

Assessing an intuitive condition for stability under a range of traffic conditions via a generalised Lu-Kumar network

forthcoming in *Journal of Applied Probability* vol. 37, no. 3, 2000

José Niño-Mora*

Kevin D. Glazebrook**

January 1997, Revised: November 1999, January 2000

*José Niño-Mora, Dept. of Economics and Business, Universitat Pompeu Fabra, E-08005 Barcelona, Spain. E-mail: jose.nino-mora@econ.upf.es. Part of this work was carried out during this author's stay at the Center for Operations Research and Econometrics (CORE) of the Université catholique de Louvain, Belgium. His research was supported by a CORE Research Fellowship, by European Commission Individual Marie Curie Postdoctoral Fellowship no. ERBFMBICT961480, and by Spain's National R & D Program Grant CICYT TAP98-0229.

**Kevin D. Glazebrook, Department of Statistics, Newcastle University, Newcastle upon Tyne NE1 7RU, UK, E-mail: kevin.glazebrook@newcastle.ac.uk. The research of this author was supported by the Engineering and Physical Sciences Research Council by means of grant nos. GR/K03043 and GR/M09308.

Abstract

We argue the importance *both* of developing simple sufficient conditions for the stability of general multiclass queueing networks *and* also of assessing such conditions under a range of assumptions on the weight of the traffic flowing between service stations. To achieve the former, we review a peak-rate stability condition and extend its range of application and for the latter, we introduce a generalisation of the Lu-Kumar network on which the stability condition may be tested for a range of traffic configurations. The peak-rate condition is close to exact when the between-station traffic is light, but degrades as this traffic increases.

Key words: Multiclass queueing networks, stability, fluid model, Lu-Kumar network.

Journal of Economic Literature classification: C60, C61.

1 Introduction

The stability problem for open multiclass queueing networks may be stated as follows: given a family of scheduling policies for such a model, characterize the parameter region under which any policy in the family is *stable* (i.e., the time-average number of customers in the network is finite). The problem has drawn extensive research attention since it was shown that the stability condition prevalent in product-form networks (i.e., that the nominal traffic intensity at each station is less than unity) does not guarantee stability more generally (see, e.g. Bramson (1994)).

Researchers have addressed the problem by developing methods to construct a Lyapunov function with negative drift (quadratic or, more recently, piecewise linear), for the network or its fluid model, that implies stability. See Dai (1995). Two kinds of results have emerged: (1) computational tests, which seek to construct such a Lyapunov function for specific model parameters, typically by solving a linear programming (LP) problem (see e.g., Kumar and Meyn (1996)), and (2) qualitative results, which establish the stability of a family of policies for a restricted network topology under some stability condition (see e.g., Dai and Weiss (1996)).

The current paper is motivated by two important considerations. Firstly, there has been a dearth in the literature of (necessary and) sufficient conditions for stability which (i) apply to reasonably general families of stochastic networks, (ii) are intuitive, (iii) can be easily checked and (iv) are well suited for addressing subsequent performance evaluation and optimization issues. Glazebrook and Niño-Mora (1999) present one such - namely, a sufficient condition for stability in open Markovian networks with single-server stations which is based around the simple notion that stations should have the capacity to process traffic arriving at a maximum possible (or peak) rate from elsewhere in the network. This so-called peak-rate stability condition is described in Section 2 and a proof of its sufficiency is given under general distributional assumptions, thus extending the previous result for Markovian networks.

Secondly, the stability literature has been preoccupied with highly connected networks, most especially reentrant lines. This is understandable, in that stability issues are at their most acute and intractable in such cases. However, in many applications there will be light traffic between stations (including those in which most customers are served at a single station, while a minority have requirements which involve an excursion within the network) and it is important not to neglect contributions to stability which are designed to deal with such situations. With this in mind, we introduce a class of simple networks in Section 3 in which

the weight of traffic between stations may be varied by the setting of a single parameter q . The networks generalise the two station network investigated by Lu and Kumar (1991), which is recovered as the case $q = 1$. A sufficient condition for stability of these networks is obtained, which is tight enough for practical purposes. This result allows us to use these networks to test proposed sufficient conditions for stability across a range of assumptions about the weight of between-station traffic.

In particular, in Section 4, we use these networks to assess the peak-rate stability condition of Section 2. We propose and evaluate a measure, $r(q)$, of the tightness of the peak-rate condition which depends on the level of between-station traffic, as measured by parameter q . In fact we have that $r(q) \rightarrow 1$, $q \rightarrow 0$ and so the condition is close to exact and will be a viable option for checking stability simply when traffic between stations is light. The condition is seen to degrade as q increases in that r is a decreasing function.

2 A peak-rate stability condition for a class of general networks with single-server stations

Glazebrook and Niño-Mora (1999) studied stability and performance issues for a general open Markovian queueing network with N customer classes and M single-server stations. Here we shall consider a larger class of networks in which interarrival and service times are drawn from general distributions. Station $m \in \mathcal{M} = \{1, \dots, M\}$ provides service to a *constituency* of customer classes $\mathcal{C}_m \subseteq \mathcal{N} = \{1, \dots, N\}$, where $\mathcal{C}_1, \dots, \mathcal{C}_M$ is a partition of class set \mathcal{N} . Class i customers (or *i -customers*) arrive exogenously at the network as a renewal process with rate α_i (equivalently, mean interarrival time $1/\alpha_i$ when $\alpha_i > 0$) and require an amount of service time at station $s(i) \in \mathcal{M}$ which has positive mean $m_i = 1/\mu_i$, $i \in \mathcal{N}$. We require that arrival streams are *either* null *or* have finite mean interarrival times and that all mean service times are finite. Further, interarrival times are “unbounded and spread out” in the sense of (1.4) and (1.5) of Dai (1995). Upon completion of its service, an i -customer may be *routed* for further service as a j -customer, with probability p_{ij} , or it may *leave* the network, with probability $p_{i0} = 1 - \sum_{j \in \mathcal{N}} p_{ij}$. To ensure that a customer entering the network leaves it with probability one we require that matrix $\mathbf{I} - \mathbf{P}$ is invertible, where \mathbf{I} denotes the identity matrix and $\mathbf{P} = (p_{ij})_{i,j \in \mathcal{N}}$ is the network *routing matrix*. We further assume that all arrival, service and routing processes are mutually independent. Note that the generalised Lu-Kumar network discussed in Section 3 belongs to this general class.

The *total arrival rate of j -customers*, denoted by λ_j , is given by the solution to the system of *traffic equations*

$$\lambda_j = \alpha_j + \sum_{i \in \mathcal{N}} \lambda_i p_{ij}, \quad \text{for } j \in \mathcal{N}.$$

The *nominal traffic intensity of j -customers*, denoted by ρ_j , is given by

$$\rho_j = \lambda_j / \mu_j, \quad \text{for } j \in \mathcal{N}.$$

We define similarly the *nominal traffic intensity at station m* , denoted by $\rho(m)$, as

$$\rho(m) = \sum_{j \in \mathcal{C}_m} \rho_j, \quad \text{for } m \in \mathcal{M}.$$

It is well known that the condition

$$\rho(m) < 1, \quad \text{for } m \in \mathcal{M}, \tag{1}$$

is *necessary* for the network to be stable, and hence we shall assume in what follows that it holds.

The network evolution is governed by a *scheduling policy*, which specifies dynamically how servers are allocated to available customers. We consider policies that are *stationary* (decisions depend on the current system *state*); *non-idling* (servers cannot idle when they have work to do); and *preemptive* (service of a customer may be interrupted, and resumed later).

To formulate our peak-rate stability condition we introduce additional network parameters, derived from the model primitives. For each class $i \in \mathcal{N}$ and class subset $S \subseteq \mathcal{N}$, we shall consider the *mean S -workload of an i -customer*, denoted by V_i^S , to be the mean remaining service time a current i -customer receives until it first leaves classes in S following completion of its current service. The V_i^S 's are determined as the unique solution of the linear system

$$V_i^S = m_i + \sum_{j \in S} p_{ij} V_j^S, \quad \text{for } i \in \mathcal{N}. \tag{2}$$

The *peak traffic intensity from class i into station m* , denoted $R(i, m)$, is the maximum rate at which work brought in by current i -customers can enter that station i.e.,

$$R(i, m) = \mu_i \sum_{j \in \mathcal{C}_m} p_{ij} V_j^{\mathcal{C}_m}, \quad \text{for } i \in \mathcal{N}, m \in \mathcal{M}. \tag{3}$$

The *peak traffic intensity from station m' into station m* , denoted $\bar{R}(m', m)$, is the maximum rate at which work can be transferred from station m' into m , i.e.,

$$\bar{R}(m', m) = \max_{i \in \mathcal{C}_{m'}} R(i, m), \text{ for } m, m' \in \mathcal{M}, \text{ with } m' \neq m. \quad (4)$$

Finally, we denote the *peak traffic intensity for station m* by $\bar{\rho}(m)$, defined as the maximum rate at which work can be transferred into that station, i.e.,

$$\bar{\rho}(m) = \sum_{j \in \mathcal{C}_m} \alpha_j V_j^{\mathcal{C}_m} + \sum_{m' \in \mathcal{M} \setminus \{m\}} \bar{R}(m', m), \text{ for } m \in \mathcal{M}. \quad (5)$$

The following result extends Theorem 2 in Glazebrook and Niño-Mora (1999).

Theorem 1

When (1) holds, the above queueing network is stable under all non-idling stationary policies when

$$\bar{\rho}(m) < 1, \quad m \in \mathcal{M}. \quad (6)$$

Comments

1. We would argue that the condition (6) in Theorem 1 has the properties (i)-(iv) outlined in the third paragraph of the Introduction.
2. It is easily shown that if $\rho(m) < 1$ then $\rho(m) \leq \bar{\rho}(m)$, and so the condition $\bar{\rho}(m) < 1$ is at least as strong as $\rho(m) < 1$. Since also $\bar{\rho}(m) - \rho(m) \rightarrow 0$ as $\max_{i \in \mathcal{N} \setminus \mathcal{C}_m, j \in \mathcal{C}_m} p_{ij} \rightarrow 0$ and moreover $\rho(m) < 1$ is necessary for stability, it follows that condition (6) is close to sharp when the flow between stations is light.
3. Glazebrook and Niño-Mora (1999) establish the result in Theorem 1 in the context of Markovian networks for which all class specific arrival streams are Poisson and all service times are exponential. They utilise the results of Kumar and Meyn (1996) who establish the boundedness of the so-called non-idling performance LP as a sufficient condition for stability and further obtain a closed form upper bound on the mean number of customers in the system. We shall prove our more general result by

means of a fluid model approach.

In brief, a non-idling fluid model of the above network is a solution $\mathbf{L}(t) = \{L_j(t), j \in \mathcal{N}\}$ and $\mathbf{T}(t) = \{T_j(t), j \in \mathcal{N}\}$ to the following equations for $t \geq 0$:

$$\mathbf{L}(t) = \mathbf{L}(0) + \boldsymbol{\alpha}t - \mathbf{T}(t)\text{Diag}(\boldsymbol{\mu})(\mathbf{I} - \mathbf{P})$$

$$\mathbf{L}(t) \geq 0$$

$$\mathbf{T}(0) = 0 \text{ and } \mathbf{T}(t) \text{ is nondecreasing componentwise} \tag{7}$$

$$B_m(t) = \sum_{i \in \mathcal{C}_m} T_i(t), \quad m \in \mathcal{M}$$

$$U_m(t) = t - B_m(t) \text{ is nondecreasing, but increases only at times } t \text{ for which } \sum_{i \in \mathcal{C}_m} L_i(t) = 0, \quad m \in \mathcal{M}.$$

Note that, in (7) $\text{Diag}(\boldsymbol{\mu})$ is diagonal matrix whose entries are the μ_i 's, while $\mathbf{L}(t)$, $\mathbf{T}(t)$ and $\boldsymbol{\alpha}$ are all taken to be row vectors. The interpretation of the quantities in (7) is clear. In particular, $L_i(t)$ (the i th component of $\mathbf{L}(t)$) is the amount of class i fluid present in the network at time t ; $B_m(t)$ is the cumulative amount of processing performed at station m by t , with $U_m(t)$ the corresponding cumulative amount of idle time, $m \in \mathcal{M}$. This fluid model is said to be *stable* if there exists at time $\delta > 0$ such that any solution of (7) having $\sum_{j \in \mathcal{N}} L_j(0) = 1$ satisfies

$$L_j(t) = 0, \text{ for } t \geq \delta, j \in \mathcal{N}.$$

Our proof of Theorem 1 is along the lines of Dai and Weiss (1996). We shall construct a Lyapunov function $G(t) = \max_{m \in \mathcal{M}} G_m(t)$ which is piecewise linear in $\mathbf{L}(t)$ and which has negative drift. Dai and Weiss (1996) show that this implies the stability of the fluid model under any non-idling policy. By Dai (1995) this implies the stability of the original stochastic network under any non-idling policy, or global stability as it is sometimes called. Lemma 1 expresses what we need. We firstly write

$$W_m(t) = \sum_{i \in \mathcal{C}_m} m_i L_i(t), \quad m \in \mathcal{M} \tag{8}$$

for the workload at station m in the fluid model at time $t \geq 0$.

Lemma 1

Let $\{G_m(t), m \in \mathcal{M}\}$ be a collection of nonnegative linear functions of $\mathbf{L}(t)$ such that

$$(a) \quad W_m(t) > 0 \Rightarrow \dot{G}_m(t) \leq -\epsilon_m \text{ for some } \epsilon_m > 0, m \in \mathcal{M};$$

$$(b) \quad W_m(t) = 0 \Rightarrow G_m(t) \leq G_{m'}(t), m' \in \mathcal{M}, m' \neq m.$$

It then follows that

(i) $G(t) = \max_{m \in \mathcal{M}} G_m(t)$ is an absolutely continuous nonnegative function such that

$$G(t) > 0 \Rightarrow \dot{G}(t) \leq -(\min_{m \in \mathcal{M}} \epsilon_m)$$

whenever t is a regular point of G ;

(ii) The fluid model is stable under all non-idling policies.

(iii) The original stochastic network is stable under all non-idling policies.

We now utilise Lemma 1 to prove Theorem 1 above.

Proof of Theorem 1

We introduce the Lyapunov function $G(t) = \max_{m \in \mathcal{M}} G_m(t)$, where

$$G_m(t) = \sum_{i \in \mathcal{C}_m} V_i^{\mathcal{C}_m} L_i(t), m \in \mathcal{M}.$$

For each class subset $S \subseteq \mathcal{N}$, let $\mathbf{L}_S(t) = \{L_j(t)\}_{j \in S}$, $\mathbf{P}_{SS} = \{p_{ij}\}_{i \in S, j \in S}$ and define $\boldsymbol{\alpha}_S$, $\mathbf{T}_S(t)$, $\text{Diag}_S(\boldsymbol{\mu})$,

\mathbf{I}_S and $\mathbf{P}_{S^c S^c}$ analogously, where $S^c = \mathcal{N} \setminus S$. It follows from (7) that

$$\mathbf{L}_S(t) = \mathbf{L}_S(0) + \boldsymbol{\alpha}_S t - \mathbf{T}_S(t) \text{Diag}_S(\boldsymbol{\mu})(\mathbf{I}_S - \mathbf{P}_{SS}) + \mathbf{T}_{S^c}(t) \text{Diag}_{S^c}(\boldsymbol{\mu}) \mathbf{P}_{S^c S}. \quad (9)$$

Postmultiplying (9) by column vector $\mathbf{V}_S^S = \{V_i^S\}_{i \in S}$, and simplifying by use of (2), (3) and (7) yields, in the special case $S = \mathcal{C}_m$, the identities

$$\begin{aligned} G_m(t) &= G_m(0) + t \sum_{j \in \mathcal{C}_m} \alpha_j V_j^{\mathcal{C}_m} - \sum_{j \in \mathcal{C}_m} T_j(t) + \sum_{i \in \mathcal{N} \setminus \mathcal{C}_m} T_i(t) \mu_i \sum_{j \in \mathcal{C}_m} p_{ij} V_j^{\mathcal{C}_m} \\ &= G_m(0) + t \sum_{j \in \mathcal{C}_m} \alpha_j V_j^{\mathcal{C}_m} - B_m(t) + \sum_{m' \in \mathcal{M} \setminus \{m\}} \sum_{i \in \mathcal{C}_{m'}} T_i(t) R(i, m). \end{aligned} \quad (10)$$

$$= G_m(0) - t\{1 - \bar{\rho}(m)\} + \{t - B_m(t)\} - \sum_{m' \in \mathcal{M} \setminus \{m\}} \{t\bar{R}(m', m) - \sum_{i \in \mathcal{C}_{m'}} T_i(t) R(i, m)\}. \quad (11)$$

Now consider part (a) of Lemma 1. At a regular time t of $G_m(t)$ for which $W_m(t) > 0$, it must be true that $\dot{B}_m(t) = 1$. Hence, taking derivatives through (11) we obtain

$$\begin{aligned} \dot{G}_m(t) &= -\{1 - \bar{\rho}(m)\} - \sum_{m' \in \mathcal{M} \setminus \{m\}} \{\bar{R}(m', m) - \sum_{i \in \mathcal{C}_{m'}} \dot{T}_i(t) R(i, m)\} \\ &\leq -\{1 - \bar{\rho}(m)\}. \end{aligned} \quad (12)$$

Inequality (12) is a straightforward consequence of (4) and (7). Hence, part (a) of Lemma 1 holds with $\epsilon_m = 1 - \bar{\rho}(m) > 0$, $m \in \mathcal{M}$. Note also that, by definition of the quantities concerned,

$$W_m(t) = 0 \Rightarrow G_m(t) = 0, \quad m \in \mathcal{M}$$

and Lemma 1 part (b) holds trivially.

Hence, under the peak-rate stability condition (6), the requirements of Lemma 3(a), (b) are met and by Lemma 3(iii) we conclude that the original stochastic network must also be stable under all non-idling policies. This concludes the proof of Theorem 1. \diamond

3 A generalised Lu-Kumar network

Motivated by the considerations outlined in the penultimate paragraph of the Introduction, we describe an extension of the two station network investigated by Lu and Kumar (1991). This generalisation is able to model a range of assumptions about the degree of connectivity between its two stations and has a parameter (q) which is a natural measure of this. Because of the simplicity of the network's structure we are able to obtain a sufficient condition for stability which comes close to determining the network's

global stability region (i.e., the region over which the network is stable under all non-idling policies) exactly. The network then becomes a natural vehicle for testing proposed sufficient conditions for stability across a range of assumptions about the weight of between-station traffic. Such an analysis of the peak-rate stability condition discussed in Section 2 is given in Section 4.

Specifically, our network is a member of the family of open networks introduced in Section 2. There are 4 customer classes, numbered $\{1, 2, 3, 4\}$ and 2 single-server stations, numbered $\{1, 2\}$. Classes 1 and 4 are served at station 1 (i.e., $\mathcal{C}_1 = \{1, 4\}$) with classes 2 and 3 served at station 2 ($\mathcal{C}_2 = \{2, 3\}$). Classes 1 and 2 arrive exogonously at the network with rates $\alpha_1 = 1$ and $\alpha_2 = 1 - q$ respectively, where $0 \leq q \leq 1$. The mean service time for class i customers is m_i and the corresponding rate $\mu_i, 1 \leq i \leq 4$, as before. The positive entries in the routing matrix \mathbf{P} are $p_{12} = q, p_{14} = 1 - q, p_{23} = 1, p_{34} = q, p_{30} = 1 - q, p_{40} = 1$. Note that when $q = 0$ the network consists of two autonomous stations, while $q = 1$ gives the reentrant line studied by Lu and Kumar (1991) under Markovian assumptions. See Figure 1.

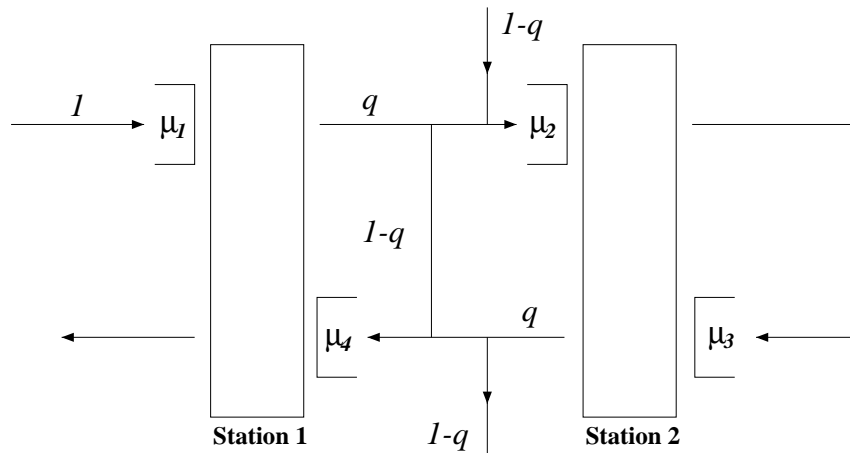


Figure 1: Generalized Lu-Kumar network.

Note that the standard necessary stability condition (1) in this case is

$$\rho(1) = m_1 + m_4 < 1 \text{ and } \rho(2) = m_2 + m_3 < 1. \quad (13)$$

With condition (13) we associate the corresponding region of $\mathbf{m} = (m_1, m_2, m_3, m_4)$ -space, namely,

$$R = \{\mathbf{m} > \mathbf{0} : \rho(1) < 1 \text{ and } \rho(2) < 1\}. \quad (14)$$

Our sufficient condition for stability for this family of networks is given as Theorem 2. The proof utilises the fluid model approach described in Section 2 and, in particular, makes heavy use of Lemma 1, adapted to this case.

Theorem 2

When (13) holds, the fluid model for the generalised Lu-Kumar network is stable under all non-idling policies when $m_2 + qm_4 < 1$.

Proof

In order to utilise Lemma 1, we introduce Lyapunov function $G(t) = \max\{G_1(t), G_2(t)\}$ given by

$$\begin{aligned} G_1(t) &= \theta_1 L_1(t) + (1 - \theta_1)[(1 - q)L_1(t) + q^2 L_1(t) + qL_2(t) + qL_3(t) + L_4(t)], \\ G_2(t) &= \theta_2[qL_1(t) + L_2(t)] + (1 - \theta_2)[qL_1(t) + L_2(t) + L_3(t)]. \end{aligned} \tag{15}$$

We shall show that, when (13) holds and $m_2 + qm_4 < 1$ then it is possible to choose $\theta_1, \theta_2 \in (0, 1)$ in such a way that G satisfies the conditions (a) and (b) of Lemma 1. By the result in Lemma 1(iii), this is enough to establish the theorem.

We firstly consider (b). Note that

$$\begin{aligned} W_2(t) = m_2 L_2(t) + m_3 L_3(t) = 0 &\Rightarrow L_2(t) = L_3(t) = 0 \Rightarrow \\ G_1(t) &= \{1 - (1 - \theta_1)q(1 - q)\}L_1(t) + (1 - \theta_1)L_4(t), \\ G_2(t) &= qL_1(t) \end{aligned}$$

and hence

$$G_2(t) \leq G_1(t), \quad 0 < \theta_1 < 1, 0 \leq q \leq 1. \tag{16}$$

Further,

$$W_1(t) = m_1 L_1(t) + m_4 L_4(t) = 0 \Rightarrow L_1(t) = L_4(t) = 0 \Rightarrow$$

$$G_1(t) = (1 - \theta_1)q\{L_2(t) + L_3(t)\},$$

$$G_2(t) = L_2(t) + (1 - \theta_2)L_3(t)$$

and so

$$(1 - \theta_1)q \leq 1 - \theta_2 \Rightarrow G_1(t) \leq G_2(t). \quad (17)$$

Now consider (a). From an appropriate version of (7) we conclude, upon taking derivatives at all regular points (almost everywhere), that

$$\dot{G}_1(t) = 1 - \theta_1 \mu_1 \dot{T}_1(t) - (1 - \theta_1) \mu_4 \dot{T}_4(t), \quad (18)$$

$$\dot{G}_2(t) = 1 - \theta_2 \mu_2 \dot{T}_2(t) - (1 - \theta_2) \mu_3 \dot{T}_3(t). \quad (19)$$

Now, if $W_1(t) > 0$ then, under all non-idling policies $\dot{T}_1(t) + \dot{T}_4(t) = 1$. Hence, from (18) we conclude that, for all such policies

$$\dot{G}_1(t) \leq -\min\{\theta_1 \mu_1 - 1, (1 - \theta_1) \mu_4 - 1\} < 0, \text{ if } m_1 < \theta_1 < 1 - m_4. \quad (20)$$

Similarly, if $W_2(t) > 0$ we conclude from (19) that for all non-idling policies

$$\dot{G}_2(t) \leq -\min\{\theta_2 \mu_2 - 1, (1 - \theta_2) \mu_3 - 1\} < 0, \text{ if } m_2 < \theta_2 < 1 - m_3. \quad (21)$$

The conclusion from Lemma 1 and (16), (17), (20) and (21) is that the stochastic network will be stable under all nonidling policies if there exists $(\theta_1, \theta_2) \in (m_1, 1 - m_4) \times (m_2, 1 - m_3)$ such that $(1 - \theta_1)q \leq (1 - \theta_2)$. It is straightforward to establish that this will be so if and only if the lower right corner of the rectangle $(m_1, 1 - m_4) \times (m_3, 1 - m_3)$ satisfies the required inequality, i.e. if and only if $m_2 + qm_4 < 1$. This concludes the proof. \diamond

As in (14), we now write

$$R_q^{fluid} = \{\mathbf{m} > 0 : \rho(1) < 1, \rho(2) < 1 \text{ and } m_2 + qm_4 < 1\} \quad (22)$$

for the region of \mathbf{m} -space corresponding to the sufficient stability condition in Theorem 2. From Theorem 2 we have

$$R_q^{fluid} \subseteq R_q^{stable} \subseteq R, \quad (23)$$

where in (23), R_q^{stable} is the global stability region of the stochastic network.

We now proceed in Section 4 to utilise the generalised Lu-Kumar network in an assessment of the peak-rate stability condition of Section 2.

4 Assessing the peak-rate stability condition on the generalised Lu-Kumar network

As indicated in Comment 2 following Theorem 1, there are sound theoretical reasons for believing that the simple peak-rate stability condition of Section 2 is close to sharp when the flow between stations is light. In this section, we explore its performance more generally by testing it on the generalised Lu-Kumar network of Section 3.

Firstly note that, for general $q \in [0, 1]$, the peak-rate condition of (6) applied to the network of Section 3 is given by

$$\bar{\rho}_q(1) = m_1 + (1 - q)m_4 + qm_4/m_3 < 1, \quad (24)$$

$$\bar{\rho}_q(2) = (1 - q + q/m_1)(m_2 + m_3) < 1.$$

In (24) we have made the q -dependence explicit in the notation $\bar{\rho}_q$. It is easy to show that

$$\bar{\rho}_q(1) < 1 \text{ and } \bar{\rho}_q(2) < 1 \Rightarrow \rho(1) < 1 \text{ and } \rho(2) < 1.$$

In keeping with the ideas and notation developed in Section 3, we use R_q^{peak} to denote the region of \mathbf{m} -space determined by the peak-rate stability condition in (6), i.e.,

$$R_q^{\text{peak}} = \{\mathbf{m} > \mathbf{0} : \bar{\rho}_q(1) < 1 \text{ and } \bar{\rho}_q(2) < 1\}.$$

We shall also use $\text{vol}(R_q^{\text{peak}})$ to denote the corresponding volume, namely

$$\text{vol}(R_q^{\text{peak}}) = \int \int \int \int_{R_q^{\text{peak}}} dm_1 dm_2 dm_3 dm_4.$$

Now, from (23) and some simple algebra we can show that for all $q \in [0, 1]$

$$R_q^{\text{peak}} \subseteq R_q^{\text{fluid}} \subseteq R_q^{\text{stable}} \subseteq R. \quad (25)$$

Taking volumes through (25) we conclude that

$$\text{vol}(R_q^{\text{peak}}) \leq \frac{1}{4} - \frac{q^2}{24} \leq \text{vol}(R_q^{\text{stable}}) \leq \frac{1}{4}, \quad (26)$$

where the second and fourth expressions in (26) can be shown by direct calculation to be $\text{vol}(R_q^{\text{fluid}})$ and $\text{vol}(R)$ respectively.

From the above discussion we propose

$$r(q) \equiv \text{vol}(R_q^{\text{peak}}) / \left(\frac{1}{4} - \frac{q^2}{24} \right), \quad q \in [0, 1]$$

as a (q -dependent) measure of the tightness of our simple peak-rate stability condition in this case. From (26) note that

$$\left(1 - \frac{q^2}{6} \right) r(q) \leq \frac{\text{vol}(R_q^{\text{peak}})}{\text{vol}(R_q^{\text{stable}})} \leq r(q), \quad q \in [0, 1] \quad (27)$$

with the right hand inequality in (27) an equality should R_q^{fluid} be the exact global stability region. Figure 2 shows a plot of $r(q)$. We note that $r(1) = 0.04$ and infer that for the Lu-Kumar network (i.e. $q = 1$) the peak-rate stability condition captures about 4% of the stability region. However, $r(q) \rightarrow 1$, $q \rightarrow 0$ and so, when there is little customer flow between stations the condition is close to exact. Hence, the quality of the

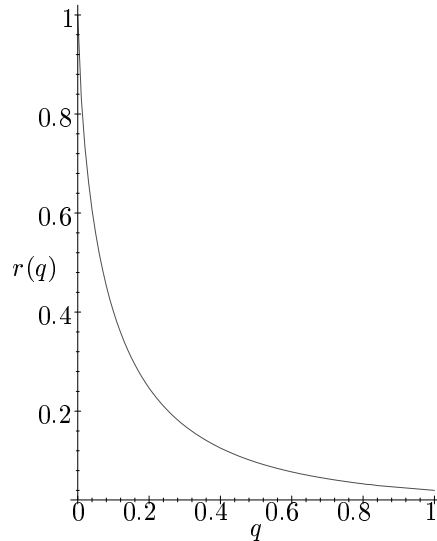


Figure 2: Plot of the ratio $r(q) = \text{vol}(R_q^{\text{peak}})/\text{vol}(R_q^{\text{fluid}})$ for the generalized Lu-Kumar network.

condition is high in conditions of light traffic between stations and is seen to degrade as q increases.

References

- Bramson, M. (1994). Instability of FIFO queueing networks. *Ann. Appl. Probab.* **4** 414–431.
- Dai, J. G. (1995). On positive Harris recurrence of multiclass queueing networks: a unified approach via fluid limit models. *Ann. Appl. Probab.* **5** 49–77.
- Dai, J. G. and G. Weiss (1996). Stability and instability of fluid models for reentrant lines. *Math. Oper. Res.* **21** 115–134.
- Glazebrook, K. D. and J. Niño-Mora (1999). A linear programming approach to stability, optimization and performance analysis for Markovian multiclass queueing networks. *Ann. Oper. Res.* **92** 1–18. Special volume on Models and Algorithms for Planning and Scheduling Problems, C.A. Glass, C.N. Potts, V.A. Strusevich and R.R. Weber (Eds.).
- Kumar, P. R. and S. P. Meyn (1995). Stability of queueing networks and scheduling policies. *IEEE Trans. Automat. Control* **40** 251–260.
- Kumar, P. R. and S. P. Meyn (1996). Duality and linear programs for stability and performance analysis of queueing networks and scheduling policies. *IEEE Trans. Autom. Control* **41** 4–16.

Lu, S. H. and P. R. Kumar (1991). Distributed scheduling based on due dates and buffer priorities.
IEEE Trans. Automat. Control **36** 1406–1416.