STABILITY OF QUEUEING NETWORKS AND SCHEDULING POLICIES*

P. R. Kumar and Sean P. Meyn†

Abstract

Usually, the stability of queueing networks is established by explicitly determining the invariant distribution. However, outside of the narrow class of queueing networks possessing a product form solution, such explicit solutions are rare, and consequently little is known concerning stability too.

We develop here a programmatic procedure for establishing the stability of queueing networks and scheduling policies. The method uses linear or nonlinear programming to determine what is an appropriate quadratic functional to use as a Lyapunov function. If the underlying system is Markovian, our method establishes not only positive recurrence and the existence of a steady–state probability distribution, but also the geometric convergence of an exponential moment.

We illustrate this method on several example problems. For an example of an open re-entrant line, we show that all stationary non-idling policies are stable for all load factors less than one. This includes the well known First Come First Serve (FCFS) policy. We determine a subset of the stability region for the Dai–Wang example, for which they have shown that the Brownian approximation does not hold. In another re-entrant line, we show that the Last Buffer First Serve (LBFS) and First Buffer First Serve (FBFS) policies are stable for all load factors less than one. Finally, for the Rybko–Stolyar example, for which a subset of the instability region has been determined by them under a certain buffer priority policy, we determine a subset of the stability region.

1 Introduction

Usually, the stability of a queueing network is established by explicitly determining an invariant distribution. However, outside of the relatively narrow class of queueing networks

*The research of the first author has been supported in part by the NSF under Grant Nos. ECS-90-25007 and ECS-92-16487, and in part by the USARO under Contract No. DAAL-03-91-G-0182.

†Department of Electrical and Computer Engineering, and the Coordinated Science Laboratory, University of Illinois, 1308 West Main Street, Urbana, IL 61801.

admitting a product form solution for the invariant distribution, such explicit solutions are rare. Consequently, stability results are also rare.

Here we develop a procedure for establishing stability of a queueing network operating under a scheduling policy. It is based on just solving a linear program on the coefficients of a quadratic form. Alternatively, a nonlinear program can also be used. The goal is to programmatically construct a quadratic Lyapunov function on buffer levels that has a negative drift, whenever the mean number of parts in the system is large. This allows one to deduce the stability–in–the–mean of a system, even if it is not Markovian. For Markovian systems, such stability is equivalent to the existence of a steady–state distribution, i.e., positive recurrence. Moreover, for Markovian systems, our method also establishes geometric convergence of an exponential moment.

Such stability results are important for a variety of reasons. First, they are a precursor to more fine grained questions concerning the performance levels of various scheduling policies. Second, several unstable scheduling policies have recently been discovered. Kumar and Seidman [2] and Chase and Ramadge [3] provide examples of deterministic systems which are unstable under all clearing, i.e., exhaustive service, policies. Lu and Kumar [1] provide an example of a re-entrant line with deterministic processing times, and deterministic bursty arrivals, for which a certain buffer priority policy is unstable. Rybko and Stolyar [4] provide an example of a stochastic network which is unstable under a certain buffer priority policy. Recently, Seidman [5] has demonstrated the instability of the well known First Come First Serve (FCFS) policy, also for a deterministic model. Bramson [6] has recently constructed a stochastic re-entrant line, that is also unstable under the FCFS policy. Third, there has been much recent interest in the use of heavy traffic Brownian approximations to construct scheduling policies for queueing networks; see Harrison [7] and Harrison and Wein [8]. Clearly, to establish heavy traffic limit theorems, it is necessary to establish the stability of the queueing networks involved. Dai and Wang [12] have constructed a counterexample where the Brownian approximation does not hold; see also Whitt [13] and Dai and Nguyen [14]. Indeed, heavy traffic limit theorems appear to be only available for systems that are already known to be stable; see Reiman [9, 10] and Peterson [11].

Quadratic Lyapunov functions find widespread use in linear system theory. For stochastic systems, Kingman [15] has used a quadratic Lyapunov function to analyze a random walk on Z_+^2. Fayolle [16] has used general quadratic forms to characterize ergodicity of random walks on Z_+^n. Piecewise linear Lyapunov functions are used in Fayolle, Malyshev, Mensikov and Sidorenko [17] for establishing the stability of Jackson networks [19], Meyn and Down [18] use the square of the workload to establish the stability of generalized Jackson networks, where the assumptions on the arrival processes and service times are relaxed. Coffman, Johnson, Shor and Weber [20] have used linear programming to find both linear and quadratic forms with negative or positive drift, and thus the stability or instability of a certain bin packing algorithm. Their linear programs test for a drift of the appropriate sign at all the states on the boundary of a prescribed hypercube; our approach may be less computationally complex. Recently, Bertsimas, Paschalidis and Tsitsiklis [21] and Kumar and Kumar [22] have used quadratic forms to obtain performance bounds for queueing networks and scheduling policies, provided the system is stable, and has a bounded second moment.

2 The Basic Open Re-Entrant Line

To expose the idea in its simplest form, we begin with the treatment of an *open re-entrant line*; see [23].

The network consists of S machines $\{1, 2, \ldots, S\}$; see Figure 1. Parts arrive as a Poisson process of rate λ to buffer b_1, located at machine $\sigma(1) \in \{1, \ldots, S\}$. Upon completing service, they proceed to buffer b_2 located at machine $\sigma(2) \in \{1, \ldots, S\}$. Let b_L at machine $\sigma(L)$ be the last buffer visited. The sequence $\{\sigma(1), \ldots, \sigma(L)\}$ is the *route* of the part. Since one may have $\sigma(i) = \sigma(j)$ for some pairs i and j with $i \neq j$, we say that the system is a *re-entrant line*. Let us suppose that parts in b_i require an exponentially distributed service time, with mean $\frac{1}{\mu_i}$, from machine $\sigma(i)$.

Let $I(i) := \{j \mid \sigma(j) = \sigma(i)\}$, i.e., the set of indices of buffers which are located at the same machine as b_i. Thus, the buffers with indices in $I(i)$ are in "contention" for the same machine.

Let us denote

$$x_i(t) \quad := \quad \text{Numbers of parts in buffer } b_i \text{ at time } t, \text{ including any in service,}$$

and

$$w_i(t) \quad := \quad 1 \text{ if machine } \sigma(i) \text{ is working on a part in buffer } b_i \text{ at time } t, \text{ and } 0 \text{ otherwise.}$$

For simplicity, we suppose that a machine works on only one part at any given time.

The key problem in "scheduling" such queueing networks is to determine which part in which buffer the machine should serve, i.e., which $w_i(t)$ should be 1. Clearly, an optimal choice for reasonable criteria will depend on the location, i.e., the buffer, occupied by the part. However, when the service priority depends on the buffer (i.e., "class" of a part), the steady–state distribution, if any, is not known, and, as mentioned earlier, neither is stability. As an example, the well known First Come First Serve Policy (FCFS) can be stable or unstable for particular systems, when the μ_i's are not the same for all buffers at a machine. Similarly, buffer priority policies can be stable or unstable for particular systems and values of parameters.

3 Copositive Matrices

Let Q be a symmetric $L \times L$ matrix that gives rise to a quadratic form which is non–negative on the positive orthant in $\mathbb{R}^L$, i.e.,

$$Q = Q^T \text{ and } (y_1, \ldots, y_L)Q(y_1, \ldots, y_L)^T \geq 0, \text{ whenever } y_i \geq 0 \text{ for all } i. \tag{1}$$

Such matrices are called symmetric *copositive* matrices. As will be shown in Section 5, our methodology will automatically confine itself to the sub–class of symmetric *strictly* copositive matrices. These are symmetric copositive matrices Q for which additionally,

$$(y_1, \ldots, y_L)Q(y_1, \ldots, y_L)^T > 0, \text{ if } y_i \geq 0 \text{ for all } i, \text{ and } y_i \neq 0 \text{ for some } i.$$

Copositive and strictly copositive matrices have been extensively studied; see Cottle, Habetler and Lemke [24]. They are characterized by the signs of certain determinants (see [24]):

Keller's Theorem. *A symmetric matrix is copositive if and only if each principal submatrix, for which the cofactors of the last row are nonnegative, has a nonnegative determinant.*

A recent algorithm for testing copositivity can be found in Andersson, Chang and Elfving [25]. However, the determination of copositivity is NP–Complete; see Murty and Kabadi [26].

The procedure we advocate below could be used with any Q satisfying (1). For concreteness, we will confine our attention to the following special types of copositive matrices,

It is easy to see that any symmetric, *non–negative* matrix, i.e., one for which, $Q^T = Q = [q_{ij}]$, with $q_{ij} \geq 0$ for all i, j, is copositive. Also, any positive semidefinite matrix, i.e., a Q for which, $Q = Q^T$ and $x^T Q x \geq 0$ for all x, is copositive. Moreover, any convex combination (or linear combination with positive weights) of such matrices is also copositive, since the set of symmetric copositive matrices is a convex cone.

4 The Basic Characterization

We shall rescale time so that,

$$\lambda + \sum_{i=1}^{L} \mu_i \quad = \quad 1, \tag{2}$$

and resort to "uniformization," see Lippman [27]. That is, we shall suppose that there is always either a real or a "virtual" part that is being served at every buffer b_i . Let $\{\tau_n\}$, with $\tau_0 = 0$, denote the sequence of all arrival and service time, real or virtual, and let $\mathcal{F}_{\tau_n}$ denote the σ-field generated by events up to time τ_n.

Let $x(t) := (x_1(t), \ldots, x_L(t))^T$ denote the vector of queue lengths. In accordance with terminology of Markov Decision Processes, we will call $x(t)$ the "state." A policy whose action at any time t depends only on $x(t)$ is called *stationary*, again in accordance with the terminology of Markov Decision Processes. Under a stationary policy, the system is described by the Markov chain $\{x(t)\}$.

We will treat a larger class of scheduling policies than stationary policies. We will consider any scheduling policy which takes a constant action in intervals of the form $[\tau_n, \tau_{n+1})$, and call such a policy *non–interruptive*. (The term non–interruptive should not be confused with the term non–preemptive). As an example, the well known First Come First Serve (FCFS) policy is non–interruptive. Note that any scheduling policy that does not change actions between *real* transition epochs is non–interruptive. Of course, all stationary policies are of this form, and are hence automatically non–interruptive.

We will allow preemptive priority at an epoch, if the scheduling policy calls for it.

Let $e_i := (0, \ldots, 0, 1, 0, \ldots, 0)^T$ be the i-th unit vector. The state transition diagram of the network is as shown in Figure 2, for any non–interruptive policy.

Let us consider the quadratic form $x^T(\tau_n)Qx(\tau_n)$. Note that since $x(\tau_n)$ can grow no faster than linearly in n, the conditional expectation $\mathrm{E}[x^T(\tau_{n+1})Qx(\tau_{n+1}) \mid \mathcal{F}_{\tau_n}]$ exists. From Figure 2, for any non–interruptive policy, we obtain

$$\begin{aligned}
\mathrm{E}[x^T(\tau_{n+1})Qx(\tau_{n+1}) \mid \mathcal{F}_{\tau_n}] &= \lambda(x(\tau_n)+e_1)^T Q(x(\tau_n)+e_1) \\
&+ \sum_{i=1}^{L-1} \mu_i w_i(\tau_n)(x(\tau_n)-e_i+e_{i+1})^T Q(x(\tau_n)-e_i+e_{i+1}) \\
&+ \mu_L w_L(\tau_n)(x(\tau_n)-e_L)^T Q(x(\tau_n)-e_L) \\
&+ \sum_{i=1}^{L} \mu_i(1-w_i(\tau_n))x^T(\tau_n)Qx(\tau_n).
\end{aligned} \tag{3}$$

Using (2) and the symmetry of Q as in (1), we obtain

$$\begin{aligned}
\mathrm{E}[x^T(\tau_{n+1})Qx(\tau_{n+1}] \mid \mathcal{F}_{\tau_n}] &= x^T(\tau_n)Qx(\tau_n) + 2\lambda e_1^T Qx(\tau_n) + \lambda e_1^T Q e_1 \\
&+ 2\sum_{i=1}^{L-1} \mu_i w_i(\tau_n)(e_{i+1}-e_i)^T Qx(\tau_n) \\
&+ \sum_{i=1}^{L-1} \mu_i w_i(\tau_n)(e_{i+1}-e_i)^T Q(e_{i+1}-e_i) \\
&- 2\mu_L w_L(\tau_n)e_L^T Qx(\tau_n) + \mu_L w_L(\tau_n)e_L^T Q e_L.
\end{aligned}$$

Note now that since $w_i(\tau_n) = 0$ or 1, all the terms not featuring $x(\tau_n)$ above are bounded above by a constant, i.e.,

$$\lambda e_1^T Q e_1 + \sum_{i=1}^{L-1} \mu_i w_i(\tau_n)(e_{i+1} - e_i)^T Q(e_{i+1} - e_i) + \mu_L w_L(\tau_n) e_L^T Q e_L \leq M < +\infty.$$

Hence,

$$\begin{aligned} \mathrm{E}[x^T(\tau_{n+1}) Q x(\tau_{n+1}) \mid \mathcal{F}_{\tau_n}] \quad \leq \quad & x^T(\tau_n) Q x(\tau_n) + 2\lambda e_1^T Q x(\tau_n) \\ & +2 \sum_{i=1}^{L-1} \mu_i w_i(\tau_n)(e_{i+1} - e_i)^T Q x(\tau_n) \\ & -2\mu_L w_L(\tau_n) e_L^T Q x(\tau_n) + M. \end{aligned}$$

Let us suppose that the initial condition is deterministic (or more generally, a bounded random variable, or even more generally, has a finite second moment). As noted earlier, $x(\tau_n)$ grows no faster than linearly in n. Hence $\mathrm{E}(x^2(\tau_n))$ exists for every n. By taking the unconditional expectation, we obtain,

$$\begin{aligned} & \mathrm{E}[x^T(\tau_{n+1}) Q x(\tau_{n+1})] \leq \mathrm{E}[x^T(\tau_n) Q x(\tau_n)] \\ & \quad +2\lambda e_1^T Q \mathrm{E}[x(\tau_n)] + 2 \sum_{i=1}^{L-1} \mu_i \mathrm{E}[w_i(\tau_n)(e_{i+1} - e_i)^T Q x(\tau_n)] \\ & \quad -2\mu_L \mathrm{E}[w_L(\tau_n) e_L^T Q x(\tau_n)] + M. \end{aligned} \tag{4}$$

Let us denote,

$$z_{ij}(\tau_n) \quad := \quad w_i(\tau_n) x_j(\tau_n). \tag{5}$$

Using (5), (4) can be written as

$$\begin{aligned} \mathrm{E}[x^T(\tau_{n+1}) Q x(\tau_{n+1})] \leq \mathrm{E}(x^T(\tau_n) Q x(\tau_n)) + 2\lambda \sum_{j=1}^{L} q_{1j} \mathrm{E}(x_j(\tau_n)) \\ +2 \sum_{i=1}^{L-1} \mu_i \sum_{j=1}^{L} (q_{i+1,j} - q_{ij}) \mathrm{E}(z_{ij}(\tau_n)) - 2\mu_L \sum_{j=1}^{L} q_{Lj} \mathrm{E}(z_{Lj}(\tau_n)) + M. \end{aligned}$$

By summing over n, and telescoping, we obtain

$$\begin{aligned} \frac{1}{N+1} \sum_{n=0}^{N} \left[-\lambda \sum_{j=1}^{L} q_{1j} \mathrm{E}(x_j(\tau_n)) - \sum_{i=1}^{L-1} \mu_i \sum_{j=1}^{L} (q_{i+1,j} - q_{ij}) \mathrm{E}(z_{ij}(\tau_n)) + \mu_L \sum_{j=1}^{L} q_{Lj} \mathrm{E}(z_{Lj}(\tau_n)) \right] \\ \leq \frac{1}{2(N+1)} (\mathrm{E}[x^T(0) Q x(0)] - \mathrm{E}[x^T(\tau_{N+1}) Q x(\tau_{N+1})]) + \frac{M}{2} \leq M' < +\infty \text{ for all } N. \end{aligned} \tag{6}$$

In the last inequality above, we have used the non–negativity of $x^T(\tau_{N+1})Qx(\tau_{N+1})$, which is guaranteed by the copositivity condition (1), since $x(\tau_{N+1})$ lies in the non–negative orthant.

Now note that *if* we can find a $\gamma > 0$ so that,

$$\lambda \sum_{j=1}^{L} q_{1j} \mathrm{E}(x_j(\tau_n)) \quad + \quad \sum_{i=1}^{L-1} \mu_i \sum_{j=1}^{L} (q_{i+1,j} - q_{ij}) \mathrm{E}(z_{ij}(\tau_n)) - \mu_L \sum_{j=1}^{L} q_{Lj} \mathrm{E}(z_{Lj}(\tau_n)) \leq -\gamma \sum_{j=1}^{L} \mathrm{E}(x_j(\tau_n)),$$

then from (6) we would have *stability–in–the–mean*, i.e.,

$$\frac{1}{N+1} \sum_{n=0}^{N} \sum_{j=1}^{L} \mathrm{E}(x_j(\tau_n)) \leq M'' < +\infty \text{ for all } N. \tag{7}$$

Before we pursue the issue of finding such a γ, we point out certain consequences of stability–in–the–mean for stationary, non–idling policies. In the rest of this paper we will restrict attention to scheduling policies that are *non–idling*, i.e., whenever one of the buffers at a machine is non-empty, then the machine cannot stay idle. For stationary, non–idling policies, $\{x(\tau_n)\}$ is a time–homogeneous, countable state, Markov chain, which has a single communicating class that is aperiodic (since the origin can be reached from every state, and the system can stay at the origin for two consecutive time steps). The condition (7) then guarantees positive recurrence, i.e., the existence of an unique steady state probability distribution. To see this, note that if the chain is not positively recurrent, then the probability that the chain is in a fixed finite set of states converges to zero as $n \to \infty$. However, then, that is so even for the finite set of states $\{x : \sum_1^L x_i \leq M''$, and all $x_i \geq 0$ and integral$\}$. This contradicts (7). Moreover, the Markov chain has bounded first moment, and the mean total number of customers converges to a finite steady state value. In fact, we will show in the next section that it even establishes the geometric convergence of an exponential moment.

Let us now see how to assure (7) for some $\gamma > 0$. We will actually work at assuring that the inequality (7) holds without the expectation being taken, i.e.,

$$\begin{aligned} \lambda \sum_{j=1}^{L} q_{1j} x_j(\tau_n) + \sum_{i=1}^{L-1} \mu_i \sum_{j=1}^{L} (q_{i+1,j} \quad &- \quad q_{ij}) z_{ij}(\tau_n) - \mu_L \sum_{j=1}^{L} q_{Lj} z_{Lj}(\tau_n) \\ &\leq \quad -\gamma \sum_{j=1}^{L} x_j(\tau_n). \end{aligned} \tag{8}$$

Let us now motivate the reason for restricting our attention to non–idling policies, in our tests for stability. Note that the coefficients λq_{1j} of the $x_j(\tau_n)$'s on the left hand side above are all non–negative, while the corresponding coefficient $(-\gamma)$ on the right hand side is negative. Clearly, to assure (8) it is necessary that there exist some choice of constants $\{\alpha_{ij}\}$ for which,

$$\sum_{j=1}^{L} x_j(\tau_n) \le \sum_{i=1}^{L}\sum_{j=1}^{L} \alpha_{ij} z_{ij}(\tau_n).$$

Focusing on a fixed index j, one will in particular need $x_j(\tau_n)$ to be bounded above by some linear combination of $\{z_{ij}(\tau_n) : 1 \le i \le L\}$. This can only be assured if some machine is guaranteed to be working whenever $x_j > 0$; hence the restriction to non–idling scheduling policies in the sequel.

Let us return to (8). For notational convenience, let us define

$$q_{L+1,j} := 0 \text{ for all } j = 1, 2, \ldots, L.$$

Focusing on a fixed value of the index j, it is clear that (8) is assured, if

$$\lambda q_{1j} x_j(\tau_n) + \sum_{i=1}^{L} \mu_i (q_{i+1,j} - q_{ij}) z_{ij}(\tau_n) \le -\gamma x_j(\tau_n) \text{ for all } j = 1, 2, \ldots, L. \tag{9}$$

Grouping the terms by machine, i.e., using $\{1, 2, \ldots, L\} = \bigcup_\sigma \{i : \sigma(i) = \sigma\}$, we see that the LHS of (9) satisfies,

$$\begin{aligned}
&\lambda q_{1j} x_j(\tau_n) + \sum_{i=1}^{L} \mu_i (q_{i+1,j} - q_{ij}) z_{ij}(\tau_n) \\
&= \lambda q_{1j} x_j(\tau_n) + \sum_{\sigma=1}^{S} \sum_{\{i:\sigma(i)=\sigma\}} \mu_i (q_{i+1,j} - q_{ij}) z_{ij}(\tau_n) \\
&\le \lambda q_{1j} x_j(\tau_n) + \sum_{\sigma=1}^{S} \max_{\{i:\sigma(i)=\sigma\}} \mu_i (q_{i+1,j} - q_{ij}) \sum_{\{i:\sigma(i)=\sigma\}} z_{ij}(\tau_n) \\
&= \lambda q_{1j} x_j(\tau_n) + \max_{\{i:\sigma(i)=\sigma(j)\}} \mu_i (q_{i+1,j} - q_{ij}) \sum_{\{i:\sigma(i)=\sigma(j)\}} z_{ij}(\tau_n) \\
&\quad + \sum_{\{\sigma':\sigma'\neq\sigma(j)\}} \max_{\{i:\sigma(i)=\sigma'\}} \mu_i (q_{i+1,j} - q_{ij}) \sum_{\{i:\sigma(i)=\sigma'\}} z_{ij}(\tau_n).
\end{aligned} \tag{10}$$

We now investigate how to assure that the RHS of (10) can be bounded above by

$-\gamma x_j(\tau_n)$. For non-idling policies,

$$\sum_{\{i:\sigma(i)=\sigma\}} x_i(\tau_n) > 0 \;\;\Rightarrow\;\; \sum_{\{i:\sigma(i)=\sigma\}} w_i(\tau_n) = 1,$$

and so,

$$x_j(\tau_n) \;\;=\;\; \sum_{\{i:\sigma(i)=\sigma(j)\}} w_i(\tau_n)x_j(\tau_n).$$

Hence,

$$x_j(\tau_n) \;\;=\;\; \sum_{\{i:\sigma(i)=\sigma(j)\}} z_{ij}(\tau_n). \tag{11}$$

Moreover, since other machines need not be working when b_j is non-empty, we have the non-idling inequalities,

$$x_j(\tau_n) \;\;\geq\;\; \sum_{\{i:\sigma(i)=\sigma'\}} z_{ij}(\tau_n) \text{ for all } \sigma' \neq \sigma(j). \tag{12}$$

Employing (11, 12), in (10), we obtain

$$\begin{aligned} &\lambda q_{1j}x_j(\tau_n) + \sum_{i=1}^{L} \mu_i(q_{i+1,j} - q_{ij})z_{ij}(\tau_n) \\ &\leq \;\; \lambda q_{1j}x_j(\tau_n) + \max_{\{i:\sigma(i)=\sigma(j)\}} (\mu_i(q_{i+1,j} - q_{ij}))x_j(\tau_n) \\ &\quad + \sum_{\{\sigma':\sigma'\neq\sigma(j)\}} \left[\max_{\{i:\sigma(i)=\sigma'\}} \mu_i(q_{i+1,j} - q_{ij})\right]^+ x_j(\tau_n). \end{aligned}$$

Above, $[y]^+ := \max\{y, 0\}$ denotes the positive part of y. The sign needs to be taken into account, since a negative sign of $\mu_i(q_{i+1,j} - q_{ij})$ may reverse the inequality.

Hence we see that if we can find an appropriate set of $\{q_{ij}\}$ for which,

$$\begin{aligned} \lambda q_{1j} + \max_{\{i:\sigma(i)=\sigma(j)\}} \mu_i(q_{i+1,j} - q_{ij}) \;\;+\;\; \sum_{\{\sigma':\sigma'\neq\sigma(j)\}} \left[\max_{\{i:\sigma(i)=\sigma'\}} \mu_i(q_{i+1,j} - q_{ij})\right]^+ \\ \leq \;\; -\gamma \quad \text{for } j = 1, 2, \ldots, L, \end{aligned}$$

for some $\gamma > 0$, then (8), and hence (7), is assured. Thus, we have arrived at the following theorem.

Theorem 1: The Basic Characterization. *Consider the basic open re-entrant line. Suppose there exists a symmetric copositive matrix* $Q = [q_{ij}]$ *which satisfies the following conditions. For every* $j = 1, 2, \ldots, L$,

$$\lambda q_{1,j} + \max_{\{i:\sigma(i)=\sigma(j)\}} \mu_i (q_{i+1,j} - q_{ij}) + \sum_{\{\sigma':\sigma' \neq \sigma(j)\}} \left[\max_{\{i:\sigma(i)=\sigma'\}} \mu_i (q_{i+1,j} - q_{ij}) \right]^+ < 0. \tag{13}$$

(Above, $q_{L+1,j} := 0$ *for all* j*). Then, every non-idling, non–interruptive, scheduling policy is* stable in the mean, *i.e., there exist constants* c, C *such that*

$$\frac{1}{N} \sum_{n=0}^{N-1} \sum_{j=1}^{L} \mathrm{E}(x_j(\tau_n)) \quad \leq \quad \frac{cE(x^T(0)Qx(0))}{N} + C \qquad \textit{for all } N.$$

Moreover, if the scheduling policy is stationary, then there is an unique steady–state probability distribution.

5 From Stability to Geometric Convergence of an Exponential Moment

In fact, for a Markovian system, the above Lyapunov based negative drift argument actually establishes the geometric convergence of an exponential moment (defined below). Thus, in particular it establishes the finiteness of all (polynomial) moments, and their geometric ergodicity.

To see this, we simply work with the square–root of the earlier Lyapunov function. From Figure 2, just as we obtained (3), we obtain,

$$\begin{aligned}
\mathrm{E}[\sqrt{x^T(\tau_{n+1})Qx(\tau_{n+1})} \mid \mathcal{F}_{\tau_n}] \quad &= \quad \lambda \sqrt{(x(\tau_n) + e_1)^T Q (x(\tau_n) + e_1)} \\
&+ \quad \sum_{i=1}^{L-1} \mu_i w_i(\tau_n) \sqrt{(x(\tau_n) - e_i + e_{i+1})^T Q (x(\tau_n) - e_i + e_{i+1})} \\
&+ \quad \mu_L w_L(\tau_n) \sqrt{(x(\tau_n) - e_L)^T Q (x(\tau_n) - e_L)} \\
&+ \quad \sum_{i=1}^{L} \mu_i (1 - w_i(\tau_n)) \sqrt{x^T(\tau_n) Q x(\tau_n)}.
\end{aligned} \tag{14}$$

From the concavity of the square–root, we obtain,

$$\mathrm{E}[\sqrt{x^T(\tau_{n+1})Qx(\tau_{n+1})} \mid \mathcal{F}_{\tau_n}] \quad \leq \quad [x^T(\tau_n)Qx(\tau_n) + 2\lambda e_1^T Q x(\tau_n)$$

$$+2\sum_{i=1}^{L-1}\mu_i w_i(\tau_n)(e_{i+1}-e_i)^T Qx(\tau_n) \tag{15}$$
$$-2\mu_L w_L(\tau_n)e_L^T Qx(\tau_n)+M]^{1/2}.$$

Now suppose the conditions of Theorem 1 are met. Then, from (8), the right–hand–side above can be bounded as follows,

$$\text{RHS of (15)} \quad \leq \quad \left[x^T(\tau_n)Qx(\tau_n)-2\gamma\sum_{j=1}^{L}x_j(\tau_n)+M\right]^{1/2} \tag{16}$$

Now, from Theorem 14.2.2 of [28], by taking τ there to be the first hitting time of the origin, it follows that there exists a $\delta > 0$, such that $x^TQx \geq \delta\|x\|^2$ for all large enough $\|x\|$ in the positive orthant.[1] (Or, in the case where all $q_{ij} \geq 0$, the fact that $q_{ii} > 0$ follows trivially from the inequality (iii) of Theorem 1). Hence, there exists an $\epsilon > 0$, small enough so that,

$$\text{RHS of (16)} \quad \leq \quad \left[x^T(\tau_n)Qx(\tau_n)\right]^{1/2}-\epsilon, \text{ whenever } \sum_{j=1}^{L}x_j(\tau_n)\geq M''',$$

for some large M'''.

Letting $W(\tau_n) := \sqrt{x(\tau_n)^TQx(\tau_n)}$, we have thus shown that,

$$\mathrm{E}[W(\tau_{n+1}) \mid \mathcal{F}_{\tau_n}] \quad \leq \quad W(\tau_n)-\epsilon, \text{ if } x(\tau_n) \text{ lies outside a compact set,}$$

and is bounded when $x(\tau_n)$ is in the compact set. Moreover, the state can jump by only a bounded amount at each transition, and hence $W(\tau_{n+1})-W(\tau_n)$ is bounded. From these two facts, it follows that the Markov chain representing the evolution of the system has a geometrically converging exponential moment; see Theorem 16.3.1 in [28].

Theorem 2: Geometric Convergence of an Exponential Moment. *Consider the basic, open re-entrant line. Suppose that the scheduling policy is stationary, and that all the conditions of Theorem 1 are satisfied. Then the Markov chain $\{x(\tau_n)\}$ has a geometrically converging exponential moment,[2] i.e., there exist $\epsilon > 0$, $r > 1$, and $C < \infty$, such that for any function f satisfying $|f(y)| \leq \exp(\epsilon\|y\|)$ for all y, and any initial condition $x(\tau_0) = x$,*

$$\sum_{n=0}^{\infty}r^n|\mathrm{E}[f(x_{\tau_n})]-\sum_{y}f(y)\pi(y)| < C\exp(\epsilon\|x\|) \textit{ for all } x.$$

[1] Thus Q is actually *strictly* copositive.

[2] This property is called "$\exp(\epsilon\|x\|)$-uniform ergodicity" in [28].

Above, $\pi(y)$ denotes the steady–state probability of the state y. Hence, in particular, the Markov chain admits a finite exponential moment. That is, for some $C' < \infty$,

$$\mathrm{E}[\exp(\epsilon\|x(\tau_n)\|)] \leq C' \exp(\epsilon\|x\|) < \infty \text{ for all } n.$$

The reader is referred to Meyn and Tweedie [29] for estimates of the rate of convergence. It is worth mentioning that the uniformization procedure is just a way of computing the drift $\mathcal{A}x^T Qx$, where $\mathcal{A}$ is the extended generator for the unsampled Markov process. Thus one actually has, for some $\rho < 1$, $|\mathrm{E}[f(x_t)] - \sum_y f(y)\pi(y)| < C\rho^t V(x)$ for all x, and all $t \geq 0$, i.e., a similar geometric convergence for the original unsampled chain.

6 A Linear Programming Characterization

As noted earlier, if Q is a symmetric non–negative matrix, then it is copositive. Note now that the LHS of (13) in Theorem 1 is homogeneous in Q. Hence, if (13) is valid, then by multiplying Q by arbitrarily large positive numbers, one can drive the value of the LHS of the inequality (13) in Theorem 1 to $-\infty$. This allows us to provide a sufficient condition for stability in terms of the unboundedness of a linear program.

Theorem 3: A Linear Programming Characterization. *Consider the basic, open re-entrant line. Suppose that the following linear program has an unbounded solution:*

$$\text{Max} \quad \gamma$$

subject to the constraints:

$$\lambda q_{1j} + r_j + \sum_{\{\sigma':\sigma' \neq \sigma(j)\}} s_{\sigma',j} + \gamma \leq 0 \quad \text{for all } j$$

$$r_j \geq \mu_i(q_{i+1,j} - q_{ij}) \quad \text{for all } i \in I(j), \text{ and for all } j$$

$$s_{\sigma,j} \geq \mu_i(q_{i+1,j} - q_{ij}) \quad \text{for all } i \text{ with } \sigma(i) = \sigma, \text{ and all } j$$

$$q_{ij} = q_{ji} \quad \text{for all } i, j$$

$$q_{L+1,j} = 0 \quad \text{for all } j$$

$$q_{ij} \geq 0 \quad \textit{for all } i, j$$

$$s_{\sigma j} \geq 0 \quad \textit{for all } \sigma, j$$

$$r_j \textit{ unrestricted in sign.}$$

Then, every non-idling, non–interruptive policy is stable in the mean. Moreover, every non–idling, stationary policy has a geometrically converging exponential moment.

The number of variables $\{q_{ij}, r_k, s_{\sigma j}, \gamma\}$ in the above linear program is $\frac{L(L+1)}{2} + L + (S - 1)L + 1$. (Note that the variables $s_{\sigma(j),j}$, q_{ij} with $i > j$, and $q_{L+1,j}$ are not really needed). The number of constraints is $L + \sum_{j=1}^{L} |I(j)| + \sum_{j=1}^{L} \sum_{\sigma:\sigma \neq \sigma(j)} |\{i : \sigma(i) = \sigma\}|$. It may be possible to rewrite the linear program more economically.

We find it convenient to slightly modify the linear program in Theorem 3 by bounding γ by 1. Thus if the value of the linear program is 1, then one deduces stability (rather than from the unboundedness of the value as in Theorem 3).

Corollary 1: A 0-1 Test of Stability. *Consider the same linear program as in Theorem 3, except that we impose the additional constraint,*

$$\gamma \leq 1.$$

If the linear program has value 1, then every non–idling non–interruptive policy is stable–in–the–mean. Moreover, every non–idling, stationary policy has a geometrically converging exponential moment. However, if the value of the linear program is 0, then no conclusion can be drawn regarding stability or instability.

7 Example: All Non-Idling Policies Stable

Consider the system shown in Figure 3. Then, to show that there exists a $Q = Q^T$ satisfying (7), it is sufficient to find $q_{ij} = q_{ji} \geq 0$, so that

$$[\lambda q_{11}, \lambda q_{12}, \lambda q_{13}] \begin{bmatrix} x_1(\tau_n) \\ x_2(\tau_n) \\ x_3(\tau_n) \end{bmatrix}$$

$$
\begin{aligned}
+ \quad & [-\mu_1 q_{11} + \mu_1 q_{12}, -\mu_1 q_{12} + \mu_1 q_{22}, -\mu_1 q_{13} + \mu_1 q_{23}, \\
& -\mu_2 q_{12} + \mu_2 q_{13}, -\mu_2 q_{22} + \mu_2 q_{23}, -\mu_2 q_{23} + \mu_2 q_{33}, \\
& -\mu_3 q_{13}, -\mu_3 q_{23}, -\mu_3 q_{33}] \begin{bmatrix} z_{11}(\tau_n) \\ z_{12}(\tau_n) \\ z_{13}(\tau_n) \\ z_{21}(\tau_n) \\ z_{22}(\tau_n) \\ z_{23}(\tau_n) \\ z_{31}(\tau_n) \\ z_{32}(\tau_n) \\ z_{33}(\tau_n) \end{bmatrix} \\
\leq \quad & -\gamma[x_1(\tau_n) + x_2(\tau_n) + x_3(\tau_n)].
\end{aligned}
$$

By the non–idling condition, $x_1(\tau_n) = z_{11}(\tau_n) + z_{31}(\tau_n), x_2(\tau_n) = z_{22}(\tau_n)$ and $x_3(\tau_n) = z_{13}(\tau_n) + z_{33}(\tau_n)$. We thus see that it suffices to show that one can choose $\{q_{11}, q_{12}, q_{13}, q_{22}, q_{23}, q_{33}\}$, all non–negative, so that $\lambda q_{11} - \mu_1 q_{11} + \mu_1 q_{12} < 0$, $-\mu_1 q_{12} + \mu_1 q_{22} < 0$, $-\mu_1 q_{13} + \mu_1 q_{23} + \lambda q_{13} < 0$, $-\mu_2 q_{12} + \mu_2 q_{13} < 0$, $-\mu_2 q_{22} + \mu_2 q_{23} + \lambda q_{12} < 0$, $-\mu_2 q_{23} + \mu_2 q_{33} < 0$, $-\mu_3 q_{13} + \lambda q_{11} < 0$, $-\mu_3 q_{23} < 0$, $-\mu_3 q_{33} + \lambda q_{13} < 0$.

Let us suppose that $\mu_1 = \mu_2 = \mu_3 =: \mu$, and $\rho := \frac{2\lambda}{\mu}$. Then the above is equivalent to, $(1-\rho/2)q_{11} > q_{12}, q_{12} > q_{22}$, $(1-\rho/2)q_{13} > q_{23}, q_{12} > q_{13}$, $q_{22} > \rho q_{12}/2 + q_{23}$, $q_{23} > q_{33}, q_{13} > \rho q_{11}/2$, $q_{23} > 0$, $q_{33} > \rho q_{13}/2$.

It is easily checked that if $\rho < 1$, then one can choose $q_{ij} > 0$ which satisfy the above conditions.

Hence, we conclude that if $\mu_1 = \mu_2 = \mu_3 \equiv \mu$ and $\rho := \frac{2\lambda}{\mu} < 1$, then all non-idling, non–interruptive, scheduling policies are stable in the mean. This includes the FCFS policy, which is already known to be stable since $\mu_1 = \mu_3$; see Kelly [30]. Moreover, every non–idling, stationary policy has a geometrically converging exponential moment.

8 The Dai–Wang Example

Dai and Wang [12], (see also Dai and Nguyen [14]), show that the system of Figure 4 does not have a Brownian approximation. It is a basic open re–entrant line with service rates $\mu_1 = 10, \mu_2 = 20, \mu_3 = 10/9, \mu_4 = 20$, and $\mu_5 = 5/4$. Let $\rho := \lambda/\mu_1 + \lambda/\mu_2 + \lambda/\mu_5 =$

$\lambda/\mu_3 + \lambda/\mu_4$ be the load factor on the two machines. Our goal is to determine whether the system is stable for *all* $\rho < 1$.

First note from the equations involving λ in (13) and Corollary 1, that if the value of the linear program is 1 for some λ' then it is 1 for all $\lambda < \lambda'$. Hence there will be critical value λ_{crit}, such that the linear program has value 1 for $\lambda < \lambda_{crit}$ and value 0 for $\lambda > \lambda_{crit}$. Equivalently, there exists such a ρ_{crit}. So we wish to see if $\rho_{crit} = 1$.

Investigating the linear program from Corollary 1, we find that its value is 1 for $\lambda < 0.55587$ (approximately), and 0 for $0.55587 < \lambda < 1$. Thus we can only assert stability for $\rho < 0.95(0.55587) = 0.528$.

9 Buffer Priority Policies

Consider the basic, open, re-entrant line. Suppose that at every machine σ there is a rank ordering of the set of buffers $\{b_i : \sigma(i) = \sigma\}$ served by the machine, according to which preemptive priority is given by the machine. To describe such a *buffer priority* policy more concisely, let $\{\theta(1), \ldots, \theta(L)\}$ be a permutation of $\{1, 2, \ldots, L\}$, with preference given to b_i over b_j if $\theta(i) < \theta(j)$, and both b_i and b_j share the same machine, i.e., $\sigma(i) = \sigma(j)$. The policy is non-idling, stationary, and preemptive.

Then, $x_i(\tau_n) > 0 \Rightarrow w_j(\tau_n) = 0$, if $\theta(i) < \theta(j)$ and $\sigma(i) = \sigma(j)$. Hence $z_{ji} = w_j(\tau_n)x_i(\tau_n) = 0$, if $\theta(i) < \theta(j)$ and $\sigma(i) = \sigma(j)$. As a consequence,

$$
\begin{aligned}
x_j(\tau_n) &= \sum_{\{i: i \in I(j), \theta(i) \leq \theta(j)\}} z_{ij}(\tau_n) && (17) \\
&\geq \sum_{\{i:\sigma(i)=\sigma'\}} z_{ij}(\tau_n) \text{ for } \sigma' \neq \sigma(j). && (18)
\end{aligned}
$$

Recall from Section 4 that our goal is to determine a symmetric copositive Q satisfying,

$$
\begin{aligned}
\lambda q_{1j} x_j(\tau_n) + \sum_{i \in I(j)} \mu_i (q_{i+1,j} - q_{ij}) z_{ij}(\tau_n) \quad &+ \sum_{\{\sigma':\sigma' \neq \sigma(j)\}} \sum_{\{i:\sigma(i)=\sigma'\}} \mu_i (q_{i+1,j} - q_{ij}) z_{ij}(\tau_n) \\
&\leq -\gamma x_j(\tau_n) \text{ for } j = 1, 2, \ldots, L
\end{aligned}
$$

for some $\gamma > 0$. (This can be seen by rewriting the LHS of (9) as in the first equality of

(10)). Using (17), (18), the conditions required to establish stability in Corollary 1 can be relaxed.

Theorem 4: Stability of Buffer Priority Policies. *Consider the basic open re-entrant line. Let* $\{\theta(1), \ldots, \theta(L)\}$ *be a permutation of* $\{1, \ldots, L\}$*. Consider the preemptive buffer priority policy which gives preference to* b_i *over* b_j*, if* $\theta(i) < \theta(j)$ *and both share the same machine. Then, the buffer priority policy* $\theta(\cdot)$ *is stable, with a geometrically converging exponential moment, if the following linear program has value 1:*

$$\text{Max} \quad \gamma$$

subject to

$$\begin{aligned} \lambda q_{1j} + \max_{\{i: i \in I(j) \text{ and } \theta(i) \le \theta(j)\}} \mu_i(q_{i+1,j} &- q_{ij}) + \sum_{\{\sigma' : \sigma' \neq \sigma(j)\}} \left[\max_{\{i: \sigma(i) = \sigma'\}} \mu_i (q_{i+1,j} - q_{ij}) \right]^+ + \gamma \\ &\le 0 \quad \text{for } j = 1, 2, \ldots, L, \\ \gamma &\le 1, \\ q_{ij} &= q_{ji} \ge 0 \quad \text{for } 1 \le i, j \le L. \end{aligned}$$

10 Example: LBFS and FBFS are stable

Consider the system of Figure 3. Unlike in Section 7, we do not require that all the μ_i's are equal.

First, let us examine the class of *all non–idling policies*. Can we prove that all non-idling, non–interruptive policies are stable? To investigate, we consider the following linear program:

$$\text{Max} \quad \gamma \tag{19}$$

subject to the constraints:

$$\lambda q_{11} + \max\{\mu_1(q_{12} - q_{11}), -\mu_3 q_{13}\} + \max\{\mu_2(q_{13} - q_{12}), 0\} + \gamma \le 0 \tag{20}$$

$$\lambda q_{12} + \mu_2(q_{23} - q_{22}) + \max\{\mu_1(q_{22} - q_{12}), -\mu_3 q_{23}, 0\} + \gamma \le 0 \tag{21}$$

$$\lambda q_{13} + \max\{\mu_1(q_{23} - q_{13}), -\mu_3 q_{33}\} + \max\{\mu_2(q_{33} - q_{23}), 0\} + \gamma \leq 0 \tag{22}$$

$$\gamma \leq 1 \tag{23}$$

$$q_{11}, q_{12}, q_{13}, q_{22}, q_{23}, q_{33} \geq 0. \tag{24}$$

Let $\rho_i := \frac{\lambda}{\mu_i}$, and consider $\rho_1 = \rho_3$. Figure 5 plots the value of the linear program as a function of $2\rho_1$ and ρ_2. It shows that there is a region where the value of the linear program is 0, and for such values of $\rho_1, \rho_2, \rho_3 (= \rho_1)$, no conclusion regarding stability or instability can be drawn.

Now consider the LBFS policy. The corresponding linear program is the same as (19-24), with exception that (22) is changed to:

$$\lambda q_{13} - \mu_3 q_{33} + \max\{\mu_2(q_{33} - q_{23}), 0\} + \gamma \leq 0. \tag{25}$$

The plot of its value as a function of $2\rho_1$ and ρ_2 is given in Figure 6. It shows that the system is stable for *all* values of $2\rho_1 < 1$ and $\rho_2 < 1$. Hence, for $\rho_1 = \rho_3$, the LBFS policy is stable in the entire *capacity* region.

Now turn to the FBFS policy. Its linear program is also almost the same as (19-24), except that (20) is changed to the following:

$$\lambda q_{11} + \mu_1(q_{12} - q_{11}) + \max\{\mu_2(q_{13} - q_{12}), 0\} + \gamma \leq 0. \tag{26}$$

The plot of its value is the same as that of LBFS, as shown in Figure 6. Hence it is also stable in the entire stability region, when $\rho_1 = \rho_3$.

11 A Nonlinear Programming Characterization

Note that every symmetric positive semidefinite Q is copositive. Thus, in our stability tests, we could use the class of symmetric positive semidefinite matrices. Every such Q possesses a square root A, i.e., $Q = A^T A$. Hence, one may search over the unrestricted space of a_{ij}'s, rather than the space of q_{ij}'s. This yields the following Theorem.

Theorem 5: A Nonlinear Programming Characterization. *Consider the basic, open, re-entrant line. Consider the nonlinear program:*

$$\text{Max} \quad \gamma$$

subject to all the constraints of Corollary 1, except that every q_{ij} *is replaced by* $\sum_{k=1}^{L} a_{ki}a_{kj}$*, and the non–negativity constraint on the* q_{ij}*'s is removed. If this nonlinear program has value 1, then, every non-idling, non–interruptive, scheduling policy is stable in the mean. Moreover, then, every non–idling, stationary policy gives rise to a Markov chain, with a geometrically converging exponential moment.*

Also, one can extend this Theorem to search over convex combinations of a non–negative matrix, and a positive semidefinite matrix. This yields the following Theorem.

Theorem 6: A More General Nonlinear Programming Characterization. *Consider the basic, open, re-entrant line. Consider the nonlinear program:*

$$\text{Max} \quad \gamma$$

subject to all the constraints of Corollary 1, except that every q_{ij} *is replaced by* $q'_{ij} + \sum_{k=1}^{L} a_{ki}a_{kj}$*, and the non–negativity constraint on the* q_{ij}*'s is replaced by non–negativity constraints on the* q'_{ij}*'s, while the* a_{ij}*'s are unrestricted. If this nonlinear program has value 1, then, every non-idling, non–interruptive, scheduling policy is stable in the mean. Moreover, then, every non–idling, stationary policy gives rise to a Markov chain, with a geometrically converging exponential moment.*

Both these theorems can be extended in the same ways as Theorem 1, to treat various kinds of systems and scheduling policies.

To go beyond Theorems 5 and 6, and obtain the most powerful test obtainable through our approach, one could simply check whether the value of the Linear Program in Theorem 3 is 1, *without imposing any sign restrictions on* q_{ij}, i.e., after removing the constraints $q_{ij} \geq 0$. If the value is indeed 1, one can then test whether the obtained Q is copositive, using an algorithm as in [25]. However, as noted in Section 3, the test of copositivity is NP-Complete, and may therefore be computationally complex for systems of large size.

12 More General Queueing Networks

Consider a queueing network with S machines, and L buffers. Let us suppose every buffer b_i has an exogenous Poisson arrival process of rate λ_i. A part leaving buffer b_i goes to buffer b_j with probability p_{ij}, and leaves the system with probability $\left(1 - \sum_{j=1}^{L} p_{ij}\right)$. The service time of parts in buffer b_i is exponential with mean $\frac{1}{\mu_i}$. Let us rescale time so that $\sum_{i=1}^{L} \lambda_i + \sum_{j=1}^{L} \mu_j = 1$. The state transition diagram is shown in Figure 7. Hence

$$\begin{aligned} \mathrm{E}(x^T(\tau_{n+1})Qx(\tau_{n+1}) \mid \mathcal{F}_{\tau_n}) &= \sum_{i=1}^{L} \lambda_i (x(\tau_n) + e_i)^T Q (x(\tau_n) + e_i) \\ &+ \sum_{i=1}^{L} \mu_i w_i(\tau_n) \sum_{j=1}^{L} p_{ij} (x(\tau_n) - e_i + e_j)^T Q (x(\tau_n) - e_i + e_j) \\ &+ \sum_{i=1}^{L} \mu_i w_i(\tau_n) \left(1 - \sum_{j=1}^{L} p_{ij}\right) (x(\tau_n) - e_i)^T Q (x(\tau_n) - e_i) \\ &+ \sum_{i=1}^{L} \mu_i (1 - w_i(\tau_n)) x^T(\tau_n) Q x(\tau_n). \end{aligned}$$

Proceeding as in Section 4, one may obtain the following theorem.

Theorem 7: Stability of a general network. *Consider the more general queueing network above, operated under any non-idling, non–interruptive policy. Then it is stable in the mean, provided there exists a symmetric copositive Q that satisfies the following inequalities for $j = 1, 2, \ldots, L$:*

$$\sum_{i=1}^{L} \lambda_i q_{i,j} + \max_{\{i:\sigma(i)=\sigma(j)\}} \mu_i (-q_{ij} + \sum_{k=1}^{L} p_{ik} q_{kj}) + \sum_{\{\sigma' : \sigma' \neq \sigma(j)\}} \left[\max_{\{i:\sigma(i)=\sigma'\}} \mu_i (-q_{ij} + \sum_{k=1}^{L} p_{ik} q_{kj} \right]^+ < 0.$$

This can also be written as a linear program if we restrict $q_{ij} \geq 0$ for all i, j, or as a nonlinear program if we take $Q = A^T A + [q'_{ij}]$, with non–negative q'_{ij}'s. Moreover, if the policy is stationary, then the system is a Markov chain with a geometrically converging exponential moment.

13 The Rybko–Stolyar Example

Consider the system shown in Figure 8. The arrivals to buffers b_1 and b_3 are Poisson of rate λ. The service times are all exponentially distributed, with mean $1/\mu_1$ at buffers b_1 and b_3, and mean $1/\mu_2$ at buffers b_2 and b_4. Consider the buffer priority policy with ordering $\{b_4, b_2, b_3, b_1\}$, i.e., with priority given to buffers earlier in this list. Let us define $\rho_1 := \lambda/\mu_1$, and $\rho_2 := \lambda/\mu_2$.

Rybko and Stolyar [4] have shown that this system is unstable for $\lambda = 1$, if $\rho_2 > 1/2$ and $\rho_1 > 0$, even if the service time requirements meet the capacity condition, $1/\mu_1 + 1/\mu_2 < 1$.

The following linear program tests the stability of the system for all ρ_1 and ρ_2.

$$\text{Max} \quad \gamma$$

subject to the constraints:

$$\text{Max}[\lambda q_{11} + \lambda q_{13} - \mu_1 q_{11} + \mu_1 q_{12}, \lambda q_{11} + \lambda q_{13} - \mu_4 q_{14}] + \text{Max}[0, -\mu_3 q_{13} + \mu_3 q_{14}] \leq \gamma$$

$$\lambda q_{12} + \lambda q_{23} - \mu_2 q_{22} + \text{Max}[0, -\mu_1 q_{12} + \mu_1 q_{22}] \leq \gamma$$

$$\text{Max}[\lambda q_{13} + \lambda q_{33} - \mu_2 q_{23}, \lambda q_{13} + \lambda q_{33} - \mu_3 q_{33} + \mu_3 q_{34}] + \text{Max}[0, -\mu_1 q_{13} + \mu_1 q_{23}] \leq \gamma$$

$$\lambda q_{14} + \lambda_{34} - \mu_4 q_{44} + \text{Max}[0, -\mu_3 q_{34} + \mu_3 q_{44}] \leq \gamma$$

$$\gamma \leq 1$$

$$q_{ij} \geq 0, \gamma \geq 0.$$

Figure 9 plots the value of the linear program for $0 < \rho_1 < 1$ and $0 < \rho_2 < 1$. (The region $\rho_1 + \rho_2 \geq 1$ should be disregarded as it lies outside the capacity region.)

The following points are salient. First, the value of the linear program is 0 in the region $\rho_2 > 1/2$, and thus non–contradictory with Rybko and Stolyar's result. Second, as shown in Figure 10, for most of the rest of the capacity region, the system is stable, since the value of the linear program is 1. However, there is a small region where the value of the linear program is 0; thus the stability remains unresolved there.

14 Concluding Remarks

We have provided here a programmatic procedure for establishing stability of queueing networks and scheduling policies.

There are several interesting questions which arise. First, it would be useful to study the structure of the linear or nonlinear programs, and thus directly establish the stability of policies. We have done so analytically for the Example of Section 7. Second, in all the examples tested by us, any Q giving a negative drift was always found to be a non–negative matrix. It would be useful to determine whether there exists an example of a system where Q is copositive but not non–negative. This should show that the more powerful tests of Section 11 are in fact valuable. Third, it would be useful to implement a multi–step drift version of the above results. Finally, it would be useful to carry out a similar development for "instability" results, as in Fayolle [16] and Coffman et al [20].

Acknowledgments The authors are grateful to Doug Down, Seetharam Gowda and John Tsitsiklis for useful comments.

References

[1] S. H. Lu and P. R. Kumar, "Distributed scheduling based on due dates and buffer priorities," *IEEE Transactions on Automatic Control*, vol. 36, pp. 1406–1416, December 1991.

[2] P. R. Kumar and T. I. Seidman, "Dynamic instabilities and stabilization methods in distributed real-time scheduling of manufacturing systems," *IEEE Trans. Automat. Control*, vol. AC-35, pp. 289–298, March 1990.

[3] C. J. Chase and P. J. Ramadge, "On the real time control of flexible manufacturing systems," in *Proc. IEEE 28th Conf. on Decision and Control*, (Tampa, FL), pp. 2026–2027, 1989.

[4] A. N. Rybko and A. L. Stolyar, "On the ergodicity of stochastic processes describing open queueing networks," *Problemy Peredachi Informatsii*, vol. 28, pp. 2–26, 1991.

[5] T. I. Seidman, ""First Come First Serve" is Unstable!," tech. rep., University of Maryland Baltimore County, 1993.

[6] M. Bramson, "Instability of FIFO queueing networks," tech. rep., Mathematics Department, University of Wisconsin, Madison, WI, 1993.

[7] J. M. Harrison, "Brownian models of heterogeneous customer populations," in *Stochastic Differential Systems, Stochastic Control Theory and Applications, IMA Volumes in Mathematics and its Applications* (W. Fleming and P. L. Lions, eds.), ch. 10, pp. 147–186, Springer-Verlag, 1988.

[8] J. M. Harrison and L. M. Wein, "Scheduling networks of queues: Heavy traffic analysis of a two-station closed network," *Operations Research*, vol. 38, no. 6, pp. 1052–1064, 1990.

[9] M. I. Reiman, "Open queueing networks in heavy traffic," *Mathematics of Operations Research*, vol. 9, pp. 441–458, 1984.

[10] M. I. Reiman, "A multiclass queue in heavy traffic," *Advances in Applied Probability*, vol. 20, pp. 179–207, 1988.

[11] W. P. Peterson, "A heavy traffic limit theorem for networks of queues with multiple customer types," *Mathematics of Operations Research*, vol. 16, pp. 90–118, 1991.

[12] J. G. Dai and Y. Wang, "Nonexistence of Brownian models of certain multiclass queueing networks," *Queueing Systems: Theory and Applications: Special Issue on Queueing Networks*, vol. 13, pp. 41–46, May 1993.

[13] W. Whitt, "An interesting example of a multiclass queueing netwrok," tech. rep., AT&T Bell Laboratories, Murray Hill, NJ, 1992.

[14] J. G. Dai and V. Nguyen, "On the convergence of multiclass queueing networks in heavy traffic," tech. rep., Georgia Institute of Technology and Massachusetts Institute of Technology, 1993. To appear in *The Annals of Applied Probability*.

[15] J. F. C. Kingman, "The ergodic behaviour of random walks," *Biometrika*, vol. 48, pp. 391–396, 1961.

[16] G. Fayolle, "On random walks arising in queueing systems: Ergodicity and transience via quadratic forms as Lyapounov functions - part I," *Queueing Systems*, vol. 5, pp. 167–184, 1989.

[17] G. Fayolle, V. A. Malyshev, M. V. Mensikov, and A. F. Sidorenko, "Lyapunov functions for Jackson networks," Rapports de Recherche No. 1380, INRIA, Rocquencourt, France, 1991.

[18] S. P. Meyn and D. Down, "Stability of generalized Jackson networks." To appear in *Annals Applied Prob.*, 1993.

[19] J. R. Jackson, "Jobshop-like queueing systems," *Management Science*, vol. 10, pp. 131–142, 1963.

[20] E. G. Coffman, Jr and D. S. Johnson and P. W. Shor and R. R. Weber, "Markov chains, computer proofs, and average-case analysis of best fit bin packing." Accepted for STOC 93, 1993.

[21] D. Bertsimas and I. Ch. Paschalidis and J. N. Tsitsiklis, "Optimization of multiclass queueing networks: Polyhedral and nonlinear characterizations of achievable performance." Laboratory for Information and Decision Systems and Operations Research Center, M. I. T., December 1992.

[22] S. Kumar and P. R. Kumar, "Performance bounds for queueing networks and scheduling policies," technical report, Coordinated Science Laboratory, University of Illinois, Urbana, IL, 1992. To appear in *IEEE Transactions on Automatic Control*, August 1994.

[23] P. R. Kumar, "Re–entrant lines," *Queueing Systems: Theory and Applications: Special Issue on Queueing Networks*, vol. 13, pp. 87–110, May 1993.

[24] R. W. Cottle, G. J. Habetler, and C. E. Lemke, "On classes of copositive matrices," *Linear Algebra and Its Applications*, vol. 3, pp. 295–310, 1970.

[25] L.-E. Andersson, G. Chang, and T. Elfving, "Criteria for copositive matrices and nonnegative Bezier patches," Tech. Rep. LiTH-MAT-R-93-27, Linkoping University and University of Science and Technology, China, August 1993.

[26] K. G. Murty and S. N. Kabadi, "Some NP-complete problems in quadratic and nonlinear programming," *Mathematical Programming*, vol. 39, pp. 117–129, 1987.

[27] S. Lippman, "Applying a new device in the optimization of exponential queueing systems," *Operations Research*, vol. 23, pp. 687–710, 1975.

[28] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*. Control and Communication in Engineering, London: Springer-Verlag, 1993.

[29] S. P. Meyn and R. L. Tweedie, "Computable bounds for convergence rates of Markov chains." 1993.

[30] F. P. Kelly, *Reversibility and Stochastic Networks*. New York, NY: John Wiley and Sons, 1979.