mathbb{E}_{X(T_F^n)} T_F \le \mathbb{E}_x T_F^n + \sup_{y \in F} \mathbb{E}_y T_F < \infty,$$

so by induction it follows that $\mathbb{E}_x T_F^n < \infty$ for all $n \in \mathbb{Z}_+$. In particular, T_F^n is almost surely finite for all n, so we may define $Y(n) = X(T_F^n)$ for $n \in \mathbb{Z}_+$. Using the strong Markov property it is not hard to check that Y is a Markov

chain on F. Further, it is clear that $\mathbb{P}_x(\exists n > 0 : Y(n) = y) = \mathbb{P}_x(\exists t > 0 : X(t) = y)$ for all $x, y \in F$, so that the irreducibility of X implies that of Y. Let $S_x = \inf\{n > 0 : Y(n) = x\}$. Being an irreducible Markov chain on the finite set F, Y is positive recurrent, so that $\mathbb{E}_x S_x < \infty$. Now,

$$\mathbb{E}_x T_x = \mathbb{E}_x \sum_{n=1}^{S_x} (T_F^n - T_F^{n-1}) = \sum_{n=1}^{\infty} \mathbb{E}_x (T_F^n - T_F^{n-1}; S_x \ge n).$$

Since $\{S_x \ge n\} \in \mathcal{F}_{T_F^{n-1}}$, the strong Markov property implies that the terms of the sum on the right-hand side equal

$$\mathbb{E}_{x}(\mathbb{E}_{x}(T_{F}^{n}-T_{F}^{n-1} \mid \mathcal{F}_{T_{F}^{n-1}}); \ S_{x} \ge n) = \mathbb{E}_{x}(\mathbb{E}_{X(T_{F}^{n-1})}T_{F}; \ S_{x} \ge n).$$

By irreducibility, the ergodicity of X now follows from

$$\mathbb{E}_x T_x \le \sup_{y \in F} \mathbb{E}_y T_F \sum_{n=1}^{\infty} \mathbb{P}_x (S_x \ge n) = \sup_{y \in F} \mathbb{E}_y T_F \mathbb{E}_x S_x < \infty.$$

Proof of Proposition 1. By symmetry, it is sufficient to show that the ergodicity of X' implies that of X. So, assume X' is 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, $\mathbb{E}_x \tau < \infty$, so by the strong Markov property,

$$\mathbb{E}_x T_D = \mathbb{E}_x \tau + \mathbb{E}_x (\mathbb{E}_{X(\tau)} T_D; X(\tau) \notin D)$$

Since q(x, y) and q'(x, y) coincide outside D, and $\mathbb{P}_x(X(\tau) \in \overline{D}) = 1$,

$$\mathbb{E}_x(\mathbb{E}_{X(\tau)}T_D; X(\tau) \notin D) = \mathbb{E}_x(\mathbb{E}_{X(\tau)}T'_D; X(\tau) \in \overline{D} \setminus D) \le \sup_{y \in \overline{D}} \mathbb{E}_yT'_D.$$

Since X' is ergodic and \overline{D} is finite, so is the right-hand side in the above inequality, and we conclude $\mathbb{E}_x T_D < \infty$. Application of Lemma 3 now completes the proof.

3.2 General input rate function

Let us now introduce another model family, denoted by $S^N(\lambda, \mu_1, \mu_2)$. For $N \in \mathbb{Z}_+$, define

$$q^{N}(x,y) = q(x,y) + \mu_{1} 1(x_{1} = 0, x_{2} < N, y = x + e_{2}).$$

It is clear that when $\lambda(0) > 0$, the transition rates $q^N(x, y)$ define using the minimal construction an irreducible Markov process on $\mathbb{Z}^2_+ \cup \{\kappa\}$.

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

$$\mathbb{E}\lambda(X_2^N) = \mu_2 \mathbb{P}(X_2^N > 0) - \mu_1 \mathbb{P}(X_1^N = 0, X_2^N < N),$$
(7)

$$\mathbb{P}(X_2^N = n) \le \mathbb{P}(X_2^N = 0) \left(\mu_1/\mu_2\right)^n \quad \text{for all } n \in \mathbb{Z}_+,\tag{8}$$

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

$$\mathbb{E}(f(X_2^N) \mid X_2^N \le N) = \mathbb{E}f(Z_N), \quad Z_N \sim \operatorname{geom}_N(\mu_1/\mu_2).$$
(9)

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

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

showing that (7) is true. Further, it is straightforward to verify that for all n,

$$\mathbb{P}(X_2^N = n+1) = \mu_1/\mu_2 \left[\mathbb{P}(X_2^N = n) - 1(n \ge N) \mathbb{P}(X_1^N = 0, X_2^N = n) \right]$$

from which (8) and (9) follow.

For $0 \leq z < 1$, we will denote $Z \sim \text{geom}(z)$ if Z is a random variable on \mathbb{Z}_+ with $\mathbb{P}(Z_n = j) = (1 - z)z^j$ for all $j \in \mathbb{Z}_+$.

Theorem 3. When $\mu_1 < \mu_2$, a necessary condition for the stability of $S(\lambda, \mu_1, \mu_2)$ is given by

$$\mathbb{E}\lambda(Z) \le \mu_1, \quad Z \sim \operatorname{geom}(\mu_1/\mu_2).$$

Proof. Let $\mu_1 < \mu_2$ and assume $S(\lambda, \mu_1, \mu_2)$ is stable. Then by Proposition 1, so is $S^N(\lambda, \mu_1, \mu_2)$ for each N. Next, applying (9) and (7),

$$\mu_{1}\mathbb{P}(X_{1}^{N} = 0, X_{2}^{N} < N) = \mu_{2}\mathbb{P}(X_{2}^{N} > 0) - \mathbb{E}\lambda(X_{2}^{N})$$

$$\leq \mu_{2}\mathbb{P}(X_{2}^{N} > 0) - \mathbb{E}(\lambda(X_{2}^{N}); X_{2}^{N} \leq N)$$

$$= \mu_{2}\mathbb{P}(X_{2}^{N} > 0) - \mathbb{P}(X_{2}^{N} \leq N)\mathbb{E}\lambda(Z_{N}),$$
(10)

where $Z_N \sim \text{geom}_N(\mu_1/\mu_2)$. Inequality (8) implies

$$\mathbb{P}(X_2^N > N) \le \sum_{n > N} (\mu_1/\mu_2)^n \quad \text{for all } N,$$

so that $\lim \mathbb{P}(X_2^N \leq N) = 1$. Next, setting f(n) = 1(n = 0) in (9) gives

$$\mathbb{P}(X_2^N = 0) = \mathbb{P}(X_2^N \le N) \mathbb{P}(Z_N = 0),$$

showing that $\lim \mathbb{P}(X_2^N > 0) = \mu_1/\mu_2$. Further, since $\mu_1 < \mu_2$, it follows that $\lim \mathbb{E}\lambda(Z_N) = \mathbb{E}\lambda(Z)$ with $Z \sim \operatorname{geom}(\mu_1/\mu_2)$. Combining these observations with (10), we may now conclude that

$$0 \le \lim[\mu_2 \mathbb{P}(X_2^N > 0) - \mathbb{P}(X_2^N \le N) \mathbb{E}\lambda(Z_N)] = \mu_1 - \mathbb{E}\lambda(Z).$$

Lemma 5. Let $Z_n \sim \text{geom}_n(z)$ with $z \ge 1$. Then for any $f : \mathbb{Z}_+ \to \mathbb{R}$,

$$\underline{\lim} f(n) \le \underline{\lim} \mathbb{E} f(Z_n) \le \lim \mathbb{E} f(Z_n) \le \lim f(n).$$

Proof. Without loss of generality, assume $\overline{\lim} f(n) < \infty$. Choose an $r \in \mathbb{R}$ so that $\lim f(n) < r$. Then there is an n_0 so that $f(n) \leq r$ for all $n > n_0$, and thus

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

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

Theorem 4. For $\mu_1 \geq \mu_2$, a necessary condition for the stability of $S(\lambda, \mu_1, \mu_2)$ is given by $\underline{\lim} \lambda(n) \leq \mu_2$.

Proof. If $S(\lambda, \mu_1, \mu_2)$ is stable with $\mu_1 \geq \mu_2$, then by Proposition 1, so is $S^{N}(\lambda, \mu_{1}, \mu_{2})$ for each N. Choose an $r \in \mathbb{R}$ so that $r < \underline{\lim} \lambda(n)$. For $Z_n \sim \operatorname{geom}_n(\mu_1/\mu_2)$ it follows by Lemma 5 that $\underline{\lim} \lambda(n) \leq \underline{\lim} \mathbb{E}\lambda(Z_n)$. Thus there is an n_0 so that $\lambda(N) \ge r$ and $\mathbb{E}\lambda(Z_N) \ge r$ for all $N > n_0$. Thus,

$$\mathbb{E}\lambda(X_2^N) = \mathbb{E}(\lambda(X_2^N); X_2^N > N) + \mathbb{E}\lambda(Z_N)\mathbb{P}(X_2^N \le N)$$
$$\geq r\mathbb{P}(X_2^N > N) + r\mathbb{P}(X_2^N \le N) = r$$

for all $N > n_0$, so that $\underline{\lim} \mathbb{E}\lambda(X_2^N) \ge r$, and by taking $r \uparrow \underline{\lim} \lambda(n)$ we

conclude that $\underline{\lim} \mathbb{E}\lambda(X_2^N) \geq \underline{\lim}\lambda(n)$. Next, $\lim \mathbb{P}(X_2^N > 0) = 1$, since $\mathbb{P}(X_2^N = 0) \leq (\sum_{j=0}^N (\mu_1/\mu_2)^j)^{-1}$ by inequality (8). Equality (7) shows that $\mathbb{E}\lambda(X_2^N) \leq \mu_2 \mathbb{P}(X_2^N > 0)$ for all N, so that we can now conclude

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

Example 1 (Diverging input rate function). Assume that $\lambda(n) = a$ for n even and b for n odd, for some a, b with $0 < a < \mu_2 < b < \mu_1$. Then the necessary condition of Theorem 4 is valid, while with $Z_n \sim \text{geom}_n(\mu_1/\mu_2)$, the sufficient condition of Theorem 2 takes the form

$$\overline{\lim} \mathbb{E}\lambda(n) = \mu_2/(\mu_1 + \mu_2)a + \mu_1/(\mu_1 + \mu_2)b < \mu_2.$$

When $\mu_1 \to \infty$, the above inequality stops being valid, so that for large μ_1 , we cannot determine the stability of the system.

The next corollary shows that, fortunately, the gap between the sufficient and necessary conditions observed in Example 1 shrinks to a single point, when we restrict ourselves to converging input rate functions.

Corollary 1. Assume $\lambda(0) > 0$ and that $\lim \lambda(n)$ exists. Then with $Z_n \sim$ $\operatorname{geom}_n(\mu_1/\mu_2)$, $\lim \mathbb{E}\lambda(Z_n)$ exists, and

$$\lim \mathbb{E}\lambda(Z_n) < \min(\mu_1, \mu_2) \implies S(\lambda, \mu_1, \mu_2) \text{ is stable,} \\ \lim \mathbb{E}\lambda(Z_n) > \min(\mu_1, \mu_2) \implies S(\lambda, \mu_1, \mu_2) \text{ is unstable.}$$

Proof. First, it is easy see that for $\mu_1 < \mu_2$, $\lim \mathbb{E}\lambda(Z_n) = \mathbb{E}\lambda(Z)$ with $Z \sim \text{geom}(\mu_1/\mu_2)$, while by Lemma 5, $\lim \mathbb{E}\lambda(Z_n) = \lim \lambda(n)$ for $\mu_1 \ge \mu_2$. The first implication is the content of Theorem 2, while for $\mu_1 < \mu_2$, the second implication follows from Theorem 3. If $\mu_1 \ge \mu_2$, then $\lim \mathbb{E}\lambda(Z_n) = \lim \lambda(n)$ combined with Theorem 4 shows that the second implication is valid. \Box

3.3 Eventually vanishing input rate function

If the input rate function eventually vanishes, we are going to show that the necessary and sufficient conditions derived in earlier sections will coincide, providing an exact characterization of the stability region for the system.

Proposition 2. Assume $\mu_1 < \mu_2$, and $\lambda(n) = 0$ eventually for 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

$$q^*(x,y) = q(x,y) + \mu_1 1(x_1 = 0, y = x + e_2).$$

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

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

Clearly, $q'(x) < \infty$ for all x, and thus the rates q'(x, y) define a Markov process X' on $\mathbb{Z}^2_+ \cup \{\kappa\}$, which is obviously irreducible. We will first show that X' is ergodic. Note that set of states where 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 inside D = D(q, q'), the behavior of X' is similar to a birth-death process with birth and death rates μ_1 and μ_2 , respectively. Denote x = (0, K + 1). Then since $\mu_1 < \mu_2$, it follows that for all $y \in D \setminus \{x\}$,

$$\mathbb{E}_y T'_x = \frac{y_2 - x_2}{\mu_2 - \mu_1}.$$
(11)

The ergodicity of X implies $\mathbb{E}_{x-e_2}T'_D = \mathbb{E}_{x-e_2}T_D < \infty$. Next, since $\mathbb{P}_{x-e_2}(T'_D \leq T'_x) = 1$, we can compute using the strong Markov property and (11),

$$\mathbb{E}_{x-e_2} T'_x = \mathbb{E}_{x-e_2} T'_D + \mathbb{E}_{x-e_2} (\mathbb{E}_{X'(T'_D)} T'_x; X'(T'_D) \neq x)$$

$$= \mathbb{E}_{x-e_2} T'_D + \mathbb{E}_{x-e_2} \frac{X'_2(T'_D) - x_2}{\mu_2 - \mu_1}$$

$$= \mathbb{E}_{x-e_2} T_D + \mathbb{E}_{x-e_2} \frac{X_2(T_D) - x_2}{\mu_2 - \mu_1}.$$
 (12)

Since $\mathbb{E}_y T_x = (y_2 - x_2)/\mu_2$ for all $y \in D \setminus \{x\}$, we find in a similar way that

$$\mathbb{E}_{x-e_2} T_x = \mathbb{E}_{x-e_2} T_D + \mathbb{E}_{x-e_2} \frac{X_2(T_D) - x_2}{\mu_2}.$$
 (13)

Since X is ergodic, comparison of (12) and (13) shows that $\mathbb{E}_{x-e_2}T'_x < \infty$. Conditioning on the first transition of X' now yields

$$\mathbb{E}_{x}T'_{x} = \frac{1}{\mu_{1} + \mu_{2}} + \frac{\mu_{1}}{\mu_{1} + \mu_{2}} \mathbb{E}_{x + e_{2}}T'_{x} + \frac{\mu_{2}}{\mu_{1} + \mu_{2}} \mathbb{E}_{x - e_{2}}T'_{x}$$
$$= \frac{1}{\mu_{1} + \mu_{2}} + \frac{\mu_{1}}{\mu_{1} + \mu_{2}} \frac{1}{\mu_{2} - \mu_{1}} + \frac{\mu_{2}}{\mu_{1} + \mu_{2}} \mathbb{E}_{x - e_{2}}T'_{x},$$

showing that $\mathbb{E}_x T'_x < \infty$. By irreducibility, it now follows that X' is ergodic.

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

Theorem 5. Assume $\lambda(0) > 0$ and $\lambda(n) = 0$ eventually for large n. Then (5) is necessary and sufficient for the stability of $S(\lambda, \mu_1, \mu_2)$. In particular, $S(\lambda, \mu_1, \mu_2)$ is always stable with $\mu_1 \ge \mu_2$, while for $\mu_1 < \mu_2$, the stability of the system is equivalent to $\mathbb{E}\lambda(Z) < \mu_1$ with $Z \sim \text{geom}(\mu_1/\mu_2)$.

Proof. That (5) is sufficient follows from Theorem 2. To prove the necessity, assume first that $\mu_1 < \mu_2$ and that $S(\lambda, \mu_1, \mu_2)$ is stable. Using Proposition 2, we conclude that $S^*(\lambda, \mu_1, \mu_2)$ is stable. From the balance equations for X^* it is easy to see that $X_2^* \sim \text{geom}(\mu_1/\mu_2)$, and for each $m \ge 0$,

$$\sum_{n=0}^{\infty} \lambda(n) \mathbb{P}(X_1^* = m, X_2^* = n) = \mu_1 \mathbb{P}(X_1^* = m + 1).$$

Summing this over m yields

$$\mathbb{E}\lambda(Z) = \mathbb{E}\lambda(X_2^*) = \mu_1 \mathbb{P}(X_1^* > 0) < \mu_1,$$

where the last inequality is strict since $\mathbb{P}(X_1^* > 0) < 1$ by the ergodicity of X^* . Now (5) follows from the fact that with $\mu_1 < \mu_2$ and $Z_n \sim \text{geom}_n(\mu_1/\mu_2)$, $\lim \mathbb{E}\lambda(Z_n) = \mathbb{E}\lambda(Z)$.

On the other hand, Lemma 5 shows that for $\mu_1 \ge \mu_2$, $\overline{\lim} \mathbb{E}\lambda(Z_n) \le \overline{\lim} \lambda(n) = 0$, so that (5) always holds.

4 Sensitivity analysis of the stability region

In this section we will investigate the stability of the system subjected to fluctuations in the system parameters. We will restrict ourselves to the case with eventually vanishing input rates, since under this assumption Theorem 5 completely characterizes the stable parameter region.

4.1 Properties of the system with eventually vanishing input rates

Lemma 6. Assume λ is non-increasing and $\lambda(n) = 0$ eventually for large n. Then the function $F : \mathbb{R}_+ \to [0, \infty]$ defined by

$$z \mapsto \lim \mathbb{E}\lambda(Z_n), \quad Z_n \sim \operatorname{geom}_n(z)$$

is non-increasing.

Proof. Assume that λ is as above. Then it is easy to check that F(z) = 0 for all $z \ge 1$, while for $0 \le z < 1$, $F(z) = (1-z) \sum_{n=0}^{\infty} \lambda(n) z^n$. To complete the proof, it suffices to note that since $\lambda(n)$ are bounded, F(z) differentiable for all $0 \le z < 1$ with

$$F'(z) = \sum_{n=0}^{\infty} (n+1) [\lambda(n+1) - \lambda(n)] z^n \le 0.$$

The next proposition shows that with non-increasing input rates, the stability of the system is preserved under speeding up of server 1.

Proposition 3. Assume λ is non-increasing and $\lambda(n) = 0$ eventually for large n. Then for all $\mu'_1 \ge \mu_1$,

$$S(\lambda, \mu_1, \mu_2)$$
 is stable $\implies S(\lambda, \mu'_1, \mu_2)$ is stable.

Proof. We know by Theorem 5 that the stability of $S(\lambda, \mu_1, \mu_2)$ is equivalent to (5), which can now be expressed as $F(\mu_1/\mu_2) < \min(\mu_1, \mu_2)$. If $\mu'_1 \ge \mu_1$, then by Lemma 6,

$$F(\mu_1'/\mu_2) \le F(\mu_1/\mu_2) < \min(\mu_1,\mu_2) \le \min(\mu_1',\mu_2),$$

showing that the triple (λ, μ'_1, μ_2) also satisfies (5).

To see why it is necessary to require λ to be non-increasing, consider the following example.

Example 2. Let $\lambda(n) = 1/8 \, 1(n=0) + (1+1/8) \, 1(n>1)$, and $\mu_2 = 1$. For $Z_n \sim \text{geom}_n(\mu_1/\mu_2)$, we now have

$$\overline{\lim} \mathbb{E}\lambda(Z_n) - \min(\mu_1, \mu_2) = \begin{cases} 1/8 - (1+1/8)\mu_1(1-\mu_1), & \mu_1 < 1, \\ 1/8, & \mu_1 \ge 1. \end{cases}$$

Denoting the above quantity by $\phi(\mu_1)$, we know by Corollary 1 that $S(\lambda, \mu_1, \mu_2)$ is stable for $\phi(\mu_1) < 0$ and unstable for $\phi(\mu_1) > 0$. The function $\phi(\mu_1)$ is plotted in Figure 2. Observe that $\mu_1 = 1/2$ is a stable point, while $\mu_1 \to 0$ or $\mu_1 \to \infty$ leads to unstability.

Alternatively, we may fix μ_1 and see what happens when μ_2 varies. The following proposition tells us a rather surprising result: even with non-increasing λ , acceleration of one of the servers may indeed *unstabilize* the system.



Figure 2: The function $\phi(\mu_1)$ of Example 2.

Proposition 4. Assume λ is non-increasing and $\lambda(n) = 0$ eventually for large n and fix a $\mu_1 > 0$. Then

- for $\lambda(0) \leq \mu_1$, $S(\lambda, \mu_1, \mu_2)$ is stable for all $\mu_2 > 0$,
- for $\lambda(0) > \mu_1$, $S(\lambda, \mu_1, \mu_2)$ becomes eventually unstable for large μ_2 .

Proof. Fix a positive integer K so that $\lambda(n) = 0$ for all n > K. Assume first $\lambda(0) \leq \mu_1$. Then by Theorem 5 we know that $S(\lambda, \mu_1, \mu_2)$ is stable for all $\mu_2 \leq \mu_1$. On the other hand, for $\mu_2 > \mu_1$ with $Z \sim \text{geom}(\mu_1/\mu_2)$,

$$\mathbb{E}\lambda(Z) = (1 - \mu_1/\mu_2) \sum_{n=0}^{K} \lambda(n)(\mu_1/\mu_2)^n \le \lambda(0)(1 - (\mu_1/\mu_2)^{K+1}) < \mu_1.$$

Thus (5) holds for all $\mu_2 > \mu_1$, proving the first claim.

For the second part, assume that $\lambda(0) > \mu_1$. Then again $S(\lambda, \mu_1, \mu_2)$ is stable for all $\mu_2 \leq \mu_1$. But with $\mu_2 \to \infty$, the above equation shows that $\mathbb{E}\lambda(Z)$ eventually becomes larger than μ_1 , making $S(\lambda, \mu_1, \mu_2)$ unstable. \Box

4.2 Phase partition of the system with threshold-based admission control

Consider the network with threshold-based admission control, and assume without loss of generality that the offered traffic to the network arrives at unit rate. Denoting the threshold level by K, this system can be modeled as $S(1(\cdot \leq K), \mu_1, \mu_2)$. Theorem 5 now implies that for each $K \in \mathbb{Z}_+ \cup \infty$, the set of (μ_1, μ_2) corresponding to a stable system is given by

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

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

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



Figure 3: The phase diagram of the system with threshold-based admission control.

This partition is depicted in Figure 3. The phase diagram clearly illustrates the content of Propositions 3 and 4, showing that accelerating server 1 drives the system towards more stable regions, while rather paradoxically, speeding up server 2 may unstabilize the network.

References

- 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-QUEIJA, R. (2004). Perturbation analysis for denumerable markov chains with application to queueing models. *Adv. Appl. Prob.* **36**, To appear.
- [3] ASMUSSEN, S. (1987). Applied Probability and Queues. Wiley.
- [4] BAMBOS, N. AND WALRAND, J. (1989). On stability of state-dependent queues and acyclic queueing networks. *Adv. Appl. Prob.* **21**, 618–701.
- [5] BRÉMAUD, P. (1999). Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues. Springer-Verlag.

- [6] FOSTER, F. G. (1953). On the stochastic matrices associated with certain queuing processes. Ann. Math. Statist. 24, 355–360.
- [7] NEUTS, M. F. (1981). *Matrix-Geometric Solutions in Stochastic Models*. John Hopkins University Press.
- [8] TWEEDIE, R. L. (1975). Sufficient conditions for regularity, recurrence and ergodicity of markov processes. *Math. Proc. Camb. Phil. Soc.* 78, 125–136.

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, and Jarmo Malinen for many helpful and inspiring discussions.

(continued from the back cover)

- A459 Timo Salin The quenching problem for the N-dimensional ball April 2003
- A458 Tuomas Hytönen Translation-invariant Operators on Spaces of Vector-valued Functions April 2003
- A457 Timo Salin On a Refined Asymptotic Analysis for the Quenching Problem March 2003
- A456 Ville Havu , Harri Hakula , Tomi Tuominen A benchmark study of elliptic and hyperbolic shells of revolution January 2003
- A455 Yaroslav V. Kurylev , Matti Lassas , Erkki Somersalo Maxwell's Equations with Scalar Impedance: Direct and Inverse Problems January 2003
- A454 Timo Eirola , Marko Huhtanen , Jan von Pfaler Solution methods for R-linear problems in C^n October 2002
- A453 Marko Huhtanen Aspects of nonnormality for iterative methods September 2002
- A452 Kalle Mikkola Infinite-Dimensional Linear Systems, Optimal Control and Algebraic Riccati Equations October 2002
- A451 Marko Huhtanen Combining normality with the FFT techniques September 2002

HELSINKI UNIVERSITY OF TECHNOLOGY INSTITUTE OF MATHEMATICS RESEARCH REPORTS

The list of reports is continued inside. Electronical versions of the reports are available at *http://www.math.hut.fi/reports/*.

- A464 Ville Turunen Sampling at equiangular grids on the 2-sphere and estimates for sobolev space interpolation November 2003
- A463 Marko Huhtanen , Jan von Pfaler The real linear eigenvalue problem in C^n November 2003
- A462 Ville Turunen Pseudodifferential calculus on the 2-sphere October 2003
- A461 Tuomas Hytönen Vector-valued wavelets and the Hardy space $H^1(\mathbb{R}^n;X)$ April 2003
- A460 Timo Eirola , Jan von Pfaler Numerical Taylor expansions for invariant manifolds April 2003

ISBN 951-22-7039-0 ISSN 0784-3143 Espoo, 31 March 2004