PERFORMANCE OF MULTICLASS MARKOVIAN QUEUEING NETWORKS VIA PIECEWISE LINEAR LYAPUNOV FUNCTIONS[1]

By Dimitris Bertsimas, David Gamarnik and John N. Tsitsiklis

Massachusetts Institute of Technology, IBM T. J. Watson Research Center and Massachusetts Institute of Technology

We study the distribution of steady-state queue lengths in multiclass queueing networks under a stable policy. We propose a general methodology based on Lyapunov functions for the performance analysis of infinite state Markov chains and apply it specifically to Markovian multiclass queueing networks. We establish a deeper connection between stability and performance of such networks by showing that if there exist linear or piecewise linear Lyapunov functions that show stability, then these Lyapunov functions can be used to establish geometric-type lower and upper bounds on the tail probabilities, and thus bounds on the expectation of the queue lengths. As an example of our results, for a reentrant line queueing network with two processing stations operating under a work-conserving policy, we show that $E[L] = O(\frac{1}{(1-\rho^*)^2})$, where L is the total number of customers in the system, and ρ^* is the maximal actual or virtual traffic intensity in the network. In a Markovian setting, this extends a recent result by Dai and Vande Vate, which states that a reentrant line queueing network with two stations is globally stable if $\rho^* < 1$. We also present several results on the performance of multiclass queueing networks operating under general Markovian and, in particular, priority policies. The results in this paper are the first that establish explicit geometric-type upper and lower bounds on tail probabilities of queue lengths for networks of such generality. Previous results provide numerical bounds and only on the expectation, not the distribution, of queue lengths.

1. Introduction. Queueing networks are used to model manufacturing, communication and computer systems, and much recent research has focused on networks with multiple customer classes. In multiclass queueing networks, the customers served at the same station have in general different service requirements and follow different paths through the network. Such networks are used to model, for example, wafer fabrication facilities, in which there is a single stream of jobs arriving into a production floor. Jobs follow a deterministic route and revisit the same station multiple times (see Figure 1). Multiclass queueing networks of this type are called *reentrant line* queueing networks (see [17, 9]).
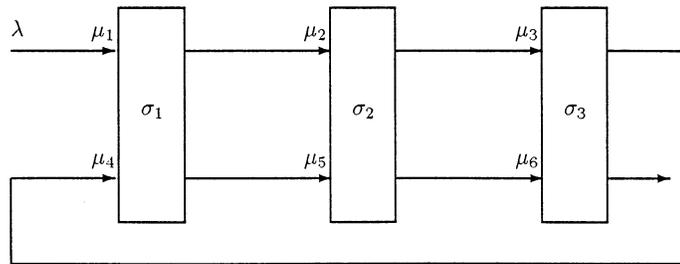
FIG. 1. *Re-entrant line queueing network.*

The focus of this paper is performance analysis of multiclass queueing networks. Specifically, we are interested in estimating the steady-state queue lengths in the network, when interarrival and service times are exponentially distributed, assuming a stable scheduling policy is used, which brings the system to steady-state.

The performance of queueing networks is largely an open research area. Some of the earlier and classical results include product form probability distributions for Jackson and BCMP-type networks (see [13]). It was realized, however, that the presence of multiple classes does not allow, in general, for a product form distribution even if the interarrival and service times have exponential distributions and the first-in-first-out policy is used. Several papers ([4, 18, 19, 15]) have analyzed the performance of multiclass queueing networks using quadratic Lyapunov functions. A certain linear program is constructed, which provides numerical bounds on the achievable performance region. The performance results obtained using quadratic Lyapunov functions were later analyzed and extended in a simpler and more intuitive way, using conservation laws ([3]).

The performance analysis of multiclass queueing networks is highly nontrivial, since it is at least as hard as the stability problem for which no general conditions are available. It is known that the natural load condition

$$\rho_\sigma < 1$$

for each station σ is necessary, but not sufficient, for stability; a variety of counterexamples have been constructed in [23, 21, 5, 24, 7]. Sufficient conditions for stability have been found using Lyapunov functions [9, 10]. Furthermore, fluid models were found to be a very useful tool for stability analysis. Dai's theorem [6] shows that the stability of a fluid model implies stability of a corresponding stochastic model. A complete characterization of fluid networks with two stations which are stable under any work-conserving policy ("globally stable") was obtained by Bertsimas, Gamarnik and Tsitsiklis [2] and subsequently by Dai and Vande Vate [8]. The second work used a very intuitive notion of virtual stations to explain instability in networks with two stations. Both works ([2] and [8]) prove that the existence of a piecewise linear

Lyapunov function is both necessary and sufficient for global stability of fluid networks with two stations.

1.1. *Our results*. The goal of this paper is to turn some of the stability analysis tools into useful performance analysis tools. We will show how linear and piecewise linear Lyapunov functions and virtual stations can be used to obtain upper and lower bounds on the steady-state queue lengths. For many examples considered in this paper, the upper bounds are finite if and only if the network is stable. Our contributions are summarized as follows. We start in Section 3 with an analysis of countably infinite Markov chains. We show that if there exists a Lyapunov function proving the stability of the Markov chain, then certain computable upper and lower bounds hold on the steady-state queue length probability distribution as well as on its expectation. We then apply this methodology, in Sections 4 and 5, to the performance analysis of multiclass queueing networks with exponentially distributed interarrival and service times. Specifically:

1. We show how linear and piecewise linear Lyapunov functions can be used to obtain lower and upper bounds, respectively, on the steady-state queue lengths. The lower bounds we obtain are explicit, while the upper bounds are numerical and depend on the solutions of a certain linear program which generates the Lyapunov function. Such linear programs were introduced in [9] and [10] for the purposes of stability analysis.

2. We use the notion of a *virtual station*, introduced in [8]. This showed that in networks with two stations, some priority policies lead to certain groups V of customer classes, called virtual stations, which cannot be served simultaneously. As a result, if the corresponding *virtual traffic intensity* $\rho(V) \equiv \sum_{i \in V} \rho_i$ is bigger than one, then the network is unstable. Here ρ_i stands for the naturally defined load factor of class i. We prove a matching performance result: for networks with two stations, if V is a virtual station, with the corresponding virtual traffic intensity $\rho(V)$, and if priority is given to the classes in V over the classes not in V then

$$\mathrm{P}\left\{L \geq \frac{1}{2}m\right\} \geq \left(\frac{\rho(V)}{2 - \rho(V)}\right)^m,$$

$$\mathrm{E}[L] \geq \frac{\rho(V)}{4(1 - \rho(V))},$$

where L is the total number of customers in the network. These lower bounds are extended to networks with more than two stations.

3. It was also proved in [8] that queueing networks with two stations are globally stable if the maximum of all actual and virtual traffic intensities, denoted by ρ^*, is less than 1 for the original network and for a certain set of subnetworks. Also if $\rho^* > 1$, then the corresponding fluid network is not globally stable. Whether this holds true for stochastic Markovian networks is not known.

We show that ρ^* is a fundamental performance parameter. For reentrant line networks with two stations, we show that if $\rho^* < 1$, then the following upper bound holds under any work-conserving policy:

$$\mathrm{E}[L] \leq \frac{C}{(1-\rho^*)^2},$$

where C is some constant, expressed explicitly in terms of the parameters of the network. An important implication of this result is that the performance region (the set of vectors of expected queue lengths obtained under different work-conserving scheduling policies) is bounded if and only if the corresponding fluid network is globally stable.

Our results show a deeper connection between stability and performance of multiclass queueing networks. Also the results in this paper are the first ones that use linear and piecewise linear Lyapunov functions for performance analysis. Previous methods for performance analysis have used quadratic Lyapunov functions, which have certain limitations. In particular, an example of a globally stable queueing network with two stations was constructed in [10] for which the quadratic Lyapunov function method leads to an infinite (inconclusive) upper bound, yet a piecewise linear Lyapunov function gives a finite upper bound. The methods developed here, on the other hand, match the sharpest known stability condition $\rho^* < 1$. The second limitation of quadratic Lyapunov functions is that the bounds constructed are in most cases only numerical and hold only for the expectations of queue lengths. In contrast, we provide bounds on the distribution of steady-state queue lengths, proving exponential decay of the tail probabilities.

2. Queueing model and assumptions.

2.1. *Multiclass Markovian queueing network.*   We consider a network consisting of J single server stations, which are denoted by σ_j, $j = 1, 2, \ldots, J$. The network includes I types of customers, where customers of type $i = 1$, $2, \ldots, I$ arrive to the network from an exogenous source. The arrival process corresponding to type i is assumed to be an independent Poisson process with rate λ_i. Let $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_I)$ denote the vector of arrival rates and let $\lambda_{\min} = \min_i\{\lambda_i\}$. Without loss of generality, we assume that $\lambda_{\min} > 0$. Similarly, we define $\lambda_{\max} = \max_i\{\lambda_i\}$. Customers of type i go through J_i stages, each of which corresponds to a service completion on a particular station. We denote these stations by $\sigma_{i,1}, \sigma_{i,2}, \ldots, \sigma_{i,J_i}$. The processing time of a type i customer at station $\sigma_{i,k}$, $k = 1, 2, \ldots, J_i$, is assumed to be exponentially distributed with rate $\mu_{i,k}$ and is independent from the processing times of all other stages of this type, from the processing times of the other types and from the interarrival times. We let $\boldsymbol{\mu} = (\mu_{i,k})_{1 \leq i \leq I, 1 \leq k \leq J_i}$ denote the vector of service rates. Customers of type i receiving service at station $\sigma_{i,k}$ are called class (i, k) customers. Let $N = \sum_{i=1}^{I} J_i$ be the total number of classes. For

convenience, we will also identify every station σ_j with the set of classes associated with this station. Let $C_{i,k} = j$ if class (i, k) customers are served at station σ_j. For $k \geq J_i$ we let $C_{i,k} = 0$. Let $\mathbf{C}$ denote the corresponding $I \times J_{\max}$ matrix, where $J_{\max} = \max_i J_i$. The matrix $\mathbf{C}$ defines the topology of the network. We assume that the buffers at each station have infinite capacity and no customers renege from the queue before receiving service.

A queueing network of the form just described is called a Markovian multiclass queueing network with deterministic routing or a multitype queueing network. Throughout the paper we consider only networks of this type. The parameters $\boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{C}$ constitute the primary parameters of the network and we denote the network by $(\boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{C})$.

For each class (i, k), we let $\rho_{i,k} = \lambda_i / \mu_{i,k}$ be the *nominal load* of this class. For each station σ_j, $j = 1, 2, \ldots, J$, we define the *nominal load (traffic intensity)* as

$$(1) \qquad \rho_{\sigma_j} \equiv \sum_{(i,k) \in \sigma_j} \rho_{i,k}.$$

The evolution of a queueing network is fully specified only when a scheduling discipline and an initial state is given. The scheduling discipline (policy) describes which customers (if any) are served at any moment at each station. Within each class, the customers are served in first-in-first-out (FIFO) fashion. Therefore, the service discipline only specifies which *customer type* is served at any given moment. We will assume throughout the paper that the scheduling policies implemented are *Markovian*, namely, scheduling decisions are purely a function of the system state, which in our case is the vector of all queue lengths. We also allow preemption. For example, preemptive priority policies are Markovian. Many important policies are not Markovian, for example FIFO or head-of-the-line-processor-sharing. We leave these policies out from the discussion in this paper, although we believe that the results hold for them as well. We will be considering mostly *work-conserving* policies: each processing station is required to work on some customer, if there are any present at this station.

Any preemptive policy w satisfying the Markovian assumption can be described by a function $w: Z_+^N \to \{0, 1\}^N$ where for any $\mathbf{q} \in Z_+^N$, the (i, k) component of the vector $\mathbf{w}(\mathbf{q})$ is 1 if station σ_j which contains class (i, k) works on a customer of class (i, k), and is zero, otherwise. Of course, $w_{i,k}(\mathbf{q}) = 1$ only if $q_{i,k} > 0$, and for each station σ_j,

$$(2) \qquad \sum_{(i,k) \in \sigma_j} w_{i,k}(\mathbf{q}) \leq 1.$$

Given a multiclass queueing network $(\boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{C})$ and some scheduling policy, we let $\mathbf{Q}(t) = (Q_{i,k}(t))_{1 \leq i \leq I, 1 \leq k \leq J_i}$ denote the vector of queue lengths at time t. Our focus is on estimating the distribution of the random vector $\mathbf{Q}(t)$ in steady-state. A necessary condition for the existence of a steady-state is the

load condition

$$(3) \qquad\qquad\qquad \rho_{\sigma_j} < 1$$

for each $j = 1, 2, \ldots, J$.

2.2. *Embedded Markov chains and uniformization.* Instead of analyzing the continuous time process $\mathbf{Q}(t)$, we will build a discrete time analogue, which has the same steady-state behavior using the standard method of *uniformization* (see [20]). We rescale the parameters, so that $\sum_i \lambda_i + \sum_{i,k} \mu_{i,k} = 1$ and consider a superposition of $I + \sum_{i=1}^{I} J_i$ Poisson processes with rates $\lambda_i, \mu_{i,k}$, respectively. The arrivals of the first I Poisson processes correspond to external arrivals into the network. The arrivals of the process with rate $\mu_{i,k}$ correspond to service completions of class (i, k) customers, if a server actually worked on a class (i, k) customer, or they correspond to "imaginary" service completions of an "imaginary" customer, if the server was idle or worked on customers from other classes. Let $\tau_s, s = 1, 2, \ldots$ be the sequence of event times for this superposition of Poisson processes. Then, as a result of this construction and the Markovian policy assumption, $\mathbf{Q}(\tau_s)$ is a discrete time Markov chain with the same steady-state distribution as $\mathbf{Q}(t)$ (assuming it exists).

We can specify the transitions of the Markov chain $\mathbf{Q}(\tau_s)$ as follows. For each class (i, k) let $\mathbf{e}_{i,k}$ be an N-dimensional unit vector with the (i, k) component equal to 1 and all other components equal to zero. We adopt the convention $\mathbf{e}_{i,0} = \mathbf{e}_{i,J_i+1} = 0$ for each i. The following proposition holds.

PROPOSITION 1. *Given a multiclass queueing network $(\boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{C})$ rescaled so that $\sum_i \lambda_i + \sum_{i,k} \mu_{i,k} = 1$ and given a Markovian policy w, the transition probabilities of the corresponding embedded Markov chain $\mathbf{Q}(\tau_s), s = 0, 1, 2, \ldots$ are given by*

$$(4) \quad \mathbf{Q}(\tau_{s+1}) = \begin{cases} \mathbf{Q}(\tau_s) + \mathbf{e}_{i,1}, & \text{with probability } \lambda_i, \\ \mathbf{Q}(\tau_s) - \mathbf{e}_{i,k} + \mathbf{e}_{i,k+1}, & \text{with probability } \mu_{i,k} w_{i,k}(\mathbf{Q}(\tau_s)), \\ \mathbf{Q}(\tau_s), & \text{with probability } \sum_{i,k} \mu_{i,k}(1 - w_{i,k}\mathbf{Q}(\tau_s)). \end{cases}$$

PROOF. Note that the change $\mathbf{Q}(\tau_{s+1}) - \mathbf{Q}(\tau_s)$ of the embedded Markov chain corresponds to either an arrival of a type i customer, or to an "actual" service completion of a class (i, k) customer and transition to the next stage $k + 1$. The first event has a probability λ_i and corresponds to a change $\mathbf{e}_{i,1}$. The second event has probability $\mu_{i,k} w_{i,k}(\mathbf{Q}(\tau_s))$ and corresponds to a change $\mathbf{e}_{i,k+1} - \mathbf{e}_{i,k}$. $\square$

DEFINITION 1. A scheduling policy w is defined to be stable if the corresponding embedded Markov chain $\mathbf{Q}(\tau_s), s = 1, 2, \ldots$, admits a stationary probability distribution $\pi = \pi(w)$ satisfying

$$(5) \qquad\qquad\qquad \sum_{i,k} \mathrm{E}[Q_{i,k}(\tau_s)] < \infty.$$

A queueing network is defined to be globally stable if every work-conserving Markovian policy is stable.

If a stationary distribution π of $\mathbf{Q}(\tau_s)$ exists, then by uniformization and by aperiodicity of our continuous time Markov chain,

$$(6) \qquad \lim_{t \to \infty} \mathrm{P}\{\mathbf{Q}(t) = \mathbf{q}\} = \mathrm{P}_\pi\{\mathbf{Q}(\tau_s) = \mathbf{q}\}.$$

Thus, for performance analysis purposes, we may concentrate on the embedded chain $\mathbf{Q}(\tau_s)$.

Throughout the paper we use standard notations $O(\cdot)$, $\Omega(\cdot)$, $\Theta(\cdot)$ in the following sense. If functions $f(s), g(s) \to \infty$ when $s \to s_0$ for some $s_0 \in [-\infty, +\infty]$, then $g = O(f)$ $(g = \Omega(f))$ means that for some fixed constant $c > 0$, $g(s) \le cf(s)$ $(g(s) \ge cf(s))$ for sufficiently large s. If both $g = O(f)$, $f = O(g)$, then we will write $g = \Theta(f)$.

3. Infinite Markov chains and Lyapunov functions. In this section, we develop a general technique for steady-state analysis of infinite Markov chains with countably many states using Lyapunov functions.

Let $\mathbf{X}(t), t = 0, 1, 2, \dots$, be a discrete time, discrete state Markov chain which takes values in some countable set $\mathscr{X}$. The transitions occur at integer times $t = 0, 1, 2, \dots$. For any two vector $\mathbf{x}, \mathbf{x}' \in \mathscr{X}$, let $p(\mathbf{x}, \mathbf{x}')$ denote the transition probabilities

$$p(\mathbf{x}, \mathbf{x}') \equiv \mathrm{P}\{\mathbf{X}(t+1) = \mathbf{x}' | \mathbf{X}(t) = \mathbf{x}\}.$$

If a stationary probability distribution π on the state space $\mathscr{X}$ exists, it satisfies

$$\sum_{\mathbf{x} \in \mathscr{X}} \pi(\mathbf{x}) = 1,$$

and for all $\mathbf{x} \in \mathscr{X}$,

$$(7) \qquad \pi(\mathbf{x}) = \sum_{\mathbf{x}' \in \mathscr{X}} \pi(\mathbf{x}') p(\mathbf{x}', \mathbf{x}).$$

The existence of a stationary distribution is usually established by constructing a certain Lyapunov function. For a survey of Lyapunov methods for stability analysis of Markov chains, see [22]. We now introduce the definitions of Lyapunov and lower Lyapunov functions. The goal is to use Lyapunov functions for the performance analysis of Markov chains, assuming a priori that the Markov chain is stable. The notion of a lower Lyapunov function is introduced exclusively as a means of getting the lower bounds on the stationary distribution of a Markov chain. In subsequent sections, we apply the results here to the embedded Markov chain of a multiclass queueing network.

DEFINITION 2. A nonnegative function

$$\Phi: \mathscr{X} \to \Re_+$$

is said to be a Lyapunov function if there exist some $\gamma > 0$ and $B \geq 0$, such that for any $t = 1, 2, \ldots$ and any $\mathbf{x} \in \mathscr{X}$, with $\Phi(\mathbf{x}) > B$,

$$(8) \qquad \mathrm{E}\big[\Phi(\mathbf{X}(t+1))|\mathbf{X}(t) = \mathbf{x}\big] \leq \Phi(\mathbf{x}) - \gamma.$$

Also a nonnegative function

$$(9) \qquad \Phi\colon \mathscr{X} \to \Re_+$$

is said to be a lower Lyapunov function if there exists some $\gamma > 0$, such that for any $t = 1, 2, \ldots$ and any $\mathbf{x} \in \mathscr{X}$, with $\Phi(\mathbf{x}) > 0$,

$$\mathrm{E}\big[\Phi(\mathbf{X}(t+1))|\mathbf{X}(t) = \mathbf{x}\big] \geq \Phi(\mathbf{x}) - \gamma.$$

REMARKS. (i) We refer to the terms γ and B as drift and exception parameters, respectively.

(ii) We could also introduce an exception parameter B for the lower Lyapunov function, but it is not required for the examples in this paper.

We assume that the Markov chain $\mathbf{X}(t)$ is positive recurrent, and we denote by π the corresponding stationary distribution. Namely, $\pi(\mathbf{x})$ is the steady-state probability $\mathrm{P}_\pi\{\mathbf{X}(t) = \mathbf{x}\}$ that the chain is in a certain state $\mathbf{x} \in \mathscr{X}$. Also, we denote by $\mathrm{E}_\pi[\cdot]$ the expectation with respect to the probability distribution π. For a given function $\Phi\colon \mathscr{X} \to \Re_+$, let

$$(10) \qquad \nu_{\max} \equiv \sup_{\mathbf{x}, \mathbf{x}' \in \mathscr{X}\colon\; p(\mathbf{x}, \mathbf{x}') > 0} |\Phi(\mathbf{x}') - \Phi(\mathbf{x})|$$

and

$$(11) \qquad \nu_{\min} \equiv \inf_{\mathbf{x}, \mathbf{x}' \in \mathscr{X}\colon\; p(\mathbf{x}, \mathbf{x}') > 0,\, \Phi(\mathbf{x}) < \Phi(\mathbf{x}')} (\Phi(\mathbf{x}') - \Phi(\mathbf{x})).$$

Namely, $\nu_{\max}$ is the largest possible change of the function Φ during an arbitrary transition, and $\nu_{\min}$ is the smallest possible increase of the function Φ. Also let

$$(12) \qquad p_{\max} = \sup_{\mathbf{x} \in \mathscr{X}} \sum_{\mathbf{x}' \in \mathscr{X},\, \Phi(\mathbf{x}) < \Phi(\mathbf{x}')} p(\mathbf{x}, \mathbf{x}')$$

and

$$(13) \qquad p_{\min} = \inf_{\mathbf{x} \in \mathscr{X}} \sum_{\mathbf{x}' \in \mathscr{X},\, \Phi(\mathbf{x}) < \Phi(\mathbf{x}')} p(\mathbf{x}, \mathbf{x}').$$

Namely, $p_{\max}$ and $p_{\min}$ are tight upper and lower bounds on the probability that the value of Φ is increasing during an arbitrary transition. In this paper, we will be interested in Lyapunov functions with finite $\nu_{\max}$, and lower Lyapunov functions with positive $\nu_{\min}$ and $p_{\min}$. We need the following lemma, the proof of which can be found in Appendix A.

LEMMA 1.   *Consider a Markov chain* $\mathbf{X}(t)$ *with stationary probability distribution* π*, and suppose that* Φ *is a Lyapunov function with drift* γ *and exception parameter* B*, such that* $\mathrm{E}_\pi[\Phi(\mathbf{X}(t)]$ *is finite. Then, for any* (*possibly negative*) $c \geq B - \nu_{\max}$,

$$(14) \qquad \mathrm{P}_\pi\big\{c + \nu_{\max} < \Phi(\mathbf{X}(t))\big\} \leq \frac{p_{\max}\nu_{\max}}{p_{\max}\nu_{\max} + \gamma} \mathrm{P}_\pi\big\{c - \nu_{\max} < \Phi(\mathbf{X}(t))\big\},$$

where $\nu_{\max}$ *and* $p_{\max}$ *are defined in* (10) *and* (12), *respectively. Also, if* Φ *is a lower Lyapunov function with drift* γ*, such that* $\mathrm{E}_\pi[\Phi(\mathbf{X}(t))]$ *is finite, then, for any* $c \geq 0$,

$$(15) \qquad \mathrm{P}_\pi\big\{c \leq \Phi(\mathbf{X}(t))\big\} \geq \frac{p_{\min}\nu_{\min}}{p_{\min}\nu_{\min} + 2\gamma} \mathrm{P}_\pi\bigg\{c - \frac{1}{2}\nu_{\min} \leq \Phi(\mathbf{X}(t))\bigg\},$$

where $\nu_{\min}$ *and* $p_{\min}$ *are defined in* (11) *and* (13), *respectively.*

This lemma shows how one can obtain a simple recurrence relation on the tail probabilities $\mathrm{P}_\pi\{c < \Phi(\mathbf{X}(t))\}$. We use this recurrence in the proof of the following result.

THEOREM 1.   *Consider a Markov chain* $\mathbf{X}(t)$ *with a stationary probability distribution* π *such that* $\mathrm{E}_\pi[\Phi(\mathbf{X}(t))] < \infty$.

(i) *If there exists a Lyapunov function* Φ *with drift* $\gamma > 0$*, and exception parameter* $B \geq 0$*, then for any* $m = 0, 1, 2, \ldots$,

$$(16) \qquad \mathrm{P}_\pi\big\{\Phi(\mathbf{X}(t)) > B + 2\nu_{\max}m\big\} \leq \left(\frac{p_{\max}\nu_{\max}}{p_{\max}\nu_{\max} + \gamma}\right)^{m+1}.$$

As a result,

$$(17) \qquad \mathrm{E}_\pi[\Phi(\mathbf{X}(t))] \leq B + \frac{2p_{\max}(\nu_{\max})^2}{\gamma}.$$

(ii) *If there exists a lower Lyapunov function* Φ *with drift* $\gamma > 0$*, then for any* $m = 0, 1, 2, \ldots$,

$$(18) \qquad \mathrm{P}_\pi\big\{\Phi(\mathbf{X}(t)) \geq (1/2)\nu_{\min}m\big\} \geq \left(\frac{(1/2)p_{\min}\nu_{\min}}{(1/2)p_{\min}\nu_{\min} + \gamma}\right)^{m}.$$

As a result,

$$(19) \qquad \mathrm{E}_\pi[\Phi(\mathbf{X}(t))] \geq \frac{p_{\min}(\nu_{\min})^2}{4\gamma}.$$

REMARK.   The bounds (16), (17) and (18), (19) are meaningful only if $\nu_{\max} < \infty$ (the Lyapunov function has uniformly bounded jumps) and $\nu_{\min}, p_{\min} > 0$, respectively.

PROOF. In order to prove (16), we let $c = B - \nu_{\max}$. By applying Lemma 1, we obtain

$$P_\pi\{B < \Phi(\mathbf{X}(t))\} \leq \frac{p_{\max}\nu_{\max}}{p_{\max}\nu_{\max} + \gamma} P_\pi\{B - 2\nu_{\max} < \Phi(\mathbf{X}(t))\}$$

$$\leq \frac{p_{\max}\nu_{\max}}{p_{\max}\nu_{\max} + \gamma}.$$

We continue similarly, using $c = B + \nu_{\max}, c = B + 3\nu_{\max}, c = B + 5\nu_{\max}, \dots$. By applying again Lemma 1, we obtain the needed upper bound on the tail distribution.

In order to prove (17), note that

$$E_\pi[\Phi(\mathbf{X}(t))] \leq B \cdot P_\pi\{\Phi(\mathbf{X}(t)) \leq B\} + \sum_{m=0}^{\infty} (B + 2\nu_{\max}(m+1))$$

$$\times P_\pi\{B + 2\nu_{\max}m < \Phi(\mathbf{X}(t)) \leq B + 2\nu_{\max}(m+1)\}$$

$$= B \cdot P_\pi\{\Phi(\mathbf{X}(t)) \leq B\}$$

$$+ B \sum_{m=0}^{\infty} P_\pi\{B + 2\nu_{\max}m < \Phi(\mathbf{X}(t)) \leq B + 2\nu_{\max}(m+1)\}$$

$$+ 2\nu_{\max} \sum_{m=0}^{\infty} (m+1)P_\pi\{B + 2\nu_{\max}m < \Phi(\mathbf{X}(t)) \leq B + 2\nu_{\max}(m+1)\}.$$

However,

$$\sum_{m=0}^{\infty} (m+1)P_\pi\{B + 2\nu_{\max}m < \Phi(\mathbf{X}(t)) \leq B + 2\nu_{\max}(m+1)\}$$

$$= \sum_{m=0}^{\infty} P_\pi\{B + 2\nu_{\max}m < \Phi(\mathbf{X}(t))\}.$$

Applying the bounds from (16), we obtain

$$E_\pi[\Phi(\mathbf{X}(t))] \leq B + 2\nu_{\max} \sum_{m=0}^{\infty} \left(\frac{p_{\max}\nu_{\max}}{p_{\max}\nu_{\max} + \gamma}\right)^{m+1}$$

$$= B + \frac{2p_{\max}(\nu_{\max})^2}{\gamma}.$$

To prove (18), let $c = \nu_{\min}/2, c = \nu_{\min}, c = 3\nu_{\min}/2, \dots$. Then, by applying Lemma 1, we obtain

$$P_\pi\{\Phi(\mathbf{X}(t)) \geq (1/2)\nu_{\min}m\} \geq \left(\frac{(1/2)p_{\min}\nu_{\min}}{(1/2)p_{\min}\nu_{\min} + \gamma}\right)^m P_\pi\{\Phi(\mathbf{X}(t)) \geq 0\}$$

$$= \left(\frac{(1/2)p_{\min}\nu_{\min}}{(1/2)p_{\min}\nu_{\min} + \gamma}\right)^m.$$

In order to prove (19), we have

$$E_\pi[\Phi(\mathbf{X}(t))] \geq \tfrac{1}{2} \sum_{m=0}^\infty \nu_{\min}\, m \mathrm{P}_\pi\big\{(1/2)\nu_{\min}\, m \leq \Phi(\mathbf{X}(t)) < (1/2)\nu_{\min}(m+1)\big\}$$

$$= \tfrac{1}{2}\nu_{\min} \sum_{m=1}^\infty \mathrm{P}_\pi\big\{(1/2)\nu_{\min}m \leq \Phi(\mathbf{X}(t))\big\}.$$

From (18) we obtain

$$E_\pi[\Phi(\mathbf{X}(t))] \geq \frac{1}{2}\nu_{\min} \sum_{m=1}^\infty \left(\frac{(1/2)p_{\min}\,\nu_{\min}}{(1/2)p_{\min}\,\nu_{\min} + \gamma} \right)^m = \frac{p_{\min}(\nu_{\min})^2}{4\gamma}.$$

This completes the proof of the theorem. $\square$

As mentioned above, the steady-state behavior of Markovian queueing networks is equivalent to the steady-state behavior of the embedded Markov chain. Applying Theorem 1, we can analyze the performance of Markovian queueing networks by constructing suitable Lyapunov functions. This is the subject of the following sections.

4. Lower bounds on queue lengths using linear lower Lyapunov functions. In this section, we use linear lower Lyapunov functions to find closed form lower bounds on the distribution and expectation of steady-state queue lengths, which hold when an arbitrary stable Markovian scheduling policy is implemented.

Given a stable scheduling policy w, let $\pi = \pi(w)$ denote the corresponding stationary distribution (of the queueing network and its embedded Markov chain). We will show that

$$E_\pi\left[\sum_{i,k} Q_{i,k}(t) \right] = \Omega\left(\sum_{j=1}^J \frac{1}{1 - \rho_{\sigma_j}} \right) = \Omega\left(\frac{1}{1 - \rho} \right),$$

where ρ_{σ_j} is the traffic intensity at station σ_j and $\rho = \max_{1 \leq j \leq J}\{\rho_{\sigma_j}\}$. We will also derive lower bounds on the distribution and expected queue lengths which hold specifically when priority policies are implemented, by using the notion of a *virtual station*. Finally, we will apply these results to some examples.

4.1. *Closed form lower bounds for arbitrary work-conserving policies.* Recall that under any Markovian scheduling policy, the transitions of the uniformized embedded Markov chain are given by Proposition 1. For each station σ_j, we now construct a lower Lyapunov function. For any class (i, k), let

$$(20) \qquad \rho_{i,k}^{\sigma_j+} = \sum_{k':\ (i,k')\in\sigma_j,\, k'\geq k} \rho_{i,k'}.$$

In words, $\rho_{i,k}^{\sigma_j+}$ is the sum of traffic intensities of classes of type i starting from stage k onward which are processed on station σ_j. Let

$$\Phi_j(\mathbf{Q}) = \sum_{i,k} \frac{\rho_{i,k}^{\sigma_j+}}{\lambda_i} Q_{i,k}.$$

PROPOSITION 2. *Let w be an arbitrary Markovian policy. Then, Φ_j is a lower Lyapunov function with drift $\gamma_j = 1 - \rho_{\sigma_j}$ and $p_{\min} = \sum_i \lambda_i, \nu_{\min} \geq \rho_{\sigma_j}/\lambda_{\max}$.*

PROOF. Using Proposition 1, we have

$$(21) \quad \mathrm{E}[\Phi_j(\mathbf{Q}(\tau_{s+1}))|\mathbf{Q}(\tau_s)] = \Phi_j(\mathbf{Q}(\tau_s)) + \sum_{i=1}^{I} \lambda_i \frac{\rho_{i,1}^{\sigma_j+}}{\lambda_i}$$
$$+ \sum_{i,k} \mu_{i,k} w_{i,k}(\mathbf{Q}(\tau_s))(\rho_{i,k+1}^{\sigma_j+} - \rho_{i,k}^{\sigma_j+})/\lambda_i.$$

Note from (20) that

$$\sum_{i=1}^{I} \lambda_i \frac{\rho_{i,1}^{\sigma_j+}}{\lambda_i} = \sum_{i=1}^{I} \sum_{k': (i,k')\in\sigma_j, k'\geq 1} \rho_{i,k'} = \rho_{\sigma_j}.$$

Observe that for any station σ_j,

$$\sum_{(i,k)\in\sigma_j} \mu_{i,k} w_{i,k}(\mathbf{Q}(\tau_s))(\rho_{i,k+1}^{\sigma_j+} - \rho_{i,k}^{\sigma_j+})/\lambda_i = \sum_{(i,k)\in\sigma_j} \mu_{i,k} w_{i,k}(\mathbf{Q}(\tau_s))\left(-\frac{\rho_{i,k}}{\lambda_i}\right)$$
$$= -\sum_{(i,k)\in\sigma_j} w_{i,k}(\mathbf{Q}(\tau_s)) \geq -1,$$

where the last inequality follows from the feasibility constraint (2) for the policy w. Also note that $\rho_{i,k+1}^{\sigma_j+} - \rho_{i,k}^{\sigma_j+} = 0$ when $(i,k) \notin \sigma_j$. Combining with (21) we obtain

$$\mathrm{E}[\Phi_j(\mathbf{Q}(\tau_{s+1}))|\mathbf{Q}(\tau_s)] - \Phi_j(\mathbf{Q}(\tau_s)) \geq \rho_{\sigma_j} - 1.$$

This proves that Φ_j is a lower Lyapunov function. We now bound the parameters $\nu_{\min}$ and $p_{\min}$. From Proposition 1, if a transition of the Markov chain $\mathbf{Q}(\tau_s)$ corresponds to a service completion in the class (i,k), then the corresponding change in the value of the Lyapunov function Φ_j is

$$-\rho_{i,k}^{\sigma_j+}/\lambda_i + \rho_{i,k+1}^{\sigma_j+}/\lambda_i,$$

which by definition is nonpositive. Therefore, the value of the Lyapunov function can increase only at the arrival times and, as a result, $p_{\min} = \sum_{i=1}^{I} \lambda_i$. At an arrival of a type i customer, the value of the Lyapunov function increases by $\rho_{\sigma_j}/\lambda_i$. Therefore $\nu_{\min} \geq \rho_{\sigma_j}/\lambda_{\max}$. □

We now are ready to state the main result of this section.

THEOREM 2. *Consider a multiclass queueing network* $(\boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{C})$ *operating under an arbitrary stable Markovian policy. The following lower bounds hold on the steady-state number of customers in the network: for each* $j = 1, 2, \ldots, J,$ *and* $m = 0, 1, 2, \ldots,$

$$\mathrm{P}\left\{\sum_{i,k} \frac{\rho_{i,k}^{\sigma_j+}}{\lambda_i} Q_{i,k}(t) \geq \frac{\rho_{\sigma_j}}{2\lambda_{\max}} m\right\} \geq \left(\frac{\rho_{\sigma_j}}{2 - \rho_{\sigma_j}}\right)^m$$

and

$$\mathrm{E}\left[\sum_{i,k} \frac{\rho_{i,k}^{\sigma_j+}}{\lambda_i} Q_{i,k}(t)\right] \geq \frac{\rho_{\sigma_j}^2}{4\lambda_{\max}(1 - \rho_{\sigma_j})}.$$

The result follows by applying Proposition 2 and Theorem 1.

REMARKS. (i) The bounds hold whether we have rescaled the parameters to $\sum_i \lambda_i + \sum_{i,k} \mu_{i,k} = 1$ or not, since ρ_{σ_j}, and the ratio $\lambda_i/\lambda_{\max}$ are insensitive to rescaling.

(ii) The bounds hold whether the policy used is work-conserving or not.

The bounds of Theorem 2 are simplified when the multiclass queueing network has a *reentrant line* structure, namely, $I = 1$. In this case, all customers follow the same route in the network. We denote by $Q_k(t)$ the queue length at the kth stage in the network. The parameters $\rho_{i,k}, \rho_{i,k}^{\sigma_j+}$ are denoted simply by ρ_k and $\rho_k^{\sigma_j+}$. The lower bounds on the queue lengths are simplified as follows.

COROLLARY 1. *Given a reentrant line-type queueing network* $(\boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{C})$, *operating under any stable Markovian policy, the following lower bounds hold on the number of customers in the network in steady-state. For each* $j = 1, 2, \ldots, J,$ *and* $m = 0, 1, 2, \ldots,$

$$\mathrm{P}\left\{\sum_k \rho_k^{\sigma_j+} Q_k(t) \geq \frac{\rho_{\sigma_j}}{2} m\right\} \geq \left(\frac{\rho_{\sigma_j}}{2 - \rho_{\sigma_j}}\right)^m$$

and

$$\mathrm{E}\left[\sum_k \rho_k^{\sigma_j+} Q_k(t)\right] \geq \frac{\rho_{\sigma_j}^2}{4(1 - \rho_{\sigma_j})}.$$

4.2. *Closed form lower bounds under a priority policy.* In this section, we derive lower bounds on the tail probabilities and the expected number of customers in a multiclass queueing network operating under a priority policy w_θ that is described by a permutation θ of the set of classes $\{(i, k)\}_{1 \leq i \leq I, 1 \leq k \leq J_i}$. For two classes $(i, k), (i', k')$ associated with the same station σ_j, we say that class (i', k') has a higher priority than class (i, k) if $\theta(i', k') < \theta(i, k)$.

A corresponding priority policy w_θ can be described as follows: for each state $\mathbf{q} \in Z_+^N$, $w_\theta(\mathbf{q})$ is an N-dimensional binary vector whose components satisfy

$$w_{i,k}^\theta(\mathbf{q}) = 1$$

if and only if $q_{i,k} > 0$ and $q_{i',k'} = 0$, whenever $(i', k') \in \sigma_j$, where j is such that $(i, k) \in \sigma_j$, and $\theta(i', k') < \theta(i, k)$. In other words, the policy w_θ at each transition epoch τ_s, selects within each station σ_j the highest priority class with a positive number of customers and works on a customer from this class. We thus assume that w_θ is a preemptive resume priority policy. Clearly, preemptive priority policies are Markovian.

The lower bounds to be derived in this section are based on the concept of a virtual station and virtual traffic intensity introduced in [8], where the virtual station concept is used for the stability analysis. We will show how virtual stations characterize the performance of multiclass queueing networks. The definitions below follow [8] very closely.

DEFINITION 3. A collection of classes $e = \{(i, k_1), (i, k_1 + 1), \ldots, (i, k_2)\}$, corresponding to a type-i customer is defined to be an excursion if all these classes are from some station σ_j, but classes $(i, k_1 - 1)$ and $(i, k_2 + 1)$ are not from station σ_j. This includes the possibility $k_1 = 1$ or $k_2 = J_i$. The classes $(i, k_1), \ldots, (i, k_2 - 1)$ are called the first classes of the excursion e and class (i, k_2) is called the last class of the excursion e.

We denote the sequence of all excursions corresponding to type i by e_1^i, $e_2^i, \ldots, e_R^i$.

DEFINITION 4. Given a multiclass queueing network $(\boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{C})$, suppose that a collection of stations $\Lambda \subset \{\sigma_1, \sigma_2, \ldots, \sigma_J\}$ with size $|\Lambda| = K$, and nonempty collections of classes $V_j \subset \sigma_j$, $j \in \Lambda$ are selected. The set of classes $V = \cup_{j \in \Lambda} V_j$ is defined to be a K-virtual (or just a virtual) station if the following conditions hold:

(i) No classes of the first excursion are in V: $e_1^i \cap V = \emptyset$, for each $i = 1, 2, \ldots, I$.

(ii) If the last class of some excursion e_l^i is in V, then all the classes of this excursion are in V, and if a first class of the excursion e_l^i is in V, then every first class of e_l^i is in V. Thus, a virtual station must have either none of the classes, all of the classes, or all but the last class of each excursion.

(iii) If a class (i, k) is the first class of an excursion e_l^i with $l \neq 1$ [that is $\sigma(i, k - 1) \neq \sigma(i, k)$], then class $(i, k) \in V$ if and only if $(i, k - 1) \notin V$.

For example, Classes 2, 3 and 5 in the reentrant line network in Figure 2, constitute a 2-virtual station. The following result was proved in [1] and [14].

PROPOSITION 3. *Suppose that a set of classes $V = \cup_{j \in \Lambda} V_j$ forms a K-virtual station for some $\Lambda \subset \{1, 2, \ldots, J\}$ and that w_θ is a stable priority policy*
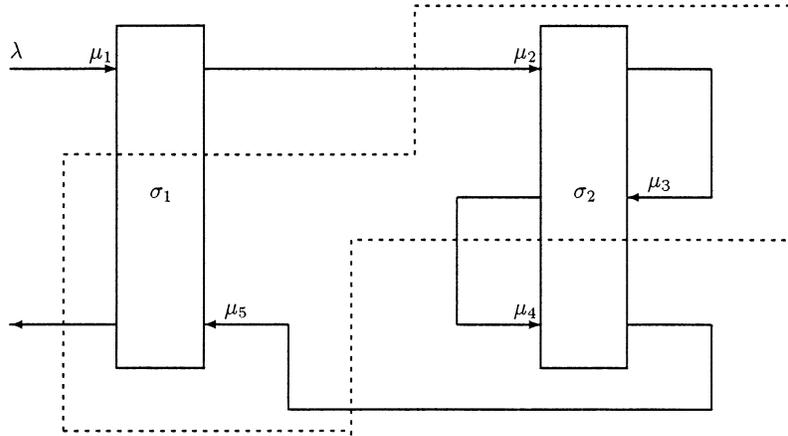
FIG. 2. *Classes 2, 3 and 5 constitute a virtual station in a reentrant line network.*

which gives priority to classes in V over classes not in V. Namely, whenever $(i, k) \in V$ and $(i', k') \notin V$, $\theta(i, k) < \theta(i', k')$. Then, the corresponding stationary distribution π_θ satisfies

$$P_{\pi_\theta}\left\{\sum_{(i, k) \in V} w_{i, k}^\theta(\mathbf{Q}(t)) \leq K - 1\right\} = 1.$$

Namely, in steady-state, at most $K - 1$ of classes in V can receive service simultaneously.

We see, in particular, that for networks with two stations, if V is a 2-virtual station for a priority policy w_θ, then, in steady-state, only one of the classes of V can receive service at a time. Thus, V acts as a station sharing its resources among its classes. This justifies the name *virtual station*.

Let $V = \cup_{j \in \Lambda} V_j$ be a K-virtual station in a multiclass queueing network with deterministic routing. Similar to Section 4.1, we introduce

$$(22) \qquad \rho(V) \equiv \sum_{(i, k) \in V} \rho_{i, k},$$

$$(23) \qquad \rho_{i, k}^{V+} \equiv \sum_{(i, k') \in V, k' \geq k} \rho_{i, k'}, \qquad 1 \leq i \leq I, 1 \leq k \leq J_i.$$

PROPOSITION 4. *Suppose w_θ is a stable priority policy in a multiclass queueing network $(\boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{C})$. Suppose also that the set of classes $V = \cup_{j \in \Lambda} V_j$ forms a K-virtual station, and w_θ gives priority to classes in V over classes not in V. Then, for the corresponding embedded Markov chain $\mathbf{Q}(\tau_s)$ the function*

$$\Phi(\mathbf{Q}) = \sum_{i, k} \frac{\rho_{i, k}^{V+}}{\lambda_i} Q_{i, k}$$

is a lower Lyapunov function with drift $K - 1 - \rho(V)$, $p_{\min} = \sum_i \lambda_i$ *and* $\nu_{\min} \geq \rho(V)/\lambda_{\max}$.

PROOF. From Proposition 1 in Section 2, we have

$$\mathrm{E}[\Phi(\mathbf{Q}(\tau_{s+1}))|\mathbf{Q}(\tau_s)] - \Phi(\mathbf{Q}(\tau_s))$$

$$= \sum_{i=1}^{I} \lambda_i \frac{\rho_{i,1}^{V+}}{\lambda_i} + \sum_{i,k} \mu_{i,k} w_{i,k}^{\theta}(\mathbf{Q}(\tau_s)) \frac{1}{\lambda_i} (\rho_{i,k+1}^{V+} - \rho_{i,k}^{V+}),$$

where we assume that $\rho_{i,J_i+1}^{V+} = 0$. Note that

$$\rho_{i,k+1}^{V+} - \rho_{i,k}^{V+} = \begin{cases} -\rho_{i,k}, & \text{if } (i,k) \in V, \\ 0, & \text{if } (i,k) \notin V. \end{cases}$$

Therefore,

$$\mathrm{E}[\Phi(\mathbf{Q}(\tau_{s+1}))|\mathbf{Q}(\tau_s)] - \Phi(\mathbf{Q}(\tau_s)) = \sum_i \rho_{i,1}^{V+} - \sum_{(i,k)\in V} w_{i,k}^{\theta}(\mathbf{Q}(\tau_s)).$$

From Proposition 3 and from $\sum_i \rho_{i,1}^{V+} = \rho(V)$ we obtain that the drift is $\gamma = K - 1 - \rho(V)$. We obtain the expressions for $p_{\min}$ and $\nu_{\min}$ as in the proof of Proposition 2. $\square$

A corollary of this result is the transience (instability) of a priority policy w_θ if for some virtual station V, we have $\rho(V) > K - 1$. This instability result was proven in [1] and [14] under more general assumptions, interarrival and service times have a general (as opposed to exponential) distribution. We now derive a matching performance result, when $\rho(V) < K - 1$. The following theorem is the main result of this section.

THEOREM 3. *Suppose we are given a multiclass queueing network* $(\boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{C})$, *and a set of classes* V *that forms a* K-*virtual station. If a stable priority policy* w_θ *gives priority to classes in* V *over the classes outside* V, *then the following lower bounds hold on the steady-state distribution and expectation of the number of customers in the network. For each* $j = 1, 2, \ldots, J$, *and* $m = 0, 1, 2, \ldots,$

$$\mathrm{P}\left\{\sum_{i,k} \frac{\rho_{i,k}^{V+}}{\lambda_i} Q_{i,k}(t) \geq \frac{\rho(V)}{2\lambda_{\max}} m\right\} \geq \left(\frac{\rho(V)}{2(K-1) - \rho(V)}\right)^m$$

and

$$\mathrm{E}\left[\sum_{i,k} \frac{\rho_{i,k}^{V+}}{\lambda_i} Q_{i,k}(t)\right] \geq \frac{\rho^2(V)}{4\lambda_{\max}(K - 1 - \rho(V))}.$$

The proof is similar to the one of Theorem 2.

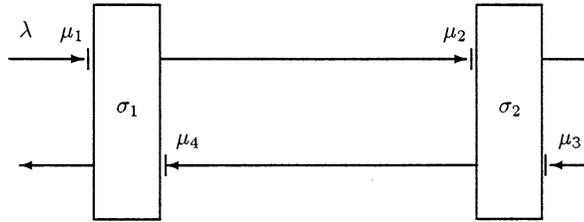The lower bounds of Theorem 3 are also simplified when the network is reentrant line-type.

FIG. 3.   *Lu–Kumar network.*

COROLLARY 2.   *Suppose that $(\boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{C})$ is a reentrant line-type queueing network and that a set of classes V is a K-virtual station. If a stable priority policy w_θ gives priority to classes in V over classes outside V, then the following lower bound holds on the number of customers in the network in steady-state. For each $m = 0, 1, 2, \ldots,$*

$$\mathrm{P}\left\{\sum_k \rho_k^{V+} Q_k(t) \geq \frac{\rho(V)}{2} m\right\} \geq \left(\frac{\rho(V)}{2(K-1) - \rho(V)}\right)^m$$

and

$$\mathrm{E}\left[\sum_k \rho_k^{V+} Q_k(t)\right] \geq \frac{\rho^2(V)}{4(K - 1 - \rho(V))}.$$

4.3. *Examples.*   In this section, we demonstrate the usage of the techniques developed in the previous sections on two specific networks.

The Lu–Kumar network.   Consider the network in Figure 3. This reentrant line network is described by the following parameters:

$$\boldsymbol{\lambda}' = (\lambda, 0, 0, 0)', \qquad \boldsymbol{\mu}' = (\mu_1, \mu_2, \mu_3, \mu_4)',$$

$$\rho_i = \lambda/\mu_i, \; i = 1, 2, 3, 4, \qquad \rho_{\sigma_1} = \rho_1 + \rho_4, \qquad \rho_{\sigma_2} = \rho_2 + \rho_3.$$

We have $\rho_1^{\sigma_1+} = \rho_1 + \rho_4$ and $\rho_i^{\sigma_1+} = \rho_4$ for $i = 2, 3, 4$. Similarly, $\rho_i^{\sigma_2+} = \rho_2 + \rho_3$ for $i = 1, 2$, $\rho_3^{\sigma_2+} = \rho_3$ and $\rho_4^{\sigma_2+} = 0$.

PROPOSITION 5.   *In the Lu–Kumar network of Figure 3, the following lower bounds on a stationary probability distribution π hold, under any stable scheduling policy:*

$$\mathrm{P}_\pi\left\{2\frac{(\rho_1 + \rho_4)Q_1(t) + \rho_4 Q_2(t) + \rho_4 Q_3(t) + \rho_4 Q_4(t)}{\rho_1 + \rho_4} \geq m\right\} \geq \left(\frac{\rho_1 + \rho_4}{2 - \rho_1 - \rho_4}\right)^m,$$

$$\mathrm{P}_\pi\left\{2\frac{(\rho_2 + \rho_3)Q_1(t) + (\rho_2 + \rho_3)Q_2(t) + \rho_3 Q_3(t)}{\rho_2 + \rho_3} \geq m\right\} \geq \left(\frac{\rho_2 + \rho_3}{2 - \rho_3 - \rho_2}\right)^m,$$

for all $m = 0, 1, 2, \ldots$ Also

$$\mathrm{E}_\pi\big[(\rho_1 + \rho_4)Q_1(t) + \rho_4 Q_2(t) + \rho_4 Q_3(t) + \rho_4 Q_4(t)\big] \geq \frac{1}{4}\frac{(\rho_1 + \rho_4)^2}{(1 - \rho_1 - \rho_4)},$$

$$\mathrm{E}_\pi\big[(\rho_2 + \rho_3)Q_1(t) + (\rho_2 + \rho_3)Q_2(t) + \rho_3 Q_3(t)\big] \geq \frac{1}{4}\frac{(\rho_2 + \rho_3)^2}{(1 - \rho_2 - \rho_3)}.$$

If, in addition, the network operates under priority policy w_θ with priority rule $\theta(4) < \theta(1)$, $\theta(2) < \theta(3)$, then

$$\mathrm{P}_{\pi_\theta}\bigg\{2\frac{(\rho_2 + \rho_4)Q_1(t) + (\rho_2 + \rho_4)Q_2(t) + \rho_4 Q_3(t) + \rho_4 Q_4(t)}{\rho_2 + \rho_4} \geq m\bigg\}$$

$$\geq \left(\frac{\rho_2 + \rho_4}{2 - \rho_2 - \rho_4}\right)^m,$$

for all $m = 0, 1, 2, \ldots$, and

(24)
$$\mathrm{E}_{\pi_\theta}\big[(\rho_2 + \rho_4)Q_1(t) + (\rho_2 + \rho_4)Q_2(t) + \rho_4 Q_3(t) + \rho_4 Q_4(t)\big]$$

$$\geq \frac{1}{4}\frac{(\rho_2 + \rho_4)^2}{(1 - \rho_2 - \rho_4)}.$$

The first part of the proposition is obtained by applying Corollary 1 to stations σ_1 and σ_2, the second part is obtained by applying Corollary 2 to the virtual station $V = \{2, 4\}$.

A 3-station, 6-class reentrant line. Consider the reentrant line queueing network with six classes and three stations described in Figure 1. This network was considered in [7], where the authors introduce the priority rule θ with $\theta(4) < \theta(1)$, $\theta(2) < \theta(5)$, $\theta(6) < \theta(3)$, and show that the set $V = \{2, 4, 6\}$ forms a 3-virtual station. Applying Corollary 2 we obtain the following result.

PROPOSITION 6. *Consider the network in Figure* 1, *under the priority rule* $\theta(4) < \theta(1)$, $\theta(2) < \theta(5)$, $\theta(6) < \theta(3)$. *Suppose in addition that the policy is stable. Then, for the corresponding stationary probability distribution π_θ,*

$$\mathrm{P}_{\pi_\theta}\bigg\{2\frac{(\rho_2 + \rho_4 + \rho_6)Q_1(t) + (\rho_2 + \rho_4 + \rho_6)Q_2(t) + (\rho_4 + \rho_6)Q_3(t)}{\rho_2 + \rho_4 + \rho_6}$$

$$+ 2\frac{(\rho_4 + \rho_6)Q_4(t) + \rho_6 Q_5(t) + \rho_6 Q_6(t)}{\rho_2 + \rho_4 + \rho_6} \geq m\bigg\} \geq \left(\frac{\rho_2 + \rho_4 + \rho_6}{4 - \rho_2 - \rho_4 - \rho_6}\right)^m,$$

for each $m = 0, 1, 2, \ldots$.
  Also,

$$\mathrm{E}_{\pi_\theta}\big[(\rho_2 + \rho_4 + \rho_6)Q_1(t) + (\rho_2 + \rho_4 + \rho_6)Q_2(t) + (\rho_4 + \rho_6)Q_3(t)$$

$$+ (\rho_4 + \rho_6)Q_4(t) + \rho_6 Q_5(t) + \rho_6 Q_6(t)\big] \geq \frac{1}{4}\frac{(\rho_2 + \rho_4 + \rho_6)^2}{2 - \rho_2 - \rho_4 - \rho_6}.$$

Observe that the lower bound on the expected number of customers in the network above has a singularity at $\rho_2 + \rho_4 + \rho_6 = 2$. This describes a heavy-traffic behavior not observed before in the literature.

5. Upper bounds on queue lengths using piecewise linear Lyapunov functions. The main focus of this section is in deriving upper bounds on the steady-state queue lengths in multiclass queueing networks by means of a piecewise linear Lyapunov function. Given that a certain linear program has a feasible solution, we construct a Lyapunov function for the embedded Markov chain $\mathbf{Q}(\tau_s)$ of the queueing network. In this way we obtain finite bounds on the tail distribution and the expectation of the queue lengths in the network, operating under any Markovian *work-conserving* policy.

5.1. *Arbitrary work-conserving policies.* Consider a multiclass Markovian queueing network $(\boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{C})$. Down and Meyn [10] (concentrating only on reentrant line case $I = 1$) showed that if the following linear program has a feasible solution with strictly positive γ, then any work-conserving policy is stable (global stability):

$$(25) \quad \text{GLP[dm]:} \quad L_{i,1}^j \lambda_i + L_{i,k+1}^j \mu_{i,k} - L_{i,k}^j \mu_{i,k} + V_j \leq -\gamma, \qquad (i,k) \in \sigma_j, 1 \leq j \leq J,$$

$$(26) \qquad\qquad L_{i,k+1}^j \mu_{i,k} - L_{i,k}^j \mu_{i,k} \leq V_j, \qquad (i,k) \notin \sigma_j,$$

$$(27) \qquad\qquad \frac{1}{J-1} \sum_{j' \neq j} L_{i,k}^{j'} \geq L_{i,k}^j, \qquad (i,k) \notin \sigma_j,$$

$$(28) \qquad\qquad L_{i,k}^j, V_j, \geq 0.$$

Specifically, using the technique of [12], Down and Meyn [10] proved that if GLP[dm] has a feasible solution with positive γ, then a random perturbation of the following piecewise linear function:

$$(29) \qquad\qquad \Phi(\mathbf{x}) \equiv \max_{1 \leq j \leq J} \{\mathbf{L}_j' \mathbf{x}\},$$

where $\mathbf{L}_j = (L_{i,k}^j)$, is a Lyapunov function with drift $\approx \gamma$ and some (unknown, but finite) exception parameter B. For networks with two stations the linear program GLP[dm] takes the following form (we denote it by LP[dm]):

$$(30) \qquad L_{i,1}^1 \lambda_i + L_{i,k+1}^1 \mu_{i,k} - L_{i,k}^1 \mu_{i,k} + V \leq -\gamma, \qquad (i,k) \in \sigma_1,$$

$$(31) \qquad\qquad L_{i,k+1}^1 \mu_{i,k} - L_{i,k}^1 \mu_{i,k} \leq V, \qquad (i,k) \in \sigma_2,$$

$$(32) \qquad L_{i,1}^2 \lambda_i + L_{i,k+1}^2 \mu_{i,k} - L_{i,k}^2 \mu_{i,k} + W \leq -\gamma, \qquad (i,k) \in \sigma_2,$$

$$(33) \qquad\qquad L_{i,k+1}^2 \mu_{i,k} - L_{i,k}^2 \mu_{i,k} \leq W, \qquad (i,k) \in \sigma_1,$$

$$(34) \qquad\qquad\qquad L_{i,k}^1 \geq L_{i,k}^2, \qquad (i,k) \in \sigma_1,$$

$$(35) \qquad\qquad\qquad L_{i,k}^1 \leq L_{i,k}^2, \qquad (i,k) \in \sigma_2,$$

$$(36) \qquad\qquad \mathbf{L}, V, W, \gamma \geq \mathbf{0}.$$

The piecewise linear function (29) is very "close" to being a Lyapunov function of the embedded Markov chain $\mathbf{Q}(\tau_s)$. It satisfies the drift condition (8) for all $\mathbf{x} \in Z_+^N$, except near the intersections of hyperplanes, namely, near the sets $\{\mathbf{x} \in Z_+^N: \mathbf{L}_i'\mathbf{x} = \mathbf{L}_j'\mathbf{x}, i, j = 1, 2, \ldots, J, i \neq j\}$. The "smoothing" random perturbation used by Down and Meyn solves this technical difficulty. We now use the same smoothing operation as in [10] but, unlike [10], our derivation is explicit and a closed form estimate of the exception parameter B will be given. In fact, we reestablish the results obtained in [10]. Again we rescale time so that

$$\sum_{i=1}^{I} \lambda_i + \sum_{i, k} \mu_{i, k} = 1.$$

Let $\mathbf{L}_1, \mathbf{L}_2, \ldots, \mathbf{L}_J, \gamma$ be any feasible solution of GLP[dm] with $\gamma > 0$. Let

$$L_{\max} \equiv \max_{i, k, j}\{L_{i, k}^j\}.$$

For all $j = 1, 2, \ldots, J$, we let

$$
\begin{aligned}
O_j = \Bigg\{ &\mathbf{z} = (z_{1,1}, z_{1,2}, \ldots, z_{I, J_I}) \in \Re_+^N: \\
&z_{i, k} \in \left[L_{i, k}^j + \frac{J-1}{2J}\gamma, L_{i, k}^j + \frac{1}{2}\gamma\right], \text{ for } (i, k) \in \sigma_j, \\
&z_{i, k} \in \left[L_{i, k}^j, L_{i, k}^j + \frac{1}{2J}\gamma\right], \text{ for } (i, k) \notin \sigma_j \Bigg\}.
\end{aligned}
$$

(37)

Consider the uniform probability density function $p_j(\mathbf{z})$ on the set O_j. We will denote by $\mathbf{Z}_j$ a random variable with distribution $p_j(\mathbf{z})$ and denote by $\mathbf{z}_j$ a sample point from the set O_j.

For any $(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_J) \in O_1 \times \cdots \times O_J$ and any $\mathbf{x} \in Z_+^N$ let

(38)
$$\Phi_0(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_J, \mathbf{x}) = \max_{1 \leq j \leq J}\{\mathbf{z}_j'\mathbf{x}\}$$

and let

(39)
$$
\begin{aligned}
\Phi_s(\mathbf{x}) &= \mathrm{E}_u\big[\Phi_0(\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_J, \mathbf{x})\big] \\
&= \int \Phi_0(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_J, \mathbf{x}) p_1(\mathbf{z}_1) \cdots p_J(\mathbf{z}_J)\, d\mathbf{z}_1\, d\mathbf{z}_2 \cdots d\mathbf{z}_J.
\end{aligned}
$$

We use a subscript u to emphasize the uniform distribution u. We next show that the modified function Φ_s is a Lyapunov function.

PROPOSITION 7. *Let $\mathbf{L}_1, \mathbf{L}_2, \ldots, \mathbf{L}_J, \gamma$ be any feasible solution of* GLP[dm] *with $\gamma > 0$. Then for any Markovian work-conserving policy w, Φ_s is a Lyapunov function of the embedded Markov chain $\mathbf{Q}(\tau_s)$ with drift equal to $\frac{1}{4}\gamma$ and exception parameter*

$$B = \frac{16NJ^2(J-1)(L_{\max} + \gamma)^3}{\gamma^2}.$$

Also $\nu_{\max} \leq L_{\max} + (1/2)\gamma$.

For the proof, see Appendix B.

We now apply Theorem 1 to obtain the following result.

THEOREM 4.   *Given a multiclass queueing network $(\boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{C})$, with parameters rescaled so that $\sum_{i=1}^{I} \lambda_i + \sum_{i,k} \mu_{i,k} = 1$, suppose that the corresponding linear program* GLP[dm] *has a feasible solution* $\mathbf{L}_1, \mathbf{L}_2, \ldots, \mathbf{L}_J, \gamma$ *with positive* γ. *Then the following upper bound holds on the stationary distribution* π *corresponding to any stable work-conserving Markovian policy* w:

$$\mathrm{P}_\pi \left\{ \frac{\mathbf{L}_j' \mathbf{Q}(t) - B}{2(L_{\max} + (1/2)\gamma)} \geq m \right\} \leq \left(\frac{L_{\max} + (1/2)\gamma}{L_{\max} + (3/4)\gamma} \right)^m,$$

for all $m = 0, 1, 2, \ldots,$ *and all* $j = 1, 2, \ldots, J,$ *where*

$$L_{\max} = \max_{1 \leq j \leq J,\, 1 \leq i \leq I,\, 1 \leq k \leq J_i} \left\{ L_{i,k}^j \right\}$$

and

$$B = \frac{16 N J^2 (J-1)(L_{\max} + \gamma)^3}{\gamma^2}.$$

Also

$$\mathrm{E}_\pi[\mathbf{L}_j' \mathbf{Q}(t)] \leq \frac{16 N J^2 (J-1)(L_{\max} + \gamma)^3}{\gamma^2} + \frac{8(L_{\max} + (1/2)\gamma)^2}{\gamma},$$

for all $j = 1, 2, \ldots, J.$

PROOF.   The bounds are a direct corollary of Proposition 7, Theorem 1, equation (38) and the fact

$$\Phi_0(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_J, \mathbf{x}) \geq \mathbf{z}_j' \mathbf{x} \geq \mathbf{L}_j' \mathbf{x},$$

for all $\mathbf{z}_j \in O_j$, $j = 1, 2, \ldots, J$. We also use $p_{\max} \leq 1$. □

REMARK.   It is known (see [2, 8]) that a fluid network with two stations is globally stable if and only if the linear program LP[dm] has a feasible solution with positive γ. Therefore, for networks with two stations, the bounds are finite if and only if the corresponding fluid network is globally stable.

5.2. *Upper bounds for networks with two stations.*   In this section, we provide explicit performance bounds for queueing networks with two stations. We will consider only reentrant line queueing networks. The reference to the type i is thus omitted. The Poisson arrival rate is denoted by λ. An explicit and tight characterization of global stability of fluid networks with two stations is given in [8]. Specifically, it is proved that a fluid queueing network with two stations is globally stable if and only if the maximal of all the real and virtual traffic intensities ρ^* (to be defined below) is smaller than 1. From this result and Dai's theorem [6] connecting fluid and stochastic stability, the condition $\rho^* < 1$ is also sufficient for global stability of the stochastic network (with arbitrary and not necessarily exponential service distribution). In this section, we

derive a matching performance result: whenever $\rho^* < 1$, we construct a finite upper bound on the tail probabilities and the expectation of queue lengths in the network. We show that ρ^* is a fundamental performance parameter of the network. In particular, we prove that under any work-conserving policy w, the corresponding stationary distribution π satisfies

$$\mathrm{E}_\pi\left[\sum_{i=1}^N Q_i(t)\right] = O\left(\frac{1}{(1-\rho^*)^2}\right).$$

Following [8], we introduce the definitions of separating sets and recall the definition of a 2-virtual station (Definition 4 with $K = 2$). In this section, we will refer to a 2-virtual station as a virtual station as we only consider networks with only two stations. Recall that a set of classes $\{k_1, \ldots, k_2\}$ is defined to be an excursion if all of these classes belong to some station σ_j, $j = 1, 2$, but classes $k_1 - 1$, $k_2 + 1$ are not from station σ_j. Let $e_1, e_2, \ldots, e_R$ denote the set of all excursions. We assume without loss of generality that $e_1 \subset \sigma_1$. For each excursion $e_r = \{k_1, \ldots, k_2\}$, the class k_2 is called the last class of excursion e_r and is denoted by $l(e_r)$. The classes $k_1, \ldots, k_2 - 1$ are called the first classes of the excursion e_r and are denoted by $f(e_r)$.

DEFINITION 5. A set of excursions S is defined to be a separating set if it contains no consecutive excursions. Namely, $e_r \in S$ implies $e_{r-1}, e_{r+1} \notin S$. We have assumed that the first excursion e_1 belongs to the first station; that is, $e_1 \subset \sigma_1$. A separating set S is defined to be strictly separating if it does not contain e_1. Two separating sets consisting only of excursions in σ_1 or of excursions in σ_2 are called trivial separating sets.

Each separating set of excursions induces a collection $V(S)$ consisting of the classes in excursions in S together with the first classes of excursions (other than e_1) whose immediate predecessor is not in S. Thus,

$$V(S) = \left(\bigcup_{e_r \in S} e_r\right) \cup \left(\bigcup_{e_r \notin S} f(e_{r+1})\right).$$

If S is in addition strictly separating, we refer to $V(S)$ as a virtual station.

It is not hard to see that if S is a strictly separating set then $V(S)$ is a virtual station as defined by Definition 4.

We now introduce some additional notations. Let S be a separating set and let us choose an excursion e_r. Denote

$$\rho(V(S), \sigma_1) \equiv \sum_{k \in \sigma_1 \cap V(S)} \rho_k,$$

(40)
$$\rho(V(S), e_r, \sigma_1) \equiv \sum_{k \in \sigma_1 \cap V(S),\, k > l(e_r)} \rho_k,$$

$$\rho(e_r, \sigma_1) \equiv \sum_{k \in \sigma_1,\, k < l(e_r)} \rho_k.$$

Similarly, we introduce $\rho(V(S), \sigma_2)$, $\rho(V(S), e_r, \sigma_2)$ and $\rho(e_r, \sigma_2)$. Let ρ^* denote the maximal actual or virtual traffic intensity:

$$(41) \qquad \rho^* \equiv \max\Big\{\max_{S, e_r}\{\rho(V(S), e_r)\}, \rho_{\sigma_1}, \rho_{\sigma_2}\Big\},$$

where

$$(42) \qquad \begin{aligned} \rho(V(S), e_r) &= \rho(V(S), e_r, \sigma_1)(1 - \rho(e_r, \sigma_2)) + \rho(V(S), e_r, \sigma_2) \\ &\quad \times (1 - \rho(e_r, \sigma_1)) + \rho(e_r, \sigma_1) + \rho(e_r, \sigma_2) - \rho(e_r, \sigma_1)\rho(e_r, \sigma_2). \end{aligned}$$

Dai and Vande Vate [8] proved that if $\rho^* < 1$, then a two-station queueing network is stable under all work-conserving policies (globally stable). Also, if there exists a virtual station $V(S)$ such that $\rho(V(S)) > 1$ then there exists an unstable priority policy.

Our goal in this section is to derive closed form upper bounds on the steady-state number of customers in the network, in terms of the parameter ρ^*. An outline of our approach is as follows. We consider a certain modification of the linear program LP[dm] from Section 5.1. We use the results in [8] to show that if $\rho^* < 1$, then this modified linear program has a feasible solution with positive γ and the result of Theorem 4 becomes applicable. In addition, by analyzing the linear program we obtain explicit bounds on the solution variables and specifically on the drift γ. The latter allows us to obtain the explicit dependence of the drift on the maximal traffic intensity ρ^*.

We consider now the following linear program considered in [8] (Equations (4.11)–(4.15), (5.1), (5.2) in [8]):

$$(43) \qquad \lambda \sum_{i \in \sigma_j} x_i - \mu_k x_k + \lambda \varepsilon \leq 0, \quad k \in \sigma_j, j = 1, 2,$$

$$(44) \qquad \sum_{i \in \sigma_1, i > l(e)} x_i - \sum_{i \in \sigma_2, i \geq l(e)} x_i \leq 0 \quad \text{for any excursion } e \subset \sigma_2,$$

$$(45) \qquad \sum_{i \in \sigma_2, i > l(e)} x_i - \sum_{i \in \sigma_1, i \geq l(e)} x_i \leq 0 \quad \text{for any excursion } e \subset \sigma_1,$$

$$(46) \qquad \sum_{k \in \sigma_1} x_k + \varepsilon = 1,$$

$$(47) \qquad \sum_{k \in \sigma_2} x_k + \varepsilon = \beta,$$

$$(48) \qquad x, \varepsilon \geq 0.$$

We denote this linear program by LP[dv]. Note that β could be treated as a variable in the linear program above. But instead, as in [8], we will treat it as a parameter. Note also that constraints (43), (46), (47) and (48) of LP[dv] imply

$$(49) \qquad x_k \geq \rho_k, \qquad k \in \sigma_1,$$

$$(50) \qquad x_k \geq \beta \rho_k, \qquad k \in \sigma_2.$$

We now show that if LP[dv] has a feasible solution with positive ε, then LP[dm] also has a feasible solution with positive γ.

PROPOSITION 8.   *Let* $\mathbf{x}$, ε *be a feasible solution to* LP[dv]. *Let also*

(51)
$$L_k^1 = \sum_{k' \in \sigma_1, \, k' \geq k} x_{k'}, \quad L_k^2 = \sum_{k' \in \sigma_2, \, k' \geq k} x_{k'},$$

$$\mathbf{L}_j = (L_1^j, \ldots, L_N^j), \quad j = 1, 2, \qquad \gamma = \lambda \varepsilon.$$

Then $\mathbf{L}_1$, $\mathbf{L}_2$, γ, $V = 0$, $W = 0$ *is a feasible solution to* LP[dm]. *In particular, if* ε *is positive then* γ *is also positive. This solution satisfies* $L_k^j \geq L_{k'}^j$ *whenever* $k' \geq k$.

For the proof, see Appendix B.

The connection between the linear program LP[dv] and $\rho^* < 1$ is established in [8] by using network flow techniques. Specifically, Section 5 of [8] shows that if there exists a β such that

(52)
$$\frac{1 - \rho(V(S), \sigma_1)}{\rho(V(S), \sigma_2)} > \beta > \frac{\rho(V(S'), \sigma_1)}{1 - \rho(V(S'), \sigma_2)}$$

for every nontrivial strictly separating set S', and every nontrivial separating set S, then there exists a feasible solution $\varepsilon = \varepsilon(\beta) > 0$ to the linear program LP[dv] with

(53)
$$\varepsilon(\beta) \equiv \min\{1 - \rho_{\sigma_1}, 1 - \rho_{\sigma_2}, \beta(1 - \rho(V(S'), \sigma_2)) - \rho(V(S'), \sigma_1),$$
$$(1 - \rho(V(S), \sigma_1)) - \beta\rho(V(S), \sigma_2)\}$$

(the minimum is over all strictly separating sets S' and all separating sets S).

In the next lemma, which is a slight modification of the argument in Section 6 of [8], we establish the connection between the linear program LP[dv] and the condition $\rho^* < 1$.

LEMMA 2.   *Suppose* $\rho^* < 1$. *Then, there exists a* β *and a feasible solution* $\mathbf{x}$, ε *of* LP[dv] *such that* $\varepsilon \geq 1 - \rho^*$.

For the proof, see Appendix B.

We now have all the necessary tools to state and prove the main result of this section.

THEOREM 5.   *We consider a reentrant line queueing network with two stations* σ_1, σ_2, *arrival rate* λ *and service rates* $\mu_1, \mu_2, \ldots, \mu_N$. *Class* 1 *is assumed to belong to station* σ_1. *If* $\rho^* < 1$, *then the following upper bounds hold on the*

steady-state number of customers in the network:

(54)
$$\mathrm{P}\left\{\sum_{i=1}^{N}\rho_i^{\sigma_1+}Q_i(t)-B\geq m\frac{1+\rho^*+2\rho^*\sum_{i=1}^N\rho_i^{-1}}{1+\sum_{i=1}^N\rho_i^{-1}}\right\}$$
$$\leq\left(\frac{\frac{1}{2}+\frac{1}{2}\rho^*+\rho^*\sum_{i=1}^N\rho_i^{-1}}{\frac{3}{4}+\frac{1}{4}\rho^*+\rho^*\sum_{i=1}^N\rho_i^{-1}}\right)^m$$

and

(55)
$$\mathrm{P}\left\{\sum_{i=1}^{N}\rho_{l(e_2)+1}^{\sigma_1+}\rho_i^{\sigma_2+}Q_i(t)-B\geq m\frac{1+\rho^*+2\rho^*\sum_{i=1}^N\rho_i^{-1}}{1+\sum_{i=1}^n\rho_i^{-1}}\right\}$$
$$\leq\left(\frac{\frac{1}{2}+\frac{1}{2}\rho^*+\rho^*\sum_{i=1}^N\rho_i^{-1}}{\frac{3}{4}+\frac{1}{4}\rho^*+\rho^*\sum_{i=1}^N\rho_i^{-1}}\right)^m$$

for all $m=0,1,2,\ldots,$ *where*

$$B=\frac{64N(\rho^*\sum_{i=1}^N\rho_i^{-1})^3}{(1+\sum_{i=1}^N\rho_i^{-1})(1-\rho^*)^2}.$$

Also

(56)
$$\mathrm{E}\left[\sum_{i=1}^{N}\rho_i^{\sigma_1+}Q_i(t)\right]\leq\frac{64N(\rho^*\sum_{i=1}^N\rho_i^{-1})^3}{(1+\sum_{i=1}^N\rho_i^{-1})(1-\rho^*)^2}+\frac{2(1+\rho^*+2\rho^*\sum_{i=1}^N\rho_i^{-1})^2}{(1+\sum_{i=1}^N\rho_i^{-1})(1-\rho^*)}$$

and

(57)
$$\mathrm{E}\left[\sum_{i=1}^{N}\rho_{l(e_2)+1}^{\sigma_1+}\rho_i^{\sigma_1+}Q_i(t)\right]\leq\frac{64N(\rho^*\sum_{i=1}^N\rho_i^{-1})^3}{(1+\sum_{i=1}^N\rho_i^{-1})(1-\rho^*)^2}$$
$$+\frac{2(1+\rho^*+2\rho^*\sum_{i=1}^N\rho_i^{-1})^2}{(1+\sum_{i=1}^N\rho_i^{-1})(1-\rho^*)}.$$

In particular,

(58)
$$\mathrm{E}\left[\sum_{i=1}^{N}Q_i(t)\right]=O\left(\frac{1}{(1-\rho^*)^2}\right).$$

REMARKS. (i) The bounds are asymmetric with respect to the order of the stations. If class 1 belongs to Station σ_2, the corresponding bounds are obtained trivially by exchanging σ_1 and σ_2.

(ii) The condition $\rho^*<1$ guarantees that

$$\frac{\frac{1}{2}+\frac{1}{2}\rho^*+\rho^*\sum_{i=1}^N\rho_i^{-1}}{\frac{3}{4}+\frac{1}{4}\rho^*+\rho^*\sum_{i=1}^N\rho_i^{-1}}<1.$$

As a result, the bounds of the theorem are nontrivial and, in particular, are of the geometric type.

For the proof of Theorem 5, see Appendix B.

6. Extensions and examples. We apply the results obtained in the previous section to several specific examples.

Feedforward networks. We start with a definition.

DEFINITION 6. A multiclass queueing network $(\boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{C})$ is defined to be feedforward (acyclic) if $(i, k) \in \sigma_{j_1}, (i, k+1) \in \sigma_{j_2}$ implies $j_1 \leq j_2$. In words, customers visit the stations in nondecreasing order.

The stability of feedforward networks under the usual load conditions $\rho_{\sigma_j} < 1$ was proved in [6] and [11]. Since the stability conditions for feedforward networks are given explicitly as $\rho_{\sigma_j} < 1$, then it is natural to expect that performance bounds can be constructed, which are finite whenever the load condition $\rho_{\sigma_j} < 1$ holds. In the next theorem we will show that this is the case.

THEOREM 6. *Consider a feedforward multiclass queueing network $(\boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{C})$ operating under an arbitrary work-conserving policy π. Let $\rho_{\min} = \min_{i, k} \rho_{i, k}$, $\rho^* = \max_j \rho_{\sigma_j}$, and let $\rho_{i, k}^{\sigma_j +}$ be defined by (20). The following upper bounds hold on the steady-state number of customers in the network*:

$$\mathrm{P}\left\{ \frac{\overline{L}_j' Q(t) - B}{2(\rho^* + (1/2)\bar{\lambda}_{\min}(1 - \rho^*))} \geq m \right\} \leq \left(\frac{\rho^* + (1/2)\bar{\lambda}_{\min}(1 - \rho^*)}{\rho^* + (3/4)\bar{\lambda}_{\min}(1 - \rho^*)} \right)^m,$$

for all $m = 0, 1, 2, \ldots$, and all $j = 1, 2, \ldots, J$, where

$$(59) \qquad \overline{L}_{i, k}^j = \left(\frac{\rho_{\min}}{(J - 1)\rho^*} \right)^{j-1} \rho_{i, k}^{\sigma_j +}, \qquad \overline{L}_j = (\overline{L}_{i, k}^j)_{i, k},$$

$$\bar{\lambda}_{\min} = \left(\frac{\rho_{\min}}{(J - 1)\rho^*} \right)^{J-1} \lambda_{\min},$$

$$(60) \qquad B = \frac{16 N J^2 (J - 1)(\rho^* + \bar{\lambda}_{\min}(1 - \rho^*))^3}{\bar{\lambda}_{\min}^2 (1 - \rho^*)^2}.$$

Also,

$$\mathrm{E}[\overline{L}_j' Q(t)] \leq \frac{16 N J^2 (J - 1)(\rho^* + \bar{\lambda}_{\min}(1 - \rho^*))^3}{\bar{\lambda}_{\min}^2 (1 - \rho^*)^2} + \frac{8(\rho^* + \bar{\lambda}_{\min}(1 - \rho^*))^2}{\bar{\lambda}_{\min}(1 - \rho^*)},$$

for all $j = 1, 2, \ldots, J$. In particular,

$$\mathrm{E}\left[\sum_{i=1}^N Q_i(t) \right] = O\left(\frac{1}{(1 - \rho^*)^2} \right).$$

For the proof, see Appendix B.

REMARK.   The bound

$$\mathrm{E}\left[\sum_{i=1}^{N} Q_i(t)\right] = O\left(\frac{1}{(1-\rho^*)^2}\right),$$

is an improvement on the bound

$$\mathrm{E}\left[\sum_{i=1}^{N} Q_i(t)\right] = O\left(\frac{1}{(1-\rho^*)^{J_0}}\right),$$

obtained in [15] using quadratic Lyapunov functions. Here J_0 denotes the number of stations with traffic intensity equal to ρ^* (the number of bottleneck stations). Note that it is possible to have $J_0 = J$.

The Lu–Kumar network.   Consider the network in Figure 3. The network is described by the following parameters:

$$\boldsymbol{\lambda}' = (\lambda, 0, 0, 0)', \qquad \boldsymbol{\mu}' = (\mu_1, \mu_2, \mu_3, \mu_4)',$$

$$\rho_i = \lambda/\mu_i, \, i = 1, 2, 3, 4, \quad \rho_{\sigma_1} = \rho_1 + \rho_4, \quad \rho_{\sigma_2} = \rho_2 + \rho_3.$$

We have $\rho_1^{\sigma_1+} = \rho_1 + \rho_4$ and $\rho_i^{\sigma_1+} = \rho_4$ for $i = 2, 3, 4$.

The set of excursions in this network is given as $e_1 = \{1\}, e_2 = \{2, 3\}$, $e_3 = \{4\}$. Then $\rho_{l(e_2)+1}^{\sigma_1+} = \rho_4$. The only two nontrivial separating sets in this network consist of the single excursions $e_1 = \{1\}$ and $e_3 = \{4\}$. The separating set e_1 with its set of classes $V(\{e_1\}) = \{1\}$ has $\rho(V(S), e_k) = 0$ for all $k = 1, 2, 3$. For the separating set e_3, with its set of classes $V(\{e_3\}) = \{2, 4\}$, we have

$$\rho(V(\{e_3\}), e_1) = \rho_2 + \rho_4, \qquad \rho(V(\{e_3\}), e_2) = \rho_1 + \rho_4.$$

However, the second term is equal to $\rho_{\sigma_2} < 1$. We conclude that

$$\rho^* = \max\{\rho_2 + \rho_4, \rho_{\sigma_1}, \rho_{\sigma_2}\}.$$

Assume now, in addition, that $\rho_2 \geq \rho_3$ and $\rho_4 \geq \rho_1$. Then $\rho^* = \rho_2 + \rho_4$. Applying now Theorem 5, we obtain the following result.

PROPOSITION 9.   *In a Lu–Kumar network satisfying $\rho_2 \geq \rho_3, \rho_4 \geq \rho_1$ and $\rho_2 + \rho_4 < 1$, the following upper bounds hold for any work-conserving policy w and the corresponding stationary probability distribution π:*

$$\mathrm{P}_\pi\Big\{(\rho_1 + \rho_4)Q_1(t) + \rho_4 Q_2(t) + \rho_4 Q_3(t) + \rho_4 Q_4(t) - B$$

$$\geq m\frac{1 + \rho_2 + \rho_4 + 2(\rho_2 + \rho_4)\sum_{i=1}^{4}\rho_i^{-1}}{1 + \sum_{i=1}^{4}\rho_i^{-1}}\Big\}$$

$$\leq \left(\frac{\frac{1}{2} + \frac{1}{2}(\rho_2 + \rho_4) + (\rho_2 + \rho_4)\sum_{i=1}^{4}\rho_i^{-1}}{\frac{3}{4} + \frac{1}{4}(\rho_2 + \rho_4) + (\rho_2 + \rho_4)\sum_{i=1}^{4}\rho_i^{-1}}\right)^{m}$$

and

$$\mathrm{P}_\pi \Bigg\{ \rho_4(\rho_2+\rho_3)Q_1(t) + \rho_4(\rho_2+\rho_3)Q_2(t) + \rho_4\rho_3 Q_3(t) - B$$

$$\geq m\, \frac{1 + \rho_2 + \rho_4 + (\rho_2+\rho_4)\sum_{i=1}^4 \rho_i^{-1}}{1 + \sum_{i=1}^4 \rho_i^{-1}} \Bigg\}$$

$$\leq \left(\frac{\frac{1}{2} + \frac{1}{2}(\rho_2+\rho_4) + (\rho_2+\rho_4)\sum_{i=1}^4 \rho_i^{-1}}{\frac{3}{4} + \frac{1}{4}(\rho_2+\rho_4) + (\rho_2+\rho_4)\sum_{i=1}^4 \rho_i^{-1}} \right)^m,$$

for all $m = 0, 1, 2, \ldots,$ *where*

$$B = \frac{256(\rho_2+\rho_4)^3 (\sum_{i=1}^4 \rho_i^{-1})^3}{(1 + \sum_{i=1}^4 \rho_i^{-1})(1 - \rho_2 - \rho_4)^2}.$$

Also,

$$\mathrm{E}_\pi\big[(\rho_1+\rho_4)Q_1(t) + \rho_4 Q_2(t) + \rho_4 Q_3(t) + \rho_4 Q_4(t) \big]$$

$$\leq \frac{256(\rho_2+\rho_4)^3 (\sum_{i=1}^4 \rho_i^{-1})^3}{(1 + \sum_{i=1}^4 \rho_i^{-1})(1 - \rho_2 - \rho_4)^2} + \frac{2(1 + \rho_2 + \rho_4 + 2(\rho_2+\rho_4)\sum_{i=1}^4 \rho_i^{-1})^2}{(1 + \sum_{i=1}^4 \rho_i^{-1})(1 - \rho_2 - \rho_4)}$$

and

$$\mathrm{E}_\pi\big[\rho_4(\rho_2+\rho_3)Q_1(t) + \rho_4(\rho_2+\rho_3)Q_2(t) + \rho_4\rho_3 Q_3(t) \big]$$

$$\leq \frac{256(\rho_2+\rho_4)^3 (\sum_{i=1}^4 \rho_i^{-1})^3}{(1 + \sum_{i=1}^4 \rho_i^{-1})(1 - \rho_2 - \rho_4)^2} + \frac{2(1 + \rho_2 + \rho_4 + 2(\rho_2+\rho_4)\sum_{i=1}^4 \rho_i^{-1})^2}{(1 + \sum_{i=1}^4 \rho_i^{-1})(1 - \rho_2 - \rho_4)}.$$

Similar bounds can be obtained for the cases $\rho_1 > \rho_4$ or $\rho_3 > \rho_2$. Note that the result above implies

$$\mathrm{E}_\pi\left[\sum_{i=1}^4 Q_i(t) \right] = O\!\left(\frac{1}{(1 - \rho_2 - \rho_4)^2} \right).$$

Contrast this with the lower bounds (24).

7. Conclusions. We have proposed a general methodology based on Lyapunov functions for the performance analysis of infinite state Markov chains and applied it specifically to multiclass queueing networks with exponentially distributed interarrival and service times.

We have proved that whenever some piecewise linear Lyapunov function is a witness for the global stability of the network, certain finite upper bounds can be derived on the probability distribution and expectation of queue lengths. Lower bounds are also constructed by means of linear lower Lyapunov functions. Thus, for certain computable constants $0 < c_1 < c_2 < 1$, we have constructed bounds of the form

$$c_1^m \leq \mathrm{P}\{L \geq m\} \leq c_2^m,$$

with L the total number of customers in the network. These bounds hold uniformly under any work conserving policy. The lower bounds are extended to priority policies as well.

Since piecewise linear Lyapunov functions provide an exact test for stability of fluid networks with two stations, our bounds for two-station networks are finite if and only if the corresponding fluid network is globally stable. Whether this remains true for the original stochastic network remains to be seen.

For reentrant line-type queueing networks with two processing stations, the constants c_1 and c_2 can be expressed explicitly in terms of traffic intensities (actual and virtual) of the network. Closed form bounds were also constructed on the total expected number of customers in the network. In particular, we have proved that

$$\mathrm{E}[L] = O\left(\frac{1}{(1 - \rho^*)^2}\right),$$

where ρ^* is a maximal (actual or virtual) traffic intensity. It would be interesting to strengthen this result, perhaps by removing the exponent 2.

The results obtained here are the first ones that establish exponential upper and lower bounds on the distribution of queue lengths in networks of such generality. Previous results on performance analysis of multiclass queueing networks can in general achieve only numerical bounds and only on the expectation of queue lengths.

APPENDIX A

PROOF OF LEMMA 1. The key to our analysis is a modified Lyapunov function, defined by

$$(61) \qquad \widehat{\Phi}(\mathbf{x}) = \max\{c, \Phi(\mathbf{x})\}$$

for some $c \in R_+$, and the corresponding equilibrium equation

$$(62) \qquad \mathrm{E}_\pi\big[\widehat{\Phi}(\mathbf{X}(t))\big] = \mathrm{E}_\pi\big[\widehat{\Phi}(\mathbf{X}(t + 1))\big].$$

We can rewrite (62) as

$$(63) \qquad \mathrm{E}_\pi\big[\widehat{\Phi}(\mathbf{X}(t + 1)) - \widehat{\Phi}(\mathbf{X}(t))\big] = 0.$$

We first prove (14). Let us fix c as in the statement of the lemma, and consider the function $\widehat{\Phi}(\mathbf{x})$ introduced in (61). Since $\mathrm{E}_\pi[\Phi(\mathbf{X}(t))]$ is finite and π is a stationary distribution, we can rewrite (63) as

$$(64) \qquad \sum_{\mathbf{x}} \pi(\mathbf{x})\big(\mathrm{E}[\widehat{\Phi}(\mathbf{X}(t + 1)) \mid \mathbf{X}(t) = \mathbf{x}] - \widehat{\Phi}(\mathbf{x})\big) = 0.$$

We decompose the left-hand side of the equation above into three parts and obtain

$$0 = \sum_{\mathbf{x}:\ \Phi(\mathbf{x}) \le c - \nu_{\max}} \pi(\mathbf{x}) \big(\mathrm{E}[\widehat{\Phi}(\mathbf{X}(t+1)) \mid \mathbf{X}(t) = \mathbf{x}] - \widehat{\Phi}(\mathbf{x}) \big)$$

(65)
$$+ \sum_{\mathbf{x}:\ c - \nu_{\max} < \Phi(\mathbf{x}) \le c + \nu_{\max}} \pi(\mathbf{x}) \big(\mathrm{E}[\widehat{\Phi}(\mathbf{X}(t+1)) \mid \mathbf{X}(t) = \mathbf{x}] - \widehat{\Phi}(\mathbf{x}) \big)$$

$$+ \sum_{\mathbf{x}:\ c + \nu_{\max} < \Phi(\mathbf{x})} \pi(\mathbf{x}) \big(\mathrm{E}[\widehat{\Phi}(\mathbf{X}(t+1)) \mid \mathbf{X}(t) = \mathbf{x}] - \widehat{\Phi}(\mathbf{x}) \big).$$

As we will see below, the advantage of using the modified function $\widehat{\Phi}(\mathbf{x}) = \max\{c, \Phi(\mathbf{x})\}$ is that it allows us to exclude the states $\mathbf{x} \in \mathscr{X}$ with low values of $\Phi(\mathbf{x})$ from the equilibrium equation (64), without affecting the states $\mathbf{x}$ with high value of $\Phi(\mathbf{x})$.

We now analyze each of the summands in (65) separately.

(i) To analyze the first summand in (65), observe that for any $\mathbf{x}, \mathbf{x}'$ with $\Phi(\mathbf{x}) \le c - \nu_{\max} < c$ and $p(\mathbf{x}, \mathbf{x}') > 0$, the definition of $\nu_{\max}$ implies that

$$\Phi(\mathbf{x}') \le \Phi(\mathbf{x}) + \nu_{\max} \le c - \nu_{\max} + \nu_{\max} = c.$$

Therefore, $\widehat{\Phi}(\mathbf{x}) = \widehat{\Phi}(\mathbf{x}') = c$. We conclude that for all $\mathbf{x}$ with $\Phi(\mathbf{x}) \le c - \nu_{\max}$,

$$\mathrm{E}\big[\widehat{\Phi}(\mathbf{X}(t+1)) \mid \mathbf{X}(t) = \mathbf{x} \big] - \widehat{\Phi}(\mathbf{x}) = c - c = 0.$$

(ii) We now analyze the third summand in (65). For any $\mathbf{x}, \mathbf{x}'$ with $\Phi(\mathbf{x}) > c + \nu_{\max}$ and $p(\mathbf{x}, \mathbf{x}') > 0$, we again obtain from the definition of $\nu_{\max}$ that

$$\Phi(\mathbf{x}') \ge \Phi(\mathbf{x}) - \nu_{\max} > c + \nu_{\max} - \nu_{\max} = c.$$

Therefore, $\widehat{\Phi}(\mathbf{x}) = \Phi(\mathbf{x})$ and $\widehat{\Phi}(\mathbf{x}') = \Phi(\mathbf{x}')$. Also, by assumption, $c + \nu_{\max} \ge B$. We conclude that for all $\mathbf{x}$ with $\Phi(\mathbf{x}) > c + \nu_{\max}$,

$$\mathrm{E}\big[\widehat{\Phi}(\mathbf{X}(t+1)) \mid \mathbf{X}(t) = \mathbf{x} \big] - \widehat{\Phi}(\mathbf{x}) = \mathrm{E}\big[\Phi(\mathbf{X}(t+1) \mid \mathbf{X}(t) = \mathbf{x} \big] - \Phi(\mathbf{x}) \le -\gamma,$$

where the last inequality holds since $\Phi(\mathbf{x}) > B$ and Φ is a Lyapunov function with drift parameter γ.

(iii) Considering the four cases $\widehat{\Phi}(\mathbf{x}) = \widehat{\Phi}(\mathbf{x}') = c$, or $(\widehat{\Phi}(\mathbf{x}) = \Phi(\mathbf{x}), \widehat{\Phi}(\mathbf{x}') = c)$, or $(\widehat{\Phi}(\mathbf{x}) = c, \widehat{\Phi}(\mathbf{x}') = \Phi(\mathbf{x}'))$, or $(\widehat{\Phi}(\mathbf{x}) = \Phi(\mathbf{x}), \widehat{\Phi}(\mathbf{x}') = \Phi(\mathbf{x}'))$, we can easily check that for any two states $\mathbf{x}, \mathbf{x}'$, there are only the following two possibilities: either

$$0 \le \widehat{\Phi}(\mathbf{x}') - \widehat{\Phi}(\mathbf{x}) \le \Phi(\mathbf{x}') - \Phi(\mathbf{x})$$

or

$$\Phi(\mathbf{x}') - \Phi(\mathbf{x}) \le \widehat{\Phi}(\mathbf{x}') - \widehat{\Phi}(\mathbf{x}) \le 0.$$

Therefore, for any $\mathbf{x}$, we have

$$\mathrm{E}\big[\widehat{\Phi}(\mathbf{X}(t+1)) \mid \mathbf{X}(t) = \mathbf{x} =\big] - \widehat{\Phi}(\mathbf{x})$$

$$= \sum_{\mathbf{x}':\ \widehat{\Phi}(\mathbf{x}')>\widehat{\Phi}(\mathbf{x})} p(\mathbf{x}, \mathbf{x}')(\widehat{\Phi}(\mathbf{x}') - \widehat{\Phi}(\mathbf{x})) + \sum_{\mathbf{x}':\ \widehat{\Phi}(\mathbf{x}')\leq\widehat{\Phi}(\mathbf{x})} p(\mathbf{x}, \mathbf{x}')(\widehat{\Phi}(\mathbf{x}') - \widehat{\Phi}(\mathbf{x}))$$

$$\leq \sum_{\mathbf{x}':\ \Phi(\mathbf{x}')>\Phi(\mathbf{x})} p(\mathbf{x}, \mathbf{x}')(\Phi(\mathbf{x}') - \Phi(\mathbf{x}))$$

$$\leq \sum_{\mathbf{x}':\ \Phi(\mathbf{x}')>\Phi(\mathbf{x})} p(\mathbf{x}, \mathbf{x}')\nu_{\max}$$

$$\leq p_{\max}\nu_{\max}.$$

We conclude that for all $\mathbf{x} \in \mathscr{X}$, and, in particular, for all $\mathbf{x}$ satisfying $c - \nu_{\max} < \Phi(\mathbf{x}) \leq c + \nu_{\max}$, we have

$$\mathrm{E}\big[\widehat{\Phi}(\mathbf{X}(t+1))|\mathbf{X}(t) = \mathbf{x}\big] - \widehat{\Phi}(\mathbf{x}) \leq p_{\max}\nu_{\max}.$$

We now incorporate the analysis of these three cases into (65), to obtain

$$0 \leq 0 + \sum_{\mathbf{x}:\ c-\nu_{\max}<\Phi(\mathbf{x})\leq c+\nu_{\max}} \pi(\mathbf{x})p_{\max}\nu_{\max} + (-\gamma) \sum_{\mathbf{x}:\ c+\nu_{\max}<\Phi(\mathbf{x})} \pi(\mathbf{x}).$$

We then use the equality

$$\sum_{\mathbf{x}:\ c-\nu_{\max}<\Phi(\mathbf{x})\leq c+\nu_{\max}} \pi(\mathbf{x}) = \sum_{\mathbf{x}:\ c-\nu_{\max}<\Phi(\mathbf{x})} \pi(\mathbf{x}) - \sum_{\mathbf{x}:\ c+\nu_{\max}<\Phi(\mathbf{x})} \pi(\mathbf{x})$$

to replace the first sum of probabilities in the previous inequality, divide by $p_{\max}\nu_{\max} + \gamma$, and obtain (14). □

We now prove (15). Let us select an arbitrary $c \geq 0$. We again consider the function $\widehat{\Phi}(\mathbf{x}) = \max\{c, \Phi(\mathbf{x})\}$, and the equilibrium equation (63), which we rewrite as

$$(66) \quad \begin{aligned} 0 = &\sum_{\mathbf{x}:\ \Phi(\mathbf{x})<c-(1/2)\nu_{\min}} \pi(\mathbf{x})\big(\mathrm{E}[\widehat{\Phi}(\mathbf{X}(t+1)) \mid \mathbf{X}(t) = \mathbf{x}] - \widehat{\Phi}(\mathbf{x})\big) \\ &+ \sum_{\mathbf{x}:\ c-(1/2)\nu_{\min}\leq\Phi(\mathbf{x})<c} \pi(\mathbf{x})\big(\mathrm{E}[\widehat{\Phi}(\mathbf{X}(t+1))|\mathbf{X}(t) = \mathbf{x}] - \widehat{\Phi}(\mathbf{x})\big) \\ &+ \sum_{\mathbf{x}:\ c\leq\Phi(\mathbf{x})} \pi(\mathbf{x})\big(\mathrm{E}[\widehat{\Phi}(\mathbf{X}(t+1))|\mathbf{X}(t) = \mathbf{x}] - \widehat{\Phi}(\mathbf{x})\big). \end{aligned}$$

We now analyze each of the summands in the above identity.

(i) For any $\mathbf{x}$ with $\Phi(\mathbf{x}) \geq c$, we have $\widehat{\Phi}(\mathbf{x}) = \Phi(\mathbf{x})$. So,

$$\sum_{\mathbf{x}:\ \Phi(\mathbf{x})\geq c} \pi(\mathbf{x})\big(\mathrm{E}[\widehat{\Phi}(\mathbf{X}(t+1)) \mid \mathbf{X}(t) = \mathbf{x}] - \widehat{\Phi}(\mathbf{x})\big)$$

$$= \sum_{\mathbf{x}:\ \Phi(\mathbf{x})\geq c} \pi(\mathbf{x})\big(\mathrm{E}[\widehat{\Phi}(\mathbf{X}(t+1)) \mid \mathbf{X}(t) = \mathbf{x}] - \Phi(\mathbf{x})\big)$$

$$\geq \sum_{\mathbf{x}:\ \Phi(\mathbf{x})\geq c} \pi(\mathbf{x})\big(E[\Phi(\mathbf{X}(t+1))\mid \mathbf{X}(t)=\mathbf{x}]-\Phi(\mathbf{x})\big)$$

$$\geq (-\gamma) \sum_{\mathbf{x}:\ \Phi(\mathbf{x})\geq c} \pi(\mathbf{x}).$$

(ii) For any $\mathbf{x}$ with $\Phi(\mathbf{x}) < c - (1/2)\nu_{\min}$, we have $\widehat{\Phi}(\mathbf{x}) = c$. So,

$$\sum_{\mathbf{x}:\ \Phi(\mathbf{x})<c-(1/2)\nu_{\min}} \pi(\mathbf{x})\big(E[\widehat{\Phi}(\mathbf{X}(t+1))\mid \mathbf{X}(t)=\mathbf{x}]-\widehat{\Phi}(\mathbf{x})\big)$$

$$= \sum_{\mathbf{x}:\ \Phi(\mathbf{x})<c-(1/2)\nu_{\min}} \pi(\mathbf{x})\big(E[\widehat{\Phi}(\mathbf{X}(t+1))\mid \mathbf{X}(t)=\mathbf{x}]-c\big) \geq\ 0.$$

(iii) We next consider the case where $\mathbf{x}$ satisfies

$$(67) \qquad\qquad\qquad c - (1/2)\nu_{\min} \leq \Phi(\mathbf{x}) < c.$$

We have $\widehat{\Phi}(\mathbf{x}) = c$. If $\mathbf{x}'$ is such that $p(\mathbf{x},\mathbf{x}') > 0$ and $\Phi(\mathbf{x}') > \Phi(\mathbf{x})$, then $\Phi(\mathbf{x}') \geq \Phi(\mathbf{x}) + \nu_{\min} \geq c + (1/2)\nu_{\min}$. In particular, $\widehat{\Phi}(\mathbf{x}') = \Phi(\mathbf{x}')$ and

$$\widehat{\Phi}(\mathbf{x}') - \widehat{\Phi}(\mathbf{x}) \geq c + (1/2)\nu_{\min} - c = (1/2)\nu_{\min}.$$

If $p(\mathbf{x},\mathbf{x}') > 0$ and $\Phi(\mathbf{x}') \leq \Phi(\mathbf{x}) < c$, then $\widehat{\Phi}(\mathbf{x}') = c$, and $\widehat{\Phi}(\mathbf{x}') - \widehat{\Phi}(\mathbf{x}) = 0$. We conclude that for any $\mathbf{x}$ satisfying (67), we have

$$E\big[\widehat{\Phi}(\mathbf{X}(t+1))\mid \mathbf{X}(t)=\mathbf{x}\big] - \widehat{\Phi}(\mathbf{x})$$

$$= \sum_{\mathbf{x}':\ \Phi(\mathbf{x}')>\Phi(\mathbf{x})} p(\mathbf{x},\mathbf{x}')\big(\widehat{\Phi}(\mathbf{x}')-\widehat{\Phi}(\mathbf{x})\big) + \sum_{\mathbf{x}':\ \Phi(\mathbf{x}')\leq\Phi(\mathbf{x})} p(\mathbf{x},\mathbf{x}')\big(\widehat{\Phi}(\mathbf{x}')-\widehat{\Phi}(\mathbf{x})\big)$$

$$= \sum_{\mathbf{x}':\ \Phi(\mathbf{x}')>\Phi(\mathbf{x})} p(\mathbf{x},\mathbf{x}')\big(\widehat{\Phi}(\mathbf{x}')-\widehat{\Phi}(\mathbf{x})\big)$$

$$\geq p_{\min}(1/2)\nu_{\min}.$$

We incorporate our analysis of the three summands into (66), to obtain

$$0 \geq \sum_{\mathbf{x}:\ c-(1/2)\nu_{\min}\leq\Phi(\mathbf{x})<c} \pi(\mathbf{x})(1/2)p_{\min}\nu_{\min} + (-\gamma) \sum_{\mathbf{x}:\ c\leq\Phi(\mathbf{x})} \pi(\mathbf{x}).$$

We also have

$$\sum_{\mathbf{x}:\ c-(1/2)\nu_{\min}\leq\Phi(\mathbf{x})<c} \pi(\mathbf{x}) = \sum_{\mathbf{x}:\ c-(1/2)\nu_{\min}\leq\Phi(\mathbf{x})} \pi(\mathbf{x}) - \sum_{\mathbf{x}:\ c\leq\Phi(\mathbf{x})} \pi(\mathbf{x}).$$

Using this and dividing the previous inequality by $(1/2)p_{\min}\nu_{\min}+\gamma$, we obtain (15). $\square$

APPENDIX B

PROOF OF PROPOSITION 7.   For any $\mathbf{x} \in Z_+^N$ we denote the vector

$$\sum_{i=1}^I \lambda_i \mathbf{e}_{i,1} + \sum_{i,k} \mu_{i,k} w_{i,k}(\mathbf{x})(\mathbf{e}_{i,k+1} - \mathbf{e}_{i,k})$$

by $\Delta(\mathbf{x}, w)$, where $\mathbf{e}_{i,k}$ is the i, kth unit vector. Let us also denote

$$\mathrm{E}\big[\Phi_s(\mathbf{Q}(\tau_{s+1}))|\mathbf{Q}(\tau_s) = \mathbf{x}\big] - \Phi_s(\mathbf{x})$$
$$= \sum_{i=1}^I \lambda_i\big(\Phi_s(\mathbf{x}+\mathbf{e}_{i,1}) - \Phi_s(\mathbf{x})\big) + \sum_{i,k} \mu_{i,k} w_{i,k}(\mathbf{x})\big(\Phi_s(\mathbf{x}+\mathbf{e}_{i,k+1} - \mathbf{e}_{i,k}) - \Phi_s(\mathbf{x})\big),$$

by $\Delta\Phi_s(\mathbf{x}, w)$. We need to show that the function Φ_s satisfies

$$(68) \qquad\qquad\qquad \Delta\Phi_s(\mathbf{x}, w) \le -(1/4)\gamma,$$

for all $\mathbf{x} \in Z_+^N$ such that $\Phi_s(\mathbf{x}) \ge B$.

LEMMA 3.   *Suppose a nonzero vector* $\mathbf{x} \in Z_+^N$ *satisfies* $\sum_{(i,k) \in \sigma_{j_0}} x_{i,k} = 0$, *for some station* σ_{j_0} *Then,*

$$(69) \qquad\qquad\qquad \mathrm{P}_u\big\{\Phi_0(\mathbf{Z}_1, \ldots, \mathbf{Z}_J, \mathbf{x}) = \mathbf{Z}_{j_0}'\mathbf{x}\big\} = 0.$$

PROOF.   Since $\sum_{(i,k)\in\sigma_{j_0}} x_{i,k} = 0$, then for any $\mathbf{z}_{j_0} \in O_{j_0}$, (37) yields

$$(70) \qquad \mathbf{z}_{j_0}'\mathbf{x} = \sum_{(i,k)\notin\sigma_{j_0}} z_{i,k}^{j_0} x_{i,k} \le \sum_{(i,k)\notin\sigma_{j_0}} L_{i,k}^{j_0} x_{i,k} + \frac{1}{2J}\gamma \sum_{(i,k)\notin\sigma_{j_0}} x_{i,k}.$$

We have from the third constraint of GLP[dm] that for all $(i,k) \notin \sigma_{j_0}$,

$$(71) \qquad\qquad\qquad L_{i,k}^{j_0} x_{i,k} \le \frac{1}{J-1} \sum_{j \ne j_0} L_{i,k}^j x_{i,k}.$$

Now consider any $(i,k) \notin \sigma_{j_0}$. Note from (37), that for j such that $(i,k) \notin \sigma_j$ and for any $\mathbf{z}_j \in O_j$

$$L_{i,k}^j x_{i,k} \le z_{i,k}^j x_{i,k},$$

where $z_{i,k}^j$ denotes the i, kth component of $\mathbf{z}_j$. Moreover, for j such that $(i,k) \in \sigma_j$ and for any $\mathbf{z}_j \in O_j$,

$$L_{i,k}^j x_{i,k} \le z_{i,k}^j x_{i,k} - \frac{J-1}{2J}\gamma x_{i,k}.$$

Combining the two inequalities above we obtain that for every $(i, k) \notin \sigma_{j_0}$,

$$
\begin{aligned}
(72) \quad \frac{1}{J-1} \sum_{j \neq j_0} L_{i, k}^j x_{i, k} &\leq \frac{1}{J-1} \left(\sum_{j \neq j_0} z_{i, k}^j x_{i, k} \right) - \frac{1}{J-1} \frac{J-1}{2J} \gamma x_{i, k} \\
&= \frac{1}{J-1} \sum_{j \neq j_0} z_{i, k}^j x_{i, k} - \frac{1}{2J} \gamma x_{i, k}.
\end{aligned}
$$

Then from (70), (71) and (72), we obtain

$$
\mathbf{z}_{j_0}' \mathbf{x} \leq \frac{1}{J-1} \sum_{(i, k) \notin \sigma_{j_0}} \sum_{j \neq j_0} z_{i, k}^j x_{i, k}.
$$

By assumption $\sum_{(i, k) \in \sigma_{j_0}} x_{i, k} = 0$, so

$$
\sum_{(i, k) \notin \sigma_{j_0}} \sum_{j \neq j_0} z_{i, k}^j x = \sum_{j \neq j_0} \mathbf{z}_j' \mathbf{x}.
$$

We conclude

$$
\mathbf{z}_{j_0}' \mathbf{x} \leq \frac{1}{J-1} \sum_{j \neq j_0} \mathbf{z}_j' \mathbf{x}
$$

for all $\mathbf{z}_j \in O_j$, $j = 1, 2, \ldots, J$. But the event

$$
\Phi_0(\mathbf{Z}_1, \ldots, \mathbf{Z}_J, \mathbf{x}) = \mathbf{Z}_{j_0}' \mathbf{x}
$$

implies

$$
\mathbf{Z}_{j_0}' \mathbf{x} \geq \frac{1}{J-1} \sum_{j \neq j_0} \mathbf{Z}_j' \mathbf{x}.
$$

Therefore, the event $\Phi_0(\mathbf{Z}_1, \ldots, \mathbf{Z}_J, \mathbf{x}) = \mathbf{Z}_{j_0}' \mathbf{x}$ implies

$$
Z_{j_0}' \mathbf{x} = \frac{1}{J-1} \sum_{j \neq j_0} Z_j' \mathbf{x}.
$$

The probability of this event is zero. □

LEMMA 4. *If* $\mathbf{z}_j \in O_j$ *and* $\mathbf{x} \in \mathscr{D}_+^N$ *satisfies* $\sum_{(i, k) \in \sigma_j} x_{i, k} > 0$, *then for any Markovian nonidling policy* w,

$$
\mathbf{z}_j' \Delta(\mathbf{x}, w) \leq -\tfrac{1}{2} \gamma.
$$

PROOF.   We have

$$\mathbf{z}'_j \Delta(\mathbf{x}, w) = \mathbf{L}'_j \Delta(\mathbf{x}, w) + (\mathbf{z}_j - \mathbf{L}_j)' \Delta(\mathbf{x}, w)$$

$$= \mathbf{L}'_j \Delta(\mathbf{x}, w) + \sum_{i=1}^{I} \lambda_i (z^j_{i,1} - L^j_{i,1})$$

$$+ \sum_{i,k} \mu_{i,k} w_{i,k}(\mathbf{x})\big((z^j_{i,k+1} - L^j_{i,k+1}) - (z^j_{i,k} - L^j_{i,k})\big).$$

From GLP[dm], it can be shown that since $\sum_{(i,k)\in\sigma_j} x_{i,k} > 0$, then $\mathbf{L}'_j \Delta(\mathbf{x}, w) \leq -\gamma$, (for details see [10]). Also since $L^j_{i,k} \leq z^j_{i,k} \leq L^j_{i,k} + (1/2)\gamma$ and, by assumption,

$$\sum_{i=1}^{I} \lambda_i + \sum_{i,k} \mu_{i,k} = 1,$$

then

$$\sum_{i=1}^{I} \lambda_i (z^j_{i,1} - L^j_{i,1}) + \sum_{i,k} \mu_{i,k} w_{i,k}(\mathbf{x})\big((z^j_{i,k+1} - L^j_{i,k+1}) - (z^j_{i,k} - L^j_{i,k})\big) \leq (1/2)\gamma.$$

This completes the proof. □

We now return to the proof of Proposition 7. Recall our convention $\mathbf{e}_{i,0} = \mathbf{e}_{i,J_i+1} = 0$. Fix $\mathbf{x} \in Z^N_+$. Denote by E_j, $j = 1, 2, \ldots, J$ the event

$$\Phi_0(\mathbf{Z}_1, \ldots, \mathbf{Z}_J, \mathbf{x} + \mathbf{e}_{i,k+1} - \mathbf{e}_{i,k}) = \mathbf{Z}'_j(\mathbf{x} + \mathbf{e}_{i,k+1} - \mathbf{e}_{i,k})$$

$$\text{for all } i = 1, 2, \ldots, I, k = 0, 1, \ldots, J_i,$$

and by E_0 the remaining event that none of E_j occurs. Note that the probability that two of the events E_{j_1} and E_{j_2} occurring simultaneously is equal to zero.

We rewrite the left-hand side of (68) as

$$\Delta\Phi_s(\mathbf{x}, w) = E_u\Bigg[\sum_{i=1}^{I} \lambda_i\big(\Phi_0(\mathbf{Z}_1, \ldots, \mathbf{Z}_J, \mathbf{x} + \mathbf{e}_{i,1}) - \Phi_0(\mathbf{Z}_1, \ldots, \mathbf{Z}_J, \mathbf{x})\big)$$

(73)
$$+ \sum_{i,k} \mu_{i,k} w_{i,k}(\mathbf{x})\big(\Phi_0(\mathbf{Z}_1, \ldots, \mathbf{Z}_J, \mathbf{x} + \mathbf{e}_{i,k+1} - \mathbf{e}_{i,k})$$

$$- \Phi_0(\mathbf{Z}_1, \ldots, \mathbf{Z}_J, \mathbf{x})\big)\Bigg],$$

and condition on $E_1, E_2, \ldots, E_J$ and E_0. We have

$$
(74) \quad
\begin{aligned}
E_u &\left[\sum_{i=1}^{I} \lambda_i \big(\Phi_0(\mathbf{Z}_1, \ldots, \mathbf{Z}_J, \mathbf{x} + \mathbf{e}_{i,1}) - \Phi_0(\mathbf{Z}_1, \ldots, \mathbf{Z}_J, \mathbf{x}) \big) \right. \\
&+ \sum_{i,k} \mu_{i,k} w_{i,k}(\mathbf{x}) \big(\Phi_0(\mathbf{Z}_1, \ldots, \mathbf{Z}_J, \mathbf{x} + \mathbf{e}_{i,k+1} - \mathbf{e}_{i,k}) \\
&\left. \qquad - \Phi_0(\mathbf{Z}_1, \ldots, \mathbf{Z}_J, \mathbf{x}) \big) \Big| E_j \right] P_u\{E_j\} \\
&= E_u \big[\mathbf{Z}'_j \Delta(\mathbf{x}, w) \big| E_j \big] P_u\{E_j\}.
\end{aligned}
$$

From Lemma 3, if $\sum_{i \in \sigma_j} x_i = 0$, then $P_u\{\Phi_0(\mathbf{Z}_1, \ldots, \mathbf{Z}_J, \mathbf{x}) = \mathbf{Z}'_j \mathbf{x}\} = 0$ and as a result, $P_u\{E_j\} = 0$. Otherwise, from Lemma 4,

$$
(75) \quad E_u \big[\mathbf{Z}'_j \Delta(\mathbf{x}, w) \big| E_j \big] P_u\{E_j\} \leq (-1/2)\gamma P_u\{E_j\}.
$$

Combining (73), (74) and (75) we obtain

$$
(76) \quad
\begin{aligned}
\Delta \Phi_s(\mathbf{x}, w) \leq &-(1/2)\gamma \sum_{j=1}^{J} P_u\{E_j\} \\
&+ E_u \left[\sum_{i=1}^{I} \lambda_i \big(\Phi_0(\mathbf{Z}_1, \ldots, \mathbf{Z}_J, \mathbf{x} + \mathbf{e}_{i,1}) - \Phi_0(\mathbf{Z}_1, \ldots, \mathbf{Z}_J, \mathbf{x}) \big) \right. \\
&+ \sum_{i,k} \mu_{i,k} w_{i,k}(\mathbf{x}) \big(\Phi_0(\mathbf{Z}_1, \ldots, \mathbf{Z}_J, \mathbf{x} + \mathbf{e}_{i,k+1} - \mathbf{e}_{i,k}) \\
&\left. \qquad - \Phi_0(\mathbf{Z}_1, \ldots, \mathbf{Z}_J, \mathbf{x}) \big) \Big| E_0 \right] P_u\{E_0\}.
\end{aligned}
$$

We now analyze the last term in the sum.

LEMMA 5. *For any $i = 1, 2, \ldots, I$, $k = 0, 1, 2, \ldots, J_i + 1$, and $\mathbf{x} \in Z_+^N$, there holds*

$$
\big| \Phi_0(\mathbf{z}_1, \ldots, \mathbf{z}_J, \mathbf{x} + \mathbf{e}_{i,k+1} - \mathbf{e}_{i,k}) - \Phi_0(\mathbf{z}_1, \ldots, \mathbf{z}_J, \mathbf{x}) \big| \leq L_{\max} + (1/2)\gamma.
$$

In particular, $\nu_{\max} \leq L_{\max} + (1/2)\gamma$.

PROOF. Let $\Phi_0(\mathbf{z}_1, \ldots, \mathbf{z}_J, \mathbf{x}) = \mathbf{z}'_{j_1} \mathbf{x}$ and $\Phi_0(\mathbf{z}_1, \ldots, \mathbf{z}_J, \mathbf{x} + \mathbf{e}_{i,k+1} - \mathbf{e}_{i,k}) = \mathbf{z}'_{j_2}(\mathbf{x} + \mathbf{e}_{i,k+1} - \mathbf{e}_{i,k})$. Then

$$
\begin{aligned}
\Phi_0(\mathbf{z}_1, \ldots, \mathbf{z}_J, &\mathbf{x} + \mathbf{e}_{i,k+1} - \mathbf{e}_{i,k}) - \Phi_0(\mathbf{z}_1, \ldots, \mathbf{z}_J, \mathbf{x}) \\
&= \mathbf{z}'_{j_2}(\mathbf{x} + \mathbf{e}_{i,k+1} - \mathbf{e}_{i,k}) - \mathbf{z}'_{j_1} \mathbf{x} \\
&\leq \mathbf{z}'_{j_2}(\mathbf{x} + \mathbf{e}_{i,k+1} - \mathbf{e}_{i,k}) - \mathbf{z}'_{j_2} \mathbf{x} \leq z^{j_2}_{i,k+1} \leq L_{\max} + (1/2)\gamma.
\end{aligned}
$$

Similarly, we show that

$$\Phi_0\big(\mathbf{z}_1, \ldots, \mathbf{z}_J, \mathbf{x} + \mathbf{e}_{i,\,k+1} - \mathbf{e}_{i,\,k}\big) - \Phi_0(\mathbf{z}_1, \ldots, \mathbf{z}_J, \mathbf{x}) \geq -(L_{\max} + (1/2)\gamma).$$

This completes the proof. $\square$

Applying Lemma 5 to (76), and using $\sum_i \lambda_i + \sum_{i,\,k} \mu_{i,\,k} w_{i,\,k}(\mathbf{x}) \leq 1$, we obtain

$$\Delta\Phi_s(\mathbf{x}, w) \leq -(1/2)\gamma \sum_{j=1}^J \mathrm{P}_u\{\mathrm{E}_j\} + (L_{\max} + (1/2)\gamma)\mathrm{P}_u\{\mathrm{E}_0\}$$

(77)
$$= -(1/2)\gamma \sum_{j=0}^J \mathrm{P}_u\{\mathrm{E}_j\} + (L_{\max} + \gamma)\mathrm{P}_u\{\mathrm{E}_0\}$$

$$= -(1/2)\gamma + (L_{\max} + \gamma)\mathrm{P}_u\{\mathrm{E}_0\}.$$

To complete the proof of the proposition, we prove the following lemma.

LEMMA 6. *For any* $\mathbf{x}$ *satisfying* $\Phi_s(\mathbf{x}) > B$, *there holds*

$$\mathrm{P}_u\{\mathrm{E}_0\} \leq (1/4)\frac{\gamma}{L_{\max} + \gamma}.$$

PROOF. We have

(78)
$$\mathrm{P}_u\{\mathrm{E}_0\} \leq \sum_{j_1 \neq j_2} \mathrm{P}\{\mathrm{E}_{j_1 j_2}\},$$

where $\mathrm{E}_{j_1,\,j_2}$ denotes the event
$\Phi_0(\mathbf{Z}_1, \ldots, \mathbf{Z}_J, \mathbf{x}) = \mathbf{Z}'_{j_1}\mathbf{x}, \Phi_0(\mathbf{Z}_1, \ldots, \mathbf{Z}_J, \mathbf{x} + \mathbf{e}_{i,\,k+1} - \mathbf{e}_{i,\,k}) = \mathbf{Z}'_{j_2}(\mathbf{x} + \mathbf{e}_{i,\,k+1} - \mathbf{e}_{i,\,k})$, for some i, k. Now fix j_1, j_2 with $j_2 \neq j_1$. Note that the event $\Phi_0(\mathbf{Z}_1, \ldots, \mathbf{Z}_J, \mathbf{x}) = \mathbf{Z}'_{j_1}\mathbf{x}, \Phi_0(\mathbf{Z}_1, \ldots, \mathbf{Z}_J, \mathbf{x} + \mathbf{e}_{i,\,k+1} - \mathbf{e}_{i,\,k}) = \mathbf{Z}'_{j_2}(\mathbf{x} + \mathbf{e}_{i,\,k+1} - \mathbf{e}_{i,\,k})$ implies

$$\mathbf{Z}'_{j_2}\mathbf{x} - \mathbf{Z}'_{j_1}\mathbf{x} \leq 0$$

and

$$\begin{aligned}
\mathbf{Z}'_{j_2}\mathbf{x} - \mathbf{Z}'_{j_1}\mathbf{x} &= \mathbf{Z}'_{j_2}(\mathbf{x} + \mathbf{e}_{i,\,k+1} - \mathbf{e}_{i,\,k}) - \mathbf{Z}'_{j_1}\mathbf{x} + Z^{j_2}_{i,\,k} - Z^{j_2}_{i,\,k+1} \\
&\geq \mathbf{Z}'_{j_1}(\mathbf{x} + \mathbf{e}_{i,\,k+1} - \mathbf{e}_{i,\,k}) - \mathbf{Z}'_{j_1}\mathbf{x} + Z^{j_2}_{i,\,k} - Z^{j_2}_{i,\,k+1} \\
&= Z^{j_1}_{i,\,k+1} - Z^{j_1}_{i,\,k} + Z^{j_2}_{i,\,k} - Z^{j_2}_{i,\,k+1} \\
&\geq -2(L_{\max} + (1/2)\gamma).
\end{aligned}$$

Therefore the probability of the event $\mathrm{E}_{j_1,\,j_2}$ is bounded by the probability of the event

$$-2(L_{\max} + (1/2)\gamma) \leq \mathbf{Z}'_{j_2}\mathbf{x} - \mathbf{Z}'_{j_1}\mathbf{x} \leq 0.$$

Fix any (i, k) such that $x_{i,\,k} > 0$. Since $Z^{j_1}_{i,\,k}$ is chosen uniformly from $[L^{j_1}_{i,\,k}, L^{j_1}_{i,\,k} + \frac{1}{2J}\gamma]$ or $[L^{j_1}_{i,\,k} + \frac{J-1}{2J}\gamma, L^{j_1}_{i,\,k} + \frac{1}{2}\gamma]$ [depending on whether $(i, k) \in \sigma_{j_1}$ or

not] and both intervals have the same length $\frac{\gamma}{2J}$, then by conditioning on $\mathbf{Z}_{j_2}$ and $Z_{i',k'}^{j_1}$, $(i',k') \neq (i,k)$ we have

$$
P_u\left\{-2(L_{\max} + (1/2)\gamma) \leq \mathbf{Z}_{j_2}'\mathbf{x} - \mathbf{Z}_{j_1}'\mathbf{x} \leq 0 \mid \mathbf{Z}_{j_2}, Z_{i',k'}^{j_1}, (i',k') \neq (i,k)\right\}
$$

$$
\leq \frac{\frac{2(L_{\max}+(1/2)\gamma)}{x_{i,k}}}{\frac{\gamma}{2J}} = 2J\frac{2(L_{\max}+(1/2)\gamma)}{x_{i,k}\gamma}.
$$

Unconditioning, we obtain

$$
x_{i,k}P_u\{E_{j_1 j_2}\} \leq \frac{4J(L_{\max}+(1/2)\gamma)}{\gamma},
$$

for all (i,k) such that $x_{i,k} > 0$. Note that this inequality holds trivially if $x_{i,k} = 0$. Then, for any j and $\mathbf{z}_j \in O_j$, there holds

$$
\mathbf{z}_j'\mathbf{x}P_u\{E_{j_1,j_2}\} \leq \sum_{i,k} z_{i,k}^j \frac{4J(L_{\max}+(1/2)\gamma)}{\gamma}
$$

$$
\leq N(L_{\max}+(1/2)\gamma)\frac{4J(L_{\max}+(1/2)\gamma)}{\gamma}
$$

$$
= \frac{4NJ(L_{\max}+(1/2)\gamma)^2}{\gamma},
$$

where in the second inequality we used $z_{i,k}^j \leq L_{\max} + (1/2)\gamma$. Then

$$
\Phi_s(\mathbf{x})P_u\{E_{j_1,j_2}\} \leq \frac{4NJ(L_{\max}+(1/2)\gamma)^2}{\gamma}.
$$

From (78), we obtain

$$
P_u\{E_0\} \leq J(J-1)\frac{4NJ(L_{\max}+(1/2)\gamma)^2}{\gamma\Phi_s(\mathbf{x})}.
$$

Thus, whenever

$$
\Phi_s(\mathbf{x}) > \frac{16J^2(J-1)N(L_{\max}+(1/2)\gamma)^2(L_{\max}+\gamma)}{\gamma^2},
$$

there holds

$$
P_u\{E_0\} \leq (1/4)\frac{\gamma}{L_{\max}+\gamma}.
$$

Note that

$$
\frac{16J^2(J-1)N(L_{\max}+(1/2)\gamma)^2(L_{\max}+\gamma)}{\gamma^2} \leq \frac{16NJ^2(J-1)(L_{\max}+\gamma)^3}{\gamma^2}.
$$

This completes the proof of the lemma. $\square$

Applying Lemma 6 to (77), we obtain

$$\Delta\Phi_s(\mathbf{x}, w) \leq -(1/4)\gamma.$$

We have also obtained the bound on $\nu_{\max}$ in Lemma 5. This completes the proof of Proposition 7. □

PROOF OF PROPOSITION 8.   Given a solution $\mathbf{x}, \gamma$ to LP[dv], set $V = W = 0$. We have by construction $L_k^j - L_{k+1}^j = x_k$ if $k \in \sigma_j$, and $L_k^j - L_{k+1}^j = 0$ if $k \notin \sigma_j$. In particular, $L_k^j \geq L_{k+1}^j$, for all j, k. Since $V = W = 0$ then the constraints (31), (33) are satisfied.

Fix any $k \in \sigma_2$, and let $e \subset \sigma_2$ be the excursion containing k. Then trivially $l(e) \geq k$. It follows that

$$L_k^2 = \sum_{k' \in \sigma_2,\, k' \geq k} x_{k'} \geq \sum_{k' \in \sigma_2,\, k' \geq l(e)} x_{k'}.$$

However,

$$L_k^1 = \sum_{k' \in \sigma_1,\, k' \geq k} x_{k'},$$

which by constraint (44) is not bigger than $\sum_{k' \in \sigma_2,\, k' \geq l(e)} x_{k'}$. Combining, we conclude $L_k^2 \geq L_k^1$, whenever $k \in \sigma_2$. Similarly, we show that $L_k^2 \leq L_k^1$, whenever $k \in \sigma_1$. This implies that constraints (34) and (35) are satisfied. Finally, constraints (43) are equivalent to constraints (30) and (32) for $\gamma = \lambda\varepsilon$. Thus $\mathbf{L}_j = (L_1^j, \dots, L_N^j)$, $\gamma = \lambda\varepsilon$ is a feasible solution to LP[dm]. □

PROOF OF LEMMA 2.   Suppose $\rho^* < 1$. Choose any nontrivial strictly separating set S' and any nontrivial separating set S. We choose an excursion e_r by the following rules (rules (a) and (b), Section 6, [8]):

   (i) Every earlier excursion e_l, $l < r$ served at station σ_2 is in S'.
   (ii) Every earlier excursion e_l, $l < r$ served at station σ_1 is in S.

We have $\rho(V(S'), e_r) \leq \rho^*$ by the definition of ρ^*. This inequality can be rewritten, using (42), in the form

(79) $$\frac{\rho(V(S'), e_r, \sigma_1)}{1 - \rho(e_r, \sigma_1)} + \frac{\rho(V(S'), e_r, \sigma_2)}{1 - \rho(e_r, \sigma_2)} + \frac{1 - \rho^*}{(1 - \rho(e_r, \sigma_1))(1 - \rho(e_r, \sigma_2))} \leq 1.$$

By writing

$$1 = \frac{1 - \rho(e_r, \sigma_2)}{1 - \rho(e_r, \sigma_2)},$$

we obtain from above

$$\frac{\rho(V(S'), e_r, \sigma_1)}{1 - \rho(e_r, \sigma_1)} + \frac{1 - \rho^*}{(1 - \rho(e_r, \sigma_1))(1 - \rho(e_r, \sigma_2))} \leq \frac{1 - \rho(V(S'), e_r, \sigma_2) - \rho(e_r, \sigma_2)}{1 - \rho(e_r, \sigma_2)}$$

or

$$\frac{\rho(V(S'),e_r,\sigma_1)}{1-\rho(V(S'),e_r,\sigma_2)-\rho(e_r,\sigma_2)} + \frac{1-\rho^*}{(1-\rho(V(S'),e_r,\sigma_2)-\rho(e_r,\sigma_2))(1-\rho(e_r,\sigma_2))}$$
$$\leq \frac{1-\rho(e_r,\sigma_1)}{1-\rho(e_r,\sigma_2)}.$$

We now use rules 1 and 2 in the construction of e_r to show that

$$\rho(V(S'), e_r, \sigma_2) + \rho(e_r, \sigma_2) \geq \rho(V(S'), \sigma_2)$$

and

$$\rho(V(S'), e_r, \sigma_1) = \rho(V(S'), \sigma_1).$$

In fact, suppose $e_r \subset \sigma_1$. The inequality above then follows immediately from (40). Also by rule 1 and since S' is a separating set, none of the excursions $e_l, l \leq r$ that belong to σ_1 can also belong to S'. Therefore,

$$(80) \qquad \rho(V(S'), \sigma_1) = \sum_{k \in \sigma_1 \cap V(S),\, k > l(e_r)} \rho_k = \rho(V(S'), e_r, \sigma_1).$$

Suppose now $e_r \subset \sigma_2$. Let us show that e_r does not belong to S'. In fact, otherwise if $r < R$ (where again R is the index of the last excursion) then e_{r+1} would be an excursion satisfying rules 1 and 2, and if $r = R$ then S' would be a trivial separating set. Both lead to contradiction. Therefore e_r does not belong to S'. Then the last class $l(e_r)$ does not belong to $V(S)$ and the inequality follows. Note again that by rule 1 and since S' is a separating set, none of the excursions $e_l, l < r$ that belong to σ_1 can also belong to S'. The equality then follows. We proved (80). Note also, $1 - \rho(e_r, \sigma_2) \leq 1$. Therefore,

$$(81) \qquad \frac{\rho(V(S'), \sigma_1)}{1 - \rho(V(S'), \sigma_2)} + \frac{1 - \rho^*}{1 - \rho(V(S'), \sigma_2)} \leq \frac{1 - \rho(e_r, \sigma_1)}{1 - \rho(e_r, \sigma_2)}.$$

Writing (79) for the separating set S, we obtain

$$\frac{1 - \rho(V(S), e_r, \sigma_1) - \rho(e_r, \sigma_1)}{1 - \rho(e_r, \sigma_1)}$$
$$\geq \frac{1 - \rho^*}{(1 - \rho(e_r, \sigma_1))(1 - \rho(e_r, \sigma_2))} + \frac{\rho(V(S), e_r, \sigma_2)}{1 - \rho(e_r, \sigma_2)}$$

or

$$\frac{1 - \rho(V(S), e_r, \sigma_1) - \rho(e_r, \sigma_1)}{\rho(V(S), e_r, \sigma_2)} \geq \frac{1 - \rho^*}{\rho(V(S), e_r, \sigma_2)(1 - \rho(e_r, \sigma_2))} + \frac{1 - \rho(e_r, \sigma_1)}{1 - \rho(e_r, \sigma_2)}.$$

Again, with our choice of e_r,

$$\rho(V(S), e_r, \sigma_1) + \rho(e_r, \sigma_1) \geq \rho(V(S), \sigma_1),$$
$$\rho(V(S), e_r, \sigma_2) = \rho(V(S), \sigma_2).$$

The proof is similar to the one for S'. Also, $1 - \rho(e_r, \sigma_2) \leq 1$. Therefore,

$$\frac{1 - \rho(V(S), \sigma_1)}{\rho(V(S), \sigma_2)} \geq \frac{1 - \rho^*}{\rho(V(S), \sigma_2)} + \frac{1 - \rho(e_r, \sigma_1)}{1 - \rho(e_r, \sigma_2)}.$$

By combining this inequality with (81), we obtain

$$(82) \quad \frac{1 - \rho(V(S), \sigma_1)}{\rho(V(S), \sigma_2)} - \frac{1 - \rho^*}{\rho(V(S), \sigma_2)} \geq \frac{\rho(V(S'), \sigma_1)}{1 - \rho(V(S'), \sigma_2)} + \frac{1 - \rho^*}{1 - \rho(V(S'), \sigma_2)}.$$

Note that the left-hand side of this inequality depends only on the separating set S, and the right-hand side depends only on the separating set S'. Therefore, there exists some β which is in between these two quantities for any S and S'. In particular, any such β satisfies (52). Also for any such β, from the inequality above, we obtain

$$1 - \rho^* \leq \beta(1 - \rho(V(S'), \sigma_2)) - \rho(V(S'), \sigma_1)$$

and

$$1 - \rho^* \leq (1 - \rho(V(S), \sigma_1)) - \beta\rho(V(S), \sigma_2).$$

By definition, $1 - \rho^* \leq 1 - \rho_{\sigma_j}$, $j = 1, 2$. We have proved the existence of β for which $\varepsilon(\beta) > 1 - \rho^*$. This completes the proof of the lemma. $\square$

PROOF OF THEOREM 5. We assume without loss of generality that $\lambda + \sum_{i=1}^{N} \mu_i = 1$. Combining Proposition 8 with Lemma 2, we conclude: if $\rho^* < 1$, then there exists a feasible solution of the form $\mathbf{L}, \mathbf{Q}, \gamma = \lambda\varepsilon$ to LP[dm] such that $\varepsilon \geq 1 - \rho^*$. Since, by assumption, class 1 is from station σ_1, then from the constraint (34) of LP[dm], $L_1^1 \geq L_1^2$. Using Proposition 8, we obtain for this solution $L_{\max} = \max_{1 \leq k, l \leq N}\{L_k^1, L_l^2\} = L_1^1$. From constraint (46) of LP[dv], $L_1^1 = \sum_{k \in \sigma_1} x_k = 1 - \varepsilon$. Since $\lambda \leq 1$, then

$$(83) \qquad L_{\max} + \tfrac{1}{2}\gamma = 1 - \varepsilon + \tfrac{1}{2}\lambda\varepsilon \leq \rho^* + \tfrac{1}{2}\lambda(1 - \rho^*)$$

and

$$L_{\max} + \gamma = 1 - \varepsilon + \lambda\varepsilon \leq \rho^* + \lambda(1 - \rho^*).$$

Therefore,

$$(84) \qquad \frac{(L_{\max} + (1/2)\gamma)^2}{\gamma} \leq \frac{(\rho^* + (1/2)\lambda(1 - \rho^*))^2}{\lambda(1 - \rho^*)}$$

and

$$(85) \qquad \frac{(L_{\max} + \gamma)^3}{\gamma^2} \leq \frac{(\rho^* + \lambda(1 - \rho^*))^3}{\lambda^2(1 - \rho^*)^2}.$$

We now show that

$$(86) \qquad \frac{L_{\max} + (1/2)\gamma}{L_{\max} + (3/4)\gamma} \leq \frac{\rho^* + (1/2)\lambda(1 - \rho^*)}{\rho^* + (3/4)\lambda(1 - \rho^*)}.$$

We have

$$1 - \frac{L_{\max} + (1/2)\gamma}{L_{\max} + (3/4)\gamma} = \frac{(1/4)\gamma}{L_{\max} + (3/4)\gamma}.$$

Since $\gamma = \lambda\varepsilon \geq \bar{\lambda}(1 - \rho^*)$, and since $L_{\max} + (3/4)\gamma = 1 - \varepsilon + (3/4)\lambda\varepsilon = 1 - (1 - (3/4)\lambda))\varepsilon \leq 1 - (1 - (3/4)\lambda))(1 - \rho^*) = \rho^* + \frac{3}{4}\lambda(1 - \rho^*)$, then

$$\frac{(1/4)\gamma}{L_{\max} + (3/4)\gamma} \geq \frac{(1/4)\bar{\lambda}(1 - \rho^*)}{\rho^* + (3/4)\bar{\lambda}(1 - \rho^*)}.$$

Then (86) follows immediately. From (49) and (50) we obtain that

$$L_i^1 = \sum_{k \in \sigma_1,\, k \geq i} x_i \geq \sum_{k \in \sigma_1,\, k \geq i} \rho_k = \rho_i^{\sigma_1+}$$

and

$$L_i^2 \geq \beta\rho_i^{\sigma_2+}.$$

Applying Theorem 4 to L_1, we obtain

$$P_\pi\left\{\frac{\sum_{i=1}^N \rho_i^{\sigma_1+} Q_i(t) - B_0}{2(\rho^* + (1/2)\lambda(1 - \rho^*))} \geq m\right\} \leq \left(\frac{\rho^* + (1/2)\lambda(1 - \rho^*)}{\rho^* + (3/4)\lambda(1 - \rho^*)}\right)^m,$$

where

$$B_0 = \frac{64N(\rho^* + \lambda(1 - \rho^*))^3}{\lambda^2(1 - \rho^*)^2}$$

and

$$E_\pi\left[\sum_{i=1}^N \rho_i^{\sigma_1+} Q_i(t)\right] \leq \frac{64N(\rho^* + \lambda(1 - \rho^*))^3}{\lambda^2(1 - \rho^*)^2} + \frac{8(\rho^* + \frac{1}{2}\lambda(1 - \rho^*))^2}{\lambda(1 - \rho^*)}.$$

Since we have assumed without loss of generality $\lambda + \sum_{i=1}^N \mu_i = 1$,

$$\lambda = \frac{1}{1 + \sum_{i=1}^N \rho_i^{-1}}.$$

Substituting, we obtain the bounds (54) and (56).

Applying Theorem 4 to L_2, we obtain

$$P_\pi\left\{\frac{\sum_{i=1}^N \beta\rho_i^{\sigma_2+} Q_i(t) - B_0}{2(\rho^* + (1/2)\lambda(1 - \rho^*))} \geq m\right\} \leq \left(\frac{\rho^* + (1/2)\lambda(1 - \rho^*)}{\rho^* + (3/4)\lambda(1 - \rho^*)}\right)^m,$$

where again

$$B_0 = \frac{64N(\rho^* + \lambda(1 - \rho^*))^3}{\lambda^2(1 - \rho^*)^2}$$

and

$$E_\pi\left[\sum_{i=1}^N \beta\rho_i^{\sigma_2+} Q_i(t)\right] \leq \frac{64N(\rho^* + \lambda(1 - \rho^*))^3}{\lambda^2(1 - \rho^*)^2} + \frac{8(\rho^* + \frac{1}{2}\lambda(1 - \rho^*))^2}{\lambda(1 - \rho^*)}.$$

However, from (47), $\beta > \sum_{i \in \sigma_2} x_i$. Recall that by assumption $e_1 \subset \sigma_1$ and $e_2 \subset \sigma_2$. Therefore, from (44),

$$\sum_{i \in \sigma_2,\, i \geq l(e_2)} x_i \geq \sum_{i \in \sigma_1,\, i > l(e_2)} x_i.$$

But then from (49),

$$\sum_{i \in \sigma_1,\, i > l(e_2)} x_i \geq \sum_{i \in \sigma_1,\, i > l(e_2)} \rho_i = \rho_{l(e_2)+1}^{\sigma_1+}.$$

As a result, $\beta > \rho_{l(e_2)+1}^{\sigma_1+}$. By substituting β, with $\rho_{l(e_2)+1}^{\sigma_1+}$ in the bounds above, we obtain (55) and (57). This completes the proof of the theorem. $\square$

PROOF OF THEOREM 6.   We start with the following lemma.

LEMMA 7. $\overline{L}_j,\, \bar{\gamma} = \lambda(1 - \rho^*),\, V_j = 0,\, 1 \leq j \leq J$ is a feasible solution to GLP[dm].

PROOF.   It is straightforward to check that for $L_{i,k}^j = \rho_{i,k}^{\sigma_j+},\, V_j = 0,\, j = 1, 2, \ldots, J$, and $\gamma = \lambda_{\min}(1 - \rho^*)$ the constraints (25) and (26) are satisfied. Multiplying the constraints by

$$\left(\frac{\rho_{\min}}{(J - 1)\rho^*} \right)^{j-1}$$

we conclude that they are satisfied by $\overline{L}_{i,k}^j$ and $\bar{\gamma}$.

We now prove that the constraints (27) are satisfied. Select j and $(i, k) \notin \sigma_j$. Let $(i, k) \in \sigma_{j_1}$. If $j_1 > j$, then from the Definition 6, $\rho_{i,k}^{\sigma_j+} = 0$. Then $\overline{L}_{i,k}^j = 0$ and (27) is satisfied automatically due to the nonnegativity of the variables. Suppose now $j_1 < j$. We have, using $\rho_{\min} \leq \rho^*$,

$$\overline{L}_{i,k}^j = \left(\frac{\rho_{\min}}{(J-1)\rho^*} \right)^{j-1} \rho_{i,k}^{\sigma_j+} = \frac{1}{(J-1)^{j-j_1}} \left(\frac{\rho_{\min}}{\rho^*} \right)^{j-j_1} \left(\frac{\rho_{\min}}{(J-1)\rho^*} \right)^{j_1-1} \rho_{i,k}^{\sigma_j+}$$

$$\leq \frac{1}{J-1} \frac{\rho_{\min}}{\rho^*} \left(\frac{\rho_{\min}}{(J-1)\rho^*} \right)^{j_1-1} \rho_{i,k}^{\sigma_j+}.$$

However, $\rho_{\min} \leq \rho_{i,k}^{\sigma_{j_1}+}$ and $\rho_{i,k}^{\sigma_j+} \leq \rho^*$. It follows that

$$\frac{\rho_{\min}}{\rho^*} \rho_{i,k}^{\sigma_j+} \leq \rho_{i,k}^{\sigma_{j_1}+}.$$

But

$$\left(\frac{\rho_{\min}}{(J-1)\rho^*} \right)^{j_1-1} \rho_{i,k}^{\sigma_{j_1}+} = \overline{L}_{i,k}^{j_1}.$$

Combining,

$$\overline{L}_{i,\,k}^{j} \le \frac{1}{J-1}\overline{L}_{i,\,k}^{j_1}.$$

This shows that the constraint (27) is satisfied. $\square$

To prove the theorem, we apply Theorem 4. Note that $\max\{\overline{L}_{i,\,k}^{j}\} \le \rho^*$. The result is then obtained immediately.

Acknowledgment. We thank the reviewer of the paper for helpful comments.

REFERENCES

[1] BERTSIMAS, D. (1996). Lecture notes on stability of multiclass queueing networks. Unpublished manuscript.

[2] BERTSIMAS, D., GAMARNIK, D. and TSITSIKLIS, J. (1996). Stability conditions for multiclass fluid queueing networks. *IEEE Trans. Automat. Control* **41** 1618–1631.

[3] BERTSIMAS, D. and NINO-MORA, J. (1999). Optimization of multiclass queueing networks with changeover times via the achievable region approach II. The multiple station case. *Math. Oper. Res.* **24** 331–361.

[4] BERTSIMAS, D., PASCHALIDIS, I. and TSITSIKLIS, J. (1994). Optimization of multiclass queueing networks: polyhedral and nonlinear characterization of achievable performance. *Ann. Appl. Probab.* **4** 43–75.

[5] BRAMSON, M. (1994). Instability of FIFO queueing networks. *Ann. Appl. Probab.* **4** 414–431. [Correction (1994) **4** 952.]

[6] DAI, J. G. (1995). (1995). On the positive Harris recurrence for multiclass queueing networks: a unified approach via fluid models. *Ann. Appl. Probab.* **5** 49–77.

[7] DAI, J. G., HASENBEIN, J. J. and VANDE VATE, J. H. (1999). Stability of a three-station fluid network. *Queueing Systems* **33** 293–325.

[8] DAI, J. G. and VANDE VATE, J. H. (2001). The stability of two-station multi-type fluid networks. *Oper. Res.* **48** 721–744.

[9] DAI, J. G. and WEISS, G. (1996). Stability and instability of fluid models for certain re-entrant lines. *Math. Oper. Res.* **21** 115–134.

[10] DOWN, D. D. and MEYN, S. P. (1987). Piecewise linear test functions for stability and instability of queueing networks. *Queueing Systems* **27** 205–226.

[11] DOWN, D. D. and MEYN, S. P. (1995). Stability of acyclic multiclass queueing networks. *IEEE Trans. Automat. Control* **40** 916–920.

[12] DUPUIS, P. and WILLIAMS, R. J. (1994). Lyapunov functions for semimartingale reflecting Brownian motions. *Ann. Probab.* **22** 680–702.

[13] GELENBE, E. and MITRANI, L. (1980). *Analysis and Synthesis of Computer Systems*. Academic Press, London.

[14] HASENBEIN, J. (1997). Necessary conditions for global stability of multiclass queueing networks. *Oper. Res. Lett.* **21** 87–94.

[15] HUMES, C. JR., OU, J. and KUMAR, P. R. (1997). The delay of open Markovian queueing networks: uniform functional bounds, heavy traffic pole multiplicities, and stability. *Math. Oper. Res.* **22** 921–954.

[16] JIN, H., OU, J. and KUMAR, P. R. (1997). The throughput of irreducible closed Markovian queueing networks: functional bounds, asymptotic loss, efficiency, and the Harrison–Wein conjectures. *Math. Oper. Res.* **22** 886–920.

[17] KUMAR, P. R. (1993). Re-entrant lines. *Queueing Systems* **13** 87–110.

[18] KUMAR, S. and KUMAR, P. R. (1994). Performance bounds for queueing networks and scheduling policies. *IEEE Trans. Automat. Control* **AC-39** 1600–1611.

[19] KUMAR, P. R. and MEYN, S. P. (1995). Stability of queueing networks and scheduling policies. *IEEE Trans. Automat. Control* **40** 251–261.

[20] LIPPMAN, S. (1983). Applying a new device in the optimization of exponential queueing systems. *Oper. Res.* **21** 652–666.

[21] LU, S. H. and KUMAR, P. R. (1991). Distributed scheduling based on due dates and buffer priorities. *IEEE Trans. Automat. Control* **36** 1406–1416.

[22] MEYN, S. P. and TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer, London.

[23] RYBKO, A. N. and STOLYAR, A. L. (1992). On the ergodicity of stochastic processes describing open queueing networks. *Problemy Peredachi Informatsii* **28** 3–26.

[24] SEIDMAN, T. I. (1994). First come first serve can be unstable. *IEEE Trans. Automat. Control* **39** 2166–2170.

D. BERTSIMAS
SLOAN SCHOOL OF MANAGEMENT
  AND OPERATIONS RESEARCH CENTER
RM E53-363
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS 02139
E-MAIL: dbertsim@mit.edu

D. GAMARNIK
IBM T. J. WATSON RESEARCH CENTER
P.O. BOX 218
YORKTOWN HEIGHTS, NEW YORK 10598
E-MAIL: gamarnik@watson.ibm.com

J. N. TSITSIKLIS
LABORATORY FOR INFORMATION
  AND DECISION SCIENCES
  AND OPERATIONS RESEARCH CENTER
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS 02139
E-MAIL: jnt@mit.edu