

DUALITY AND LINEAR PROGRAMS FOR STABILITY AND PERFORMANCE ANALYSIS OF QUEUEING NETWORKS AND SCHEDULING POLICIES*†

P. R. Kumar and Sean Meyn‡

Abstract

We consider the problems of performance analysis and stability/instability determination of queueing networks and scheduling policies. We exhibit a strong duality relationship between the performance of a system, and its stability analysis via mean drift. We obtain a variety of linear programs to conduct such stability and performance analyses.

A certain LP, called the Performance LP, bounds the performance of all stationary non-idling scheduling policies. If it is bounded, then its dual, called the Drift LP, has a feasible solution, which is a copositive matrix. The quadratic form associated with this copositive matrix has a negative drift, allowing us to conclude that all stationary non-idling scheduling policies are stable in the very strong sense of having a geometrically converging exponential moment.

Some systems satisfy an auxiliary set of linear constraints. Examples are systems operating under some special scheduling policies such as buffer priority policies, or systems incorporating models of machine failures. Their performance is also bounded by a Performance LP, provided that they are stable, i.e., have a finite first moment for the number of parts. If the Performance LP is infeasible, then the system is unstable. Any feasible solution to the dual of the Performance LP provides a quadratic function with a negative drift. If this quadratic form is copositive, then the system is strongly stable as above. If not, the system is either unstable, or else is highly non-robust in that arbitrarily small perturbations can lead to an unstable system. One can use known algorithms to check the copositivity of the required matrix.

*Please address all correspondence to the first author.

†The research reported here has been supported in part by the National Science Foundation under Grant Nos. ECS-90-25007 and ECS-92-16487, and in part by the Joint Services Electronics Program under Contract No. N00014-90-J1270.

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

These results carry over to fluid models, allowing the study of networks with non-exponential distributions.

There is another LP test of stability, that avoids a copositivity check. If a modification of the Performance LP, called the Monotone LP, is bounded, then the system is stable. Moreover, the stability holds for all smaller arrival rates.

Finally, there is another modification of the Performance LP, called the Finite Time LP. It provides transient bounds on the performance of the system, from any initial condition.

1 Introduction

Queueing networks are a useful class of models in many application domains, e.g., communication networks [1], manufacturing systems [2], [3], and computer systems [4]. Control is typically exercised over such systems by the use of scheduling policies; see [5]. However, if one ventures outside a certain class of special systems for which the steady state distribution has a “product form,” very little is known concerning their performance, or even their stability; see Baskett, Chandy, Muntz and Palacios [6] and Kelly [7].

This paper is concerned with these two issues, the performance analysis of such systems, and the determination of their stability or instability. By performance analysis, we mean answers to questions of the following type: What is a lower bound on the feasible mean manufacturing lead time in a manufacturing system? Or, given a particular scheduling policy, what is its performance? By stability analysis, we mean the following: Given a scheduling policy, is the queueing network stable, i.e., does the mean number of parts in the system stay bounded?

We obtain the following results.

- (i) If the system is stable, i.e., the mean number of parts in the system is bounded, then the mean values of certain important random variables satisfy a set of linear equalities (Theorem 2). This result generalizes the results in Kumar [8], Bertsimas, Paschalidis and Tsitsiklis [9], and Kumar and Kumar [10], all of which require that the *second* moment on the number of parts be finite.
- (ii) The mean number of parts under any stationary non-idling scheduling policy is bounded above by a certain linear program, called the Performance LP, and below by the minimum of

the Performance LP (Theorem 4). This improves upon the results in Kumar and Kumar [10], since those do not allow one to draw any conclusion without first ascertaining the finiteness of the second moment.

(iii) Certain scheduling policies, such as buffer priority policies, or certain systems, satisfy an auxiliary set of linear constraints. If the corresponding Performance LP including these constraints is infeasible, then all the policies or systems captured by the constraints are unstable (Theorems 3 and 13).

(iv) We establish a surprising and strong duality relationship between the problems of performance analysis and stability analysis. The dual of the Performance LP, called the Drift LP, has the following property. Any feasible solution generates a quadratic form with a negative drift. (Theorems 5, 6 and 13). This Drift LP is less restrictive than the corresponding LP in Kumar and Meyn [11], since it does not require non-negativity of the coefficients of the quadratic form.

(v) Consider any quadratic form with a negative drift. If the weighting matrix is copositive, it is known from Kumar and Meyn [11] that the system then has a geometrically converging exponential moment (Theorem 7). However, if the matrix is not copositive, then either the system is unstable, or it is highly non-robust in that arbitrarily small perturbations can lead to instability (Theorem 10). Moreover, there is no Lipschitz Lyapunov function satisfying Foster's drift criterion. In particular, if the quadratic form is unbounded below on the communicating class of the Markov chain, then the system is unstable (Theorem 10). We note that the copositivity of a matrix can be checked by a finite algorithm (Theorem 11, due to Keller). These results improve upon the results in Kumar and Meyn [11] since they decouple the copositivity test from the LP to construct a quadratic form with negative drift.

(vi) If the Performance LP, for the class of all non-idling policies, is bounded, then the system has a geometrically converging exponential moment under all stationary non-idling scheduling policies (Theorem 12). This allows us to conclude a very strong form of stability directly from performance analysis. Moreover, every feasible solution of the dual Drift LP is a copositive matrix (Theorem 12).

(vii) If the dual Drift LP of the Performance LP, for systems with auxiliary constraints, has a feasible copositive solution, then the system has a geometrically converging exponential moment, and the mean number of parts in the system is bounded above and below by the maximum and minimum of the Performance LP. Otherwise, it is highly non-robust (Theorem 13).

(viii) To avoid a copositivity check for systems with auxiliary constraints, one can use a slight modification of the Performance LP, called the Monotone LP. If it is bounded, then the system is stable. Moreover, it is stable for all smaller arrival rates (Theorem 14).

(ix) To study the transient performance of the system, there is another modification of the Performance LP, called the Finite Time LP. It provides bounds on the transient behavior of the system, from any initial condition (Theorem 15).

(x) The above results carry over to fluid models. The fluid model is L_2 -stable if and only if there is a positive, nearly quadratic functional with a negative drift (Theorem 16). Hence, if the dual Drift LP is feasible, then the fluid model is stable if and only if the solution is copositive (Theorems 17 and 18). From the important recent work of Rybko and Stolyar [12], Dai [13], and Chen [14], it is now known that the stability of the fluid model implies the stability of queueing networks with renewal type arrival and service processes.

Finally, our stability results may be relevant in the study of Brownian motion approximations of queueing networks, as in Harrison [15]. For example, for the Dai and Wang [16] system that has been shown not to have a Brownian approximation, our results show that there is no quadratic form with a negative drift, for load factors near 1.

2 The Basic Open Re-Entrant Line

For simplicity, we present all the results in the context of a basic open re-entrant line. They can all be extended to more general systems, such as networks with probabilistic routes, etc., as studied in Kumar and Kumar [10].

Consider a system with S machines $\{1, 2, \dots, S\}$. Parts arrive as a Poisson process of rate λ . They first visit machine $\sigma(1) \in \{1, 2, \dots, S\}$. There they are stored in a buffer b_1

while awaiting service or being served. Then they proceed to machine $\sigma(2)$, where they are stored in a buffer b_2 , etc. Let buffer b_L at machine $\sigma(L)$ be the last buffer visited. The service times for parts in buffer b_i are exponentially distributed with mean $1/\mu_i$. All service times and interarrival times are independent. A machine can share its capacity between its buffers, and any service can be pre-empted, if the scheduling policy calls for it. We assume that the arrival rate is within the capacity of the system, i.e., $\sum_{\{i:\sigma(i)=\sigma\}} \lambda/\mu_i < 1$, for all σ .

Let $x_i(t)$ denote the number of parts in buffer b_i at time t , and $x(t) := (x_1(t), \dots, x_L(t))^T$ the “state” of the system. Let $w_i(t)$ be the fraction of machine $\sigma(i)$ ’s service capacity that is allocated to buffer b_i at time t , i.e., $w_i(t) \geq 0$ and $\sum_{\{i:\sigma(i)=\sigma\}} w_i(t) \leq 1$ for every σ . We suppose that all stochastic processes are right continuous with left hand limits.

We consider scheduling policies which are non-idling and stationary. By non-idling, we mean that a machine cannot stay idle if one of its buffers is nonempty, i.e., $\sum_{\{i:\sigma(i)=\sigma\}} w_i(t) = 1$, whenever $\sum_{\{i:\sigma(i)=\sigma\}} x_i(t) \geq 1$. By stationary, we mean that the decision on which buffer to serve at time t , i.e., the allocation of $w_i(t)$ ’s, is purely a function of $x(t)$.

We rescale time so that $\lambda + \sum_{i=1}^L \mu_i = 1$, and resort to uniformization; see Lippmann [17]. That is, we pretend that every buffer is always being served, and sample the system at the times $\{\tau_n\}$ corresponding to either arrivals, real service completions, or virtual service completions. We take $\tau_0 = 0$. Let \mathcal{F}_{τ_n} denote the σ -algebra of events up to time τ_n .

3 The Equality Constraints Implied by a Finite First Moment

Consider any scheduling policy that is non-idling and stationary. The resulting Markov chain $\{x(\tau_n)\}$ is aperiodic, since the system can stay at the origin for two consecutive time steps. It also has a single communicating class, say \mathcal{C} , since the origin is reachable from every state. However, it need not be irreducible. (A simple example where $\mathcal{C} \neq Z_+^L$ is a machine visited two times in immediate succession, under the LBFS policy described below in Section 9.1).

Let us say that the Markov chain has a *finite first moment*, if $\sup_n E|x(\tau_n)| < +\infty$, for initial conditions in \mathcal{C} . Above, by $|x(\tau_n)|$, we mean the ℓ_1 -norm, $\sum_{i=1}^L |x_i(\tau_n)|$.

We now show that if the first moment is finite, then the mean values of certain random variables satisfy a system of linear equations. These linear equations constrain the performance of the system, and are fundamental to the procedure for obtaining bounds on system performance in Kumar [8], Bertsimas, Paschalidis and Tsitsiklis [9], and Kumar and Kumar [10]. Our result here improves upon the corresponding results in [8, 9, 10], which all require a finite *second* moment, $\sup_n E|x(\tau_n)|^2 < +\infty$. This is important because it allows us to develop self-contained procedures to determine performance bounds, without the need for any additional analysis of stability, as we shall see in Theorem 4, and allows us to establish very strong stability properties, as we see in Theorem 12.

Theorem 1 *Consider any stationary, non-idling scheduling policy under which the resulting Markov chain has a finite first moment. Then, for $x(\tau_0) \in \mathcal{C}$,*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} E[x^T(\tau_{n+1})Qx(\tau_{n+1}) - x^T(\tau_n)Qx(\tau_n)] = 0 \text{ for every matrix } Q. \quad (1)$$

Proof. The LHS of (1) telescopes to $\frac{1}{N}E[x^T(\tau_N)Qx(\tau_N) - x^T(\tau_0)Qx(\tau_0)]$, and so it suffices to show that $\lim_{N \rightarrow \infty} \frac{1}{N}E|x(\tau_N)|^2 = 0$.

Define $V(x) := E[\sum_{n=0}^{\sigma} |x(\tau_n)| \mid x(\tau_0) = x]$, where $\tau_{\sigma} := \min\{\tau_n : n \geq 0 \text{ and } x(\tau_n) = 0\}$ is the first hitting time to the origin. We will see shortly that $V(x) < +\infty$ for all x in the communicating class \mathcal{C} . Regardless, from the one-step dynamic programming recursion, see Blackwell [18], it satisfies,

$$V(x) = |x| + \sum_y p_{xy}V(y) \text{ for } x \neq 0; \text{ with } V(0) = 0. \quad (2)$$

Here, p_{xy} is the transition probability of the chain. Note that if for $x \neq 0$, $V(x) < +\infty$, then $V(y) < +\infty$ for every y such that $p_{xy} > 0$; i.e., for every “neighbor” of x .

Let $\pi(x)$ be the steady-state probability of state x . Define $\tau_{\bar{\sigma}} := \min\{\tau_n : n \geq 1 \text{ and } x(\tau_n) = 0\}$, the first *entrance* time to state 0. From the reciprocal relationship between steady state probabilities and mean recurrence times, we have $E[\bar{\sigma} \mid x(\tau_0) = 0] = \frac{1}{\pi(0)}$. Note also that from the renewal relationship (see Theorem 7.5 of Ross [19]), $\sum_x \pi(x)|x| =$

$\frac{E[\sum_{n=0}^{\bar{\sigma}} |x(\tau_n)| | x(\tau_0)=0]}{E[\bar{\sigma} | x(\tau_0)=0]}$. From the one-step dynamic programming recursion, the numerator above satisfies $E[\sum_{n=0}^{\bar{\sigma}} |x(\tau_n)| | x(\tau_0) = 0] = \sum_y p_{0y} V(y)$.

Hence $\sum_x \pi(x)|x| = \pi(0) \sum_y p_{0y} V(y)$. Since the LHS is finite, so is the sum on the RHS. Hence, we see that $V(y) < +\infty$ for every neighbor of 0, i.e., every y with $p_{0y} > 0$. Moreover, every z , such that $p_{yz} > 0$ for such a y , also has $V(z) < +\infty$, as noted earlier. By induction on the neighbors, we conclude that $V(x) < +\infty$ for all $x \in \mathcal{C}$.

Now let $x(\tau_0) \in \mathcal{C}$ be arbitrary. Then, from the dynamic programming recursion (2), by accounting for both the cases $x(\tau_n) = 0$ and $x(\tau_n) \neq 0$, we get, $V(x(\tau_n)) = |x(\tau_n)| + \sum_y p_{x(\tau_n)y} V(y) - 1(x(\tau_n) = 0) \sum_y p_{0y} V(y)$. Also, $\sum_y p_{x(\tau_n)y} V(y) = E[V(x(\tau_{n+1})) | \mathcal{F}_{\tau_n}]$. By taking expectations, and then telescoping, we obtain $\frac{V(x(\tau_0)) - EV(x(\tau_N))}{N} = \frac{1}{N} \sum_{n=0}^{N-1} E|x(\tau_n)| - \frac{1}{N} \sum_{n=0}^{N-1} E[1(x(\tau_n) = 0)] \cdot \frac{1}{\pi(0)} \sum_x \pi(x)|x| \rightarrow 0$. Thus $\frac{EV(x(\tau_N))}{N} \rightarrow 0$.

Since $|x(\tau_n)|$ can decrease by at most 1 at each time step, the hitting time τ_σ to 0 from a state x satisfies $\sigma \geq |x|$, and so $V(x) \geq \frac{1}{2}|x|^2$. Hence $\frac{E|x(\tau_N)|^2}{N} \rightarrow 0$. \square

We can now show that any system with a finite first moment has its performance constrained by a set of linear equalities.

Theorem 2 *Consider any stationary non-idling scheduling policy with a finite first moment. Define the steady state values, $z_{ij} := \lim_{n \rightarrow \infty} E[w_j(\tau_n)x_i(\tau_n)]$, and $\bar{x}_i := \lim_{n \rightarrow \infty} E[x_i(\tau_n)]$. Then the z_{ij} 's satisfy the linear equalities:*

$$2\lambda\bar{x}_1 + 2\lambda - 2\mu_1 z_{11} = 0, \quad (3)$$

$$2\mu_{j-1} z_{j-1,j} + 2\lambda - 2\mu_j z_{j,j} = 0, \quad \text{for } 2 \leq j \leq L, \quad (4)$$

$$\lambda\bar{x}_2 - \lambda - \mu_1 z_{12} + \mu_1 z_{11}) - \mu_2 z_{21} = 0, \quad (5)$$

$$\lambda\bar{x}_j - \mu_1 z_{1j} - \mu_j z_{j1} + \mu_{j-1} z_{j-1,1} = 0, \quad \text{for } 3 \leq j \leq L, \quad (6)$$

$$\mu_{i-1} z_{i-1,i+1} - \mu_i z_{i,i+1} - \lambda + \mu_i z_{ii} - \mu_{i+1} z_{i+1,i} = 0, \quad \text{for } 2 \leq i \leq L-1, \text{ and } \quad (7)$$

$$\mu_{i-1} z_{i-1,j} - \mu_i z_{i,j} + \mu_{j-1} z_{j-1,i} - \mu_j z_{j,i} = 0, \quad \text{for } 2 \leq i \leq L-2 \text{ and } i+2 \leq j \leq L. \quad (8)$$

Proof. Consider any i, j . From Theorem 1, $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} E[x_i(\tau_{n+1})x_j(\tau_{n+1}) - x_i(\tau_n)x_j(\tau_n)] = 0$. Note that for $j \geq 2$ and $i \geq j+2$, $E[x_i(\tau_{n+1})x_j(\tau_{n+1}) \mid \mathcal{F}_{\tau_n}] - x_i(\tau_n)x_j(\tau_n) = w_{i-1}(\tau_n)x_j(\tau_n)$ with probability μ_{i-1} ; $= -w_i(\tau_n)x_j(\tau_n)$ with probability μ_i ; $= w_{j-1}(\tau_n)x_i(\tau_n)$ with probability μ_{j-1} ; $= -w_j(\tau_n)x_i(\tau_n)$ with probability μ_j ; and $= 0$ with probability $(1 - \mu_{i-1} - \mu_i - \mu_{j-1} - \mu_j)$. Similarly, other values of i, j can also be analyzed. One can thus evaluate $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} E[x_i(\tau_{n+1})x_j(\tau_{n+1}) - x_i(\tau_n)x_j(\tau_n) \mid \mathcal{F}_{\tau_n}]$. Then one can compute the unconditional expectation, utilizing also the additional relationship $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} E[w_i(\tau_n)] = \lambda/\mu_i$, necessary in order for a steady-state to hold. From this, as in [10] or [9], one obtains the equalities shown above. \square

4 Performance Bounds for Systems: The Performance LP

In addition to the above equality constraints, every non-idling policy with a finite first moment also satisfies an additional set of *non-idling inequalities*. Whenever $x_j(\tau_n) \geq 1$, machine $\sigma(j)$ is not allowed to stay idle; hence $\sum_{\{i:\sigma(i)=\sigma(j)\}} w_i(\tau_n) = 1$. Thus $\bar{x}_j = \lim_{n \rightarrow \infty} E[x_j(\tau_n)] = \lim_{n \rightarrow \infty} \sum_{\{i:\sigma(i)=\sigma(j)\}} E[w_i(\tau_n)x_j(\tau_n)] = \sum_{\{i:\sigma(i)=\sigma(j)\}} z_{ij}$, i.e.,

$$\bar{x}_j - \sum_{\{i:\sigma(i)=\sigma(j)\}} z_{ij} = 0, \text{ for } 1 \leq j \leq L. \quad (9)$$

However, for all other σ 's, $\sum_{\{i:\sigma(i)=\sigma\}} w_i(\tau_n) \leq 1$, and so $\sum_{\{i:\sigma(i)=\sigma\}} z_{ij} = \lim_{n \rightarrow \infty} \sum_{\{i:\sigma(i)=\sigma\}} E[w_i(\tau_n)x_j(\tau_n)] \leq \lim_{n \rightarrow \infty} E[x_j(\tau_n)] = \bar{x}_j$. Hence, the \bar{x}_j 's and the z_{ij} 's satisfy the *non-idling inequalities*,

$$\sum_{\{i:\sigma(i)=\sigma\}} z_{ij} - \bar{x}_j \leq 0, \text{ for } 1 \leq j \leq L, \text{ and all } \sigma \neq \sigma(j). \quad (10)$$

Frequently, systems also satisfy a set of *auxiliary constraints* even without an assumption on the finiteness of the first moment,

$$\limsup_{N \rightarrow \infty} \frac{1}{N} E \sum_{n=0}^{N-1} \sum_{ij} a_{ij}^{(k)} w_i(\tau_n) x_j(\tau_n) \leq b^{(k)}, \text{ for } 1 \leq k \leq K. \quad (11)$$

This occurs when restricting attention to a particular class of scheduling policies of interest, such as the buffer priority policies treated in Section 9.1, or from modeling GI arrivals, GI service times, closed queueing networks, machine failures, or even from prior analytical knowledge of system behavior. We refer the reader to Kumar and Kumar [10] for techniques on how to obtain such additional constraints.

If the constraints above are infeasible, then all systems captured by the constraints are unstable.

The following Theorem obtains performance bounds for such systems with or without such auxiliary constraints (where one can trivially set $a_{ij}^{(k)} = b^{(k)} = 0$ if there are none), subject to the restriction that the system have a finite first moment. It extends all earlier results in Kumar and Kumar [10].

Theorem 3 *Consider a system with a finite first moment, operating under any stationary non-idling scheduling policy and satisfying the auxiliary constraints (11). The mean number of parts in the system is bounded above by the linear program:*

$$\text{Max} \quad \sum_{j=1}^L \bar{x}_j \quad (12)$$

subject to the linear equality constraints (9,3,4,5,6,7,8), the non-idling inequality constraints (10), the auxiliary inequalities,

$$\sum_{ij} a_{ij}^{(k)} z_{ij} \leq b^{(k)} \text{ for } 1 \leq k \leq K, \quad (13)$$

and the non-negativity constraints,

$$z_{ij} \geq 0 \quad \text{for all } i, j. \quad (14)$$

It is bounded below by the same linear program with Max replaced by Min. If this program is infeasible, then all systems captured by the above constraints are unstable in that they do not have a finite first moment.

We shall call the above linear program (12,9,3,4,5,6,7,8,10,13,14), the *Performance LP*.

5 Performance and Stability of All Non-Idling Policies

Before invoking Theorem 3 for obtaining a bound on all the policies in a certain class, one needs to first show that all the policies in the class have a finite first moment. Naturally, the following question arises. If the Performance LP is bounded, then is it automatically true that all the policies in the class included have a finite first moment? We can indeed show this when one considers the class of all non-idling policies without any auxiliary constraints.

The reason is that the class of all non-idling stationary policies possesses two special properties. First, it always contains a policy with a finite first moment, and an irreducible state-space. Second, it contains composite policies that use one policy in one region of the state-space, and another policy in another region of the state-space. Thus, one can approximate any arbitrary policy by a composite policy which uses a stable policy whenever the state gets large.

To show the first point, consider a policy where machine $\sigma(i)$ allocates the fraction $\frac{1}{\mu_i} \left(\sum_{\{j:\sigma(j)=\sigma(i)\}} \frac{1}{\mu_j} \right)^{-1}$ of its capacity to buffer b_i , wasting this capacity if b_i is empty, i.e., $w_i(t) = \frac{1}{\mu_i} \left(\sum_{\{j:\sigma(j)=\sigma(i)\}} \frac{1}{\mu_j} \right)^{-1} 1(x_i(t) \geq 1)$. Due to the possible wastage of capacity, this policy is not non-idling. However, it is stable, since the re-entrant line behaves just like a tandem of L M/M/1 queues. Thus, the first moment is finite.

Now let us modify this policy by allocating any unused server capacity to all the non-empty buffers in proportion to their mean service times, i.e., $w_i(t) = \frac{1}{\mu_i} \left(\sum_{\{j:\sigma(j)=\sigma(i)\}} \frac{1(x_j(t) \geq 1)}{\mu_j} \right)^{-1}$ if $x_i(t) \geq 1$; and = 0 otherwise. Under this more efficient policy, all parts exit the system earlier than they would have under the unmodified policy. Hence, $E|x_i|$ is smaller in steady-state. Thus we have constructed a non-idling stationary policy with a finite first moment. Moreover, the Markov chain is clearly irreducible. Call this policy p .

Now consider any arbitrary stationary non-idling policy \bar{p} . Construct a composite policy \tilde{p}_N as follows: \tilde{p}_N takes the same action as \bar{p} whenever the number of parts in the system is no more than N ; otherwise it follows p .

Since we have only modified the stable policy p on a finite number of states, \tilde{p}_N also has a finite first moment. Let us suppose that the value of the Performance LP is $M < \infty$. By Theorem 3, it follows that the mean number in the system is less than M . This is moreover true irrespective of N , i.e., for all the policies $\{\tilde{p}_N : N \geq 1\}$. Thus, the invariant distributions of the policies in $\{\tilde{p}_N : N \geq 1\}$ are tight.

In addition, the transition probabilities for \tilde{p}_N converge uniformly on compact sets to those of \bar{p} , as $N \rightarrow \infty$. Hence the invariant distributions for the policies \tilde{p}_N converge weakly to the invariant distribution of \bar{p} , which consequently exists; see Theorem 6, pp. 157 of Kushner [20].

Moreover, due to the weak convergence, the mean number under \bar{p} is also bounded above by M .

Now note that the bound of the Performance LP is vacuous if it is infinity. Also, the lower bound obtained by replacing “Max” by “Min” in the Performance LP is vacuous if a mean value does not exist for the number of parts, since we then regard the mean number as infinity.

Thus we have obtained the following Theorem. It is a stand alone result, not requiring any separate stability check. It consequently improves upon the corresponding result in Kumar and Kumar [10].

Theorem 4 *The mean number of parts in the system for any stationary non-idling policy is bounded above by the value of the Performance LP (12,9,3,4,5,6,7,8,10,14). It is bounded below by the same LP, where we replace “Max” by “Min.”*

Later, in Theorem 12, we will considerably extend the above result by showing that if the Performance LP is bounded, then all moments, including an exponential one, exist, and converge geometrically.

6 Performance and Stability are Duals: The Duality of the Performance and Drift LPs

We now show the very interesting result that performance and stability are related by duality. Specifically, the dual of the Performance LP is an LP for which every feasible solution provides a quadratic function with negative drift. We accordingly call the dual as the *Drift LP*.

The Drift LP was developed in Kumar and Meyn [11], as a programmatic procedure for constructing a Lyapunov function with a negative drift. By recognizing this duality here, we deduce that whenever the Performance LP is bounded, the Drift LP has a feasible solution, i.e., there exists a quadratic function with a negative drift. We note that Fayolle [21] has earlier used general quadratic forms to characterize ergodicity of random walks on Z_+^n . We also refer the reader to Fayolle et al [22], and Botvich and Zamyatin [23], for some previous studies on the stability of networks.

One caution is necessary. The existence of a quadratic function with negative drift does not necessarily imply stability. We will show in the next section that the system may be stable, unstable, or stable but highly non-robust, depending on whether the quadratic form is positive on the state-space.

For simplicity, we will consider the class of all non-idling policies, with no auxiliary constraints. Let us associate the dual variables $p_j, r_{11}, r_{12}, r_{1j}, r_{jj}, r_{i,i+1}, r_{ij}, w_{\sigma,j}$, respectively, with the constraints (9,3,5,6,4,7,8,10). Note that the dual variables r_{ij} are defined only in the upper triangular block, i.e., for $j \geq i$. The dual LP, see Murty [24], with its dual variables shown in parentheses, is,

$$\text{Min} \quad \sum_{i=1}^{L-1} \lambda r_{i,i+1} - \sum_{i=1}^L \lambda r_{ii}$$

subject to,

$$\lambda r_{1i} + p_i - \sum_{\{\sigma:\sigma \neq \sigma(i)\}} w_{\sigma,i} \geq 1 \text{ for } 1 \leq i \leq L, (\bar{x}_i)$$

$$-\mu_j r_{jj} + \mu_j r_{j,j+1} - p_j \geq 0 \text{ for } 1 \leq j \leq L, (z_{jj})$$

$$- \mu_1 r_{1j} + \mu_1 r_{2,j} - p_j 1_{\{\sigma(1)=\sigma(j)\}} + w_{\sigma(1),j} 1_{\{\sigma(1)\neq\sigma(j)\}} \geq 0 \text{ for } 2 \leq j \leq L, (z_{1j}) \quad (15)$$

$$- \mu_j r_{1j} + \mu_j r_{1,j+1} - p_1 1_{\{\sigma(j)=\sigma(1)\}} + w_{\sigma(j),1} 1_{\{\sigma(j)\neq\sigma(1)\}} \geq 0 \text{ for } 2 \leq j \leq L, (z_{j1}) \quad (16)$$

$$\mu_{j-1} r_{jj} - \mu_{j-1} r_{j-1,j} - p_j 1_{\{\sigma(j-1)=\sigma(j)\}} + w_{\sigma(j-1),j} 1_{\{\sigma(j-1)\neq\sigma(j)\}} \geq 0 \text{ for } 2 \leq j \leq L, (z_{j-1,j}) \quad (17)$$

$$\begin{aligned} \mu_i r_{i+1,j} & 1_{\{j \geq i+2\}} - \mu_i r_{ij} 1_{\{j \geq i+1\}} + \mu_i r_{j,i+1} 1_{\{i \geq j\}} - \mu_i r_{ji} 1_{\{i \geq j+1\}} - p_j 1_{\{\sigma(i)=\sigma(j)\}} \\ & + w_{\sigma(i),j} 1_{\{\sigma(i)\neq\sigma(j)\}} \geq 0 \text{ for } i \neq 1, j \neq 1, j \neq i+1, i \neq j, (z_{ij}) \end{aligned} \quad (18)$$

$$w_{\sigma,i} \geq 0.$$

Let us symmetrically extend $[r_{ij}]$ by defining $r_{ij} = r_{ji}$ for $j < i$. Then, we can simplify (15,16,17,18), and write them all together as,

$$- \mu_i r_{ij} + \mu_i r_{i+1,j} - p_j 1_{\{\sigma(i)=\sigma(j)\}} + w_{\sigma(i),j} 1_{\{\sigma(i)\neq\sigma(j)\}} \geq 0, \text{ for } j \geq i+1. (z_{ij})$$

By defining $q_{ij} := -r_{ij}$, further simplifying by eliminating the p_i 's and $w_{\sigma,i}$'s, and using the notation $[x]^+ := \max(x, 0)$, we obtain the following duality Theorem.

Theorem 5 *The dual of the Performance LP (12,9,3,4,5,6,7,8, 10,14) for the class of all non-idling policies is the following LP, called the Drift LP:*

$$\text{Min} \quad \sum_{i=1}^L \lambda q_{ii} - \sum_{i=1}^{L-1} \lambda q_{i,i+1} \quad (19)$$

subject to,

$$\begin{aligned} \lambda q_{1j} + \max_{\{i:\sigma(i)=\sigma(j)\}} \mu_i (q_{i+1,j} - q_{ij}) + \sum_{\{\sigma:\sigma\neq\sigma(j)\}} \left[\max_{\{i:\sigma(i)=\sigma\}} \mu_i (q_{i+1} - q_{ij}) \right]^+ \leq -1, \\ \text{for } 1 \leq j \leq L, \end{aligned} \quad (20)$$

$$q_{ij} = q_{ji}, \text{ and } q_{L+1,j} = 0, \text{ for } 1 \leq i, j \leq L. \quad (21)$$

It is important to note that the variables q_{ij} are allowed to be sign indefinite in the Drift LP.

We note that any auxiliary inequalities (13), if present, also carry over in the natural way to the dual Drift LP, as we show in Section 9.1.

The significance of the Drift LP is this. Suppose $Q = [q_{ij}]$ is a feasible solution. Then it has been shown in Kumar and Meyn [11] that the quadratic form $x^T Q x$ has a negative drift, in the sense that for some $\gamma > 0$ and $c < \infty$,

$$E[x^T(\tau_{n+1})Qx(\tau_{n+1}) \mid x(\tau_n) = x] \leq x^T Q x - \gamma|x| + c \text{ for all } x \in Z_+^L. \quad (22)$$

To see this, note that,

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

where e_i is the i -th unit vector. Since all the terms not involving x_i are bounded, one has,

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

for some $c < \infty$. Hence,

$$\begin{aligned} E[x^T(\tau_{n+1})Qx(\tau_{n+1}) \mid \mathcal{F}_{\tau_n}] - x^T \tau_n Q x \tau_n &\leq 2\lambda \sum_{j=1}^L q_{1j} x_j \tau_n \\ &+ 2 \sum_{i=1}^L \mu_i \sum_{j=1}^L (q_{i+1,j} - q_{ij}) w_i \tau_n x_j \tau_n + c, \end{aligned}$$

where for simplicity we have defined and used $q_{L+1,j} := 0$. One can regroup the above double summations machine by machine to give,

$$\begin{aligned} E[x^T(\tau_{n+1})Qx(\tau_{n+1}) \mid \mathcal{F}_{\tau_n}] - x^T \tau_n Q x \tau_n &\leq 2 \sum_{j=1}^L x_j \tau_n [\lambda q_{1j} + \sum_{\{i:\sigma(i)=\sigma(j)\}} \mu_i (q_{i+1,j} - q_{ij}) w_i \tau_n \\ &+ \sum_{\{\sigma:\sigma \neq \sigma(j)\}} \sum_{\{i:\sigma(i)=\sigma\}} \mu_i (q_{i+1,j} - q_{ij}) w_i \tau_n] + c. \end{aligned}$$

Now note that due to the non-idling policy, $\sum_{\{i:\sigma(i)=\sigma(j)\}} w_i \tau_n = 1$, whenever $x_j \tau_n \geq 1$. Hence,

$$\begin{aligned} \mathbb{E}[x^T(\tau_{n+1})Qx(\tau_{n+1}) \mid \mathcal{F}_{\tau_n}] - x^T \tau_n Qx \tau_n &\leq 2 \sum_{j=1}^L x_j \tau_n [\lambda q_{1j} \\ &+ \text{Max}_{\{i:\sigma(i)=\sigma(j)\}} \mu_i (q_{i+1,j} - q_{ij}) + \sum_{\{\sigma:\sigma \neq \sigma(j)\}} [\text{Max}_{\{i:\sigma(i)=\sigma\}} (q_{i+1,j} - q_{ij})]^+ + c. \end{aligned}$$

Thus we obtain the following Theorem.

Theorem 6 *Any feasible solution of the constraints (20,21) of the Drift LP gives a quadratic form $x^T Qx$ with a negative drift as in (22).*

7 Stability and Instability Testing by Negative Drift

The negative drift condition (22) can imply either stability or instability, depending on whether the quadratic form is positive in the positive orthant.

A symmetric matrix Q is said to be *copositive* if $x^T Qx \geq 0$ for all $x \geq 0$. (By $x \geq 0$ we mean every component of x is nonnegative). It has been shown in Kumar and Meyn [11] that if Q is copositive, then the system is stable in a very strong sense.

Theorem 7 *If Q is a copositive matrix for which the negative drift condition (22) is satisfied, then*

(i) $W(x) := (x^T Qx)^{1/2}$ satisfies,

$$\mathbb{E}[W(x(\tau_{n+1})) \mid \mathcal{F}_{\tau_n}] \leq \begin{cases} W(x(\tau_n)) - \epsilon, & |x(\tau_n)| \geq K \\ c, & |x(\tau_n)| < K \end{cases}, \quad (23)$$

for some $\epsilon > 0$, and $c, K < \infty$. Also, $W(x(\tau_{n+1})) - W(x(\tau_n))$ is bounded.

(ii) As a consequence of (i), the Markov chain $\{x(\tau_n)\}$ has a finite exponential moment that is geometrically ergodic, i.e., V -uniformly ergodic with $V(x) = \exp(\epsilon|x|)$ for some $\epsilon > 0$, in the notation of Meyn and Tweedie [25]. That is, there exists an $r > 1$, $C < +\infty$, such that for any function f satisfying $|f(y)| \leq V(y)$ for all y , and any initial condition $x(\tau_0) = x$,

$$\sum_{n=0}^{\infty} r^n |E[f(x(\tau_n))] - \sum_y f(y)\pi(y)| < CV(x).$$

A common approach to establish stability is to directly search for a non-negative $W(x)$ satisfying (23), which is known as Foster's criterion. For it then follows that for all x , the first entrance time $\bar{\sigma}$ to the origin (see Section 3) satisfies $\epsilon \mathbb{E}[\bar{\sigma} \mid x(\tau_0) = x] \leq W(x) + d$ for some $d < \infty$, and hence that the Markov chain is positive recurrent. Conversely, the function $W(x) := \epsilon \mathbb{E}[\bar{\sigma} \mid x(\tau_0) = x]$ always satisfies (23) if it is finite valued, see [25]. Hence the task of finding such a Lyapunov function W satisfying (23) is related to approximating the mean emptying time. In our system, $|x(\tau_n)|$ decreases by at most one at each time step, and so $W(x) \geq \epsilon|x| - d$. In most networks one expects the mean emptying time to be roughly proportional to the initial total queue size, and indeed, in all of the examples that we have encountered, W closely approximates a norm. We now digress a bit to show a partial converse to (i) of Theorem 7, which indicates that functions that are almost quadratic are the natural candidates for satisfying a drift condition as in (22).

Theorem 8 *Suppose that W is any non-negative function satisfying (23) which is Lipschitz, i.e., for some $M < \infty$, $|W(x) - W(y)| \leq M|x - y|$ for all states x, y . Then $W^2(x)$ has the negative drift of (22), i.e.,*

$$\mathbb{E}[W^2(x(\tau_{n+1})) \mid x(\tau_n) = x] \leq W^2(x) - \gamma|x| + c \text{ for all } x \in Z_+^L. \quad (24)$$

and it is equivalent to a quadratic in the sense that for some $e_1 < \infty$, and $e_2 > 0$,

$$-e_1 + e_2|x|^2 \leq W^2(x) \leq e_1 + e_2^{-1}|x|^2 \text{ for all } x \in Z_+^L. \quad (25)$$

Proof. Note first that $\epsilon|x| - d \leq W(x) \leq |W(x) - W(0)| + W(0) \leq M|x| + W(0)$, and so $W^2(x)$ is equivalent to a quadratic. Also, for $|x| \geq K$, by expanding the square within the expectation,

$$\begin{aligned} \mathbb{E}[W^2(x(\tau_{n+1})) \mid x(\tau_n) = x] &= \mathbb{E}[(W(x(\tau_n)) + W(x(\tau_{n+1})) - W(x(\tau_n)))^2 \mid x(\tau_n) = x] \\ &= W^2(x) + 2\mathbb{E}[W(x(\tau_{n+1})) - W(x(\tau_n)) \mid x(\tau_n) = x]W(x) \\ &\quad + \mathbb{E}[(W(x(\tau_{n+1})) - W(x(\tau_n)))^2 \mid x(\tau_n) = x] \\ &\leq W^2(x) - 2\epsilon W(x) + 4M^2 \quad (\text{using } |x(\tau_{n+1}) - x(\tau_n)| \leq 2) \\ &\leq W^2(x) - 2\epsilon^2|x| + 2\epsilon d + 4M^2 \quad (\text{using } W(x) \geq \epsilon|x| - d). \square \end{aligned}$$

Now we turn to systems satisfying (22) for which Q is *not* copositive. Let us begin by showing a preliminary result on systems that have a *positive* drift. Consider a countable state Markov chain $\{x_n\}$, with state space X , and containing just a single communicating class \mathcal{C} . Suppose that there is a non-negative function V satisfying,

$$E[V(x_{n+1}) \mid \mathcal{F}_n] \geq V(x_n) \text{ for all } x_n \in X \setminus C_m, \text{ where } C_m := \{x \in X : V(x) \leq m\}, \quad (26)$$

and \mathcal{F}_n is the past σ -algebra of $\{x_n\}$. Let us suppose that \mathcal{C} intersects both C_m and its complement $X \setminus C_m$. Suppose now that f is a nonnegative function satisfying

$$|V(x_{n+1}) - V(x_n)| \leq f(x_n) \text{ for all } n. \quad (27)$$

Then, we will show that either $\{x_n\}$ is not positive recurrent, or it is positive recurrent but with an infinite mean for $f(x_n)$ in steady-state.

Suppose to the contrary that $\{x_n\}$ is in fact positive recurrent with a finite mean value for $f(x_n)$ in steady-state. Then, as we have seen in Section 3, $E[\sum_{n=0}^{\sigma} f(x_n) \mid x_0 = x] < +\infty$ for all $x \in \mathcal{C}$, where $\sigma := \min\{n \geq 0 \mid x_n \in C_m\}$ is the first hitting time to C_m . Define, $M_n := V(x_{n \wedge \sigma})$, and note $\{M_n, \mathcal{F}_n\}$ is a positive submartingale. Moreover, for $x_0 \in \mathcal{C}$, $\{M_n\}$ is uniformly integrable, since $M_n \leq \sum_{k=0}^{\sigma} f(x_k) + M_0$ for all n . Thus, by the optional sampling theorem (see Doob [26]), $E[M_{\sigma}] \geq M_0$. However, if $x_0 \in \mathcal{C} \setminus C_m$, then $M_0 > m$, while $M_{\sigma} = V(x_{\sigma}) \leq m$, yielding a contradiction. \square

Thus we have the following Theorem.

Theorem 9 *Let $\{x_n\}$ be a countable state Markov chain with state space X , with a single communicating class \mathcal{C} . Let V be a non-negative function satisfying (26). Suppose f is a nonnegative function satisfying (27).¹ Then, the following are true.*

- (i) $E_x[\sum_{n=0}^{\sigma} f(x_n)] = \infty$, where $\sigma := \min\{n \geq 0 \mid x_n \in C_m\}$ is the first hitting time to C_m .
- (ii) *If, in addition, \mathcal{C} intersects both C_m and its complement, then either $\{x_n\}$ is not positive recurrent, or it is positive recurrent but with an infinite mean for $f(x_n)$ in steady-state.*

¹By using the notion of f -regularity from [25], one can relax this to $E[|V(x_{n+1}) - V(x_n)| \mid \mathcal{F}_n] \leq f(x_n)$ for all n . The proof then uses Fubini's Theorem.

Now let us return to our re-entrant line for which we suppose that there is a quadratic form $x^T Q x$ that is *not* copositive, but having the negative drift (22). Since Q is not copositive, there exists an $x \geq 0$ such that $x^T Q x < 0$. By scaling x we can drive the value of the quadratic form arbitrarily negative. Hence $\inf_{x \geq 0} x^T Q x = -\infty$. Define $V(x) := \max\{-x^T Q x, 0\}$, and note that $\sup_{x \geq 0} V(x) = +\infty$. Now,

$$\begin{aligned}
E[V(x(\tau_{n+1})) \mid \mathcal{F}_{\tau_n}] &= E[\max\{-x^T(\tau_{n+1})Qx(\tau_{n+1}), 0\} \mid \mathcal{F}_{\tau_n}] \\
&\geq \max\{-E[x^T(\tau_{n+1})Qx(\tau_{n+1}) \mid \mathcal{F}_{\tau_n}], 0\} \\
&\geq \max\{-x^T(\tau_n)Qx(\tau_n) + \gamma \mid x(\tau_n) \mid -c, 0\} \quad (\text{from (22)}) \\
&\geq \max\{-x^T(\tau_n)Qx(\tau_n), 0\}, \quad \text{if } |x(\tau_n)| \geq \frac{c}{\gamma} \\
&= V(x(\tau_n)), \quad \text{if } |x(\tau_n)| \geq \frac{c}{\gamma}. \tag{28}
\end{aligned}$$

Also, for $x \geq 0$, $V(x) = \max\{-x^T Q x, 0\} \leq (\max_{ij}(-q_{ij}, 0))|x|^2 = q|x|^2$ for some $q > 0$ (since $\max_{ij}(-q_{ij}) := q > 0$ due to the non-copositivity of Q). Hence, (28) holds whenever $V(x) \geq m$, or equivalently, when $x \in X \setminus C_m$, for m large enough. Finally, note that $V(x(\tau_{n+1})) - V(x(\tau_n)) = O(|x(\tau_n)|)$.

Thus, if $\inf_{x \in C} x^T Q x = -\infty$, then the hypotheses of Theorem 9 are satisfied, and the system does not have a finite first moment.

The other possibility is $\inf_C x^T Q x > -\infty$. Then, even if the system is stable, it is at best highly non-robust. To see this, perturb the scheduling policy ever so slightly, so that all non-empty buffers get a minute fraction of the machine's attention, while the bulk of it is used according to the original scheduling policy; i.e., the new allocation is $\bar{w}_i(t) := (w_i(t) + \epsilon)[\sum_{\{j:\sigma(j)=\sigma(i) \text{ and } x_j(t) \geq 1\}} (w_j(t) + \epsilon)]^{-1} 1(x_i(t) \geq 1)$, where $w_i(t)$ is the allocation that the original scheduling policy would use. Then the new policy is irreducible. Moreover, the drift being only slightly perturbed, remains negative. Thus the perturbed policy is unstable.

Theorem 10 *Suppose there exists a non-copositive matrix Q satisfying the negative drift condition (22). Then the stationary non-idling scheduling policy gives rise to a Markov chain*

which is either unstable, where by “unstable” we mean a system without a finite first moment, or it is highly non-robust in that arbitrarily small perturbations of the scheduling policy can lead to an unstable system. In particular, if $\inf_{x \in \mathcal{C}} x^T Q x = -\infty$, then the system is unstable.

Our approach to stability/instability determination will be to use linear programming as in Kumar and Meyn [11], to construct a Q which gives a negative drift. However, unlike [11], we use the linear program of Theorem 5, which imposes no sign restrictions on the signs of the q_{ij} ’s. It will thus remain to conduct a separate copositivity test on the constructed Q . This can be done by using the following characterization due to Keller; see Theorem 4.2 of Cottle, Habetler and Lemke [27].

Theorem 11 *A symmetric matrix Q is copositive if and only if each principal submatrix, for which the cofactors of the last row are nonnegative, has a nonnegative determinant.*

As a consequence, there exists a finite procedure for determining whether a matrix is copositive. A recent algorithm, and references to recent literature on the topic, can be found in Andersson, Chang and Elfving [28]. Unfortunately, for large matrices, the copositivity test may be computationally intractable, since it has been shown that the determination of copositivity is NP-Complete; see Murty and Kabadi [29].

8 The Stability of All Non-Idling Stationary Policies

Now we can considerably extend Theorem 4. Suppose that the Performance LP is bounded. Then the system is stable, i.e., it has a finite first moment, as seen from Theorem 4. Moreover, from duality (Theorem 5), the Drift LP is feasible, and so we know that there exists a Q which satisfies the negative drift condition (22). Either this Q is copositive or not copositive. If it is not copositive, then since the original policy is stable, a slight perturbation of it is unstable. However, the perturbed policy is still non-idling, and is therefore known to be stable. Thus, the matrix Q is indeed copositive (Theorem 10). Theorem 7 then shows that there is a geometrically converging exponential moment. Thus, we have arrived at the following Theorem.

Theorem 12 *Suppose the Performance LP (12,9,3,4,5,6,7,8,10,14) for the class of all non-idling policies is bounded. Then every feasible solution of the dual Drift LP (20,21) is copositive. Thus, the Markov chain has a geometrically converging exponential moment. i.e., it is V -uniformly ergodic with $V(x) = \exp(\epsilon|x|)$ for some $\epsilon > 0$, in the sense of Theorem 7.*

If however the Performance LP is unbounded, then the Drift LP is infeasible, and hence there exists no quadratic form with a negative drift. One will then need to examine other Lyapunov functions with different growth rates, e.g., *piecewise linear* Lyapunov functions, as in Botvich and Zamyatin [23].

9 The Performance and Stability of Systems With Auxiliary Constraints: Buffer Priority Policies

We now address systems with the auxiliary constraints (11). Then the set of scheduling policies captured by the inequalities need not possess the two special properties used in Section 5, i.e., it need not contain a policy known to be stable, and need not be closed under the “composition” of policies. Hence we are unable to deduce stability merely from the boundedness of a Performance LP incorporating the additional inequalities (13). We thus need to perform a copositivity check on the feasible solution $Q = [q_{ij}]$ for the dual Drift LP obtained from the Performance LP, and then apply either Theorem 7 or Theorem 10.

To illustrate how one deals with such systems with auxiliary constraints, we consider the special class of *buffer priority* policies.

9.1 Buffer Priority Policies

Let $\eta : \{1, \dots, S\} \rightarrow Z$ be an ordering of the buffers. where η is one-to-one, and suppose that pre-emptive priority is provided by each machine to buffers earlier in the ordering. The policy is non-idling. The choice $\eta(b_i) = -i$ gives the popular Last Buffer First Serve Policy, also known as the SERPT policy, see Lu and Kumar [5].

Note therefore that, for any b_j with priority over b_i , i.e., with $\sigma(i) = \sigma(j)$ but $\eta(b_j) < \eta(b_i)$, $w_i(\tau_n) = 0$ whenever $x_j(\tau_n) \geq 1$. Hence $E[w_i(\tau_n)x_j(\tau_n)] = 0$. Thus we obtain the following

particular auxiliary equalities (13) for the policy η :

$$z_{ij} = 0 \text{ for all } i, j \text{ with } \sigma(i) = \sigma(j) \text{ and } \eta(b_j) < \eta(b_i). \quad (29)$$

The dual Drift LP is (19), subject to (21), and

$$\begin{aligned} \lambda q_{1j} + \max_{\substack{\{i:\sigma(i)=\sigma(j)\} \\ \text{and } \eta(i) \leq \eta(j)}}} \mu_i(q_{i+1,j} - q_{ij}) + \sum_{\{\sigma:\sigma \neq \sigma(j)\}} \left[\max_{\{i:\sigma(i)=\sigma\}} \mu_i(q_{i+1,j} - q_{ij}) \right]^+ \\ \leq -1 \quad \text{for } 1 \leq j \leq L. \end{aligned} \quad (30)$$

As shown in Kumar and Meyn [11], any feasible $Q = [q_{ij}]$ satisfies the negative drift condition (22). Thus we obtain the following Theorem.

Theorem 13 *Consider a buffer priority policy η .*

- (i) *The dual of the Performance LP (12,9,3,4,5,6,7,8,10,29,14) is the Drift LP (19,30,21).*
- (ii) *If the η has a finite first moment, then the mean number of parts in the system is bounded above by the value of the Performance LP, and below by the same LP with a “Min” replacing the “Max.”*
- (iii) *If the Performance LP is infeasible, then η does not have a finite first moment.*
- (iv) *If the Performance LP is bounded, the dual Drift LP has a feasible solution $Q = [q_{ij}]$, which has the negative drift (22).*
- (v) *If a feasible solution Q to the Drift LP is copositive, then the system has an exponential moment that converges geometrically, i.e., the system is V -uniformly ergodic, with $V(x) = \exp(\epsilon|x|)$ for some $\epsilon > 0$.*
- (vi) *If a feasible solution Q to the Drift LP is not copositive, then either the system does not have a finite first moment, or else an arbitrarily small perturbation of it does not have a finite first moment. In particular, if $\inf_{x \in \mathcal{C}} x^T Q x = -\infty$, then the system is unstable in that it does not have a finite first moment.*

We have been unable to establish that boundedness of a Performance LP including auxiliary inequalities, such as those arising from buffer priority constraints, automatically implies the stability of the system. (For the class of all non-idling policies, not featuring any auxiliary constraints, such a result was shown in Theorem 4). This is an important open problem.

10 Stability of Systems With Auxiliary Constraints: The Monotone Linear Program

For systems with auxiliary constraints, one needs to perform a copositivity check, as in Theorem 13. As noted earlier, this test is NP-Complete. To avoid this, we now construct an LP, called the Monotone LP, which is obtained from the Performance LP by relaxing some of the equality constraints to inequalities, and slightly decreasing their RHS's. We will show that if this new Monotone LP is bounded, then the system is stable. The reason the LP is called "Monotone" is because its value is monotone increasing in the arrival rate λ .

Suppose the system satisfies the auxiliary constraints (11), irrespective of the arrival process.

Define $w_0(t) := 1$ if $|x(t)| \leq N$; and $= 0$ if $|x(t)| > N$. Let us control the flow of parts into the system by discarding all arrivals to the system which arrive when $w_0 = 0$. Thus, we ensure that the number of parts in the system is always less than N . Clearly this new "flow-controlled" system has a finite first moment.

The performance of this modified system satisfies the constraints (9,3,4,5,6,7,8,10,13,14), of the Performance LP, except that any term $E[\lambda x_j]$ is replaced by $E[\lambda w_0 x_j]$, and any constant term λ is replaced by $E[\lambda w_0]$. Thus, the constraints (3,4,5,6,7) are replaced by:

$$\lambda E[w_0 x_1] - \mu_1 z_{11} = -E[\lambda w_0],$$

$$\mu_{j-1} z_{j-1,j} - \mu_j z_{jj} = -E[\lambda w_0], \text{ for } 2 \leq j \leq L,$$

$$\lambda E[w_0 x_2] - \mu_1 z_{12} + \mu_1 z_{11} - \mu_2 z_{21} = E[\lambda w_0],$$

$$\lambda E[w_0 x_j] - \mu_1 z_{1j} - \mu_j z_{j1} + \mu_{j-1} z_{j-1,1} = 0, \text{ for } 3 \leq j \leq L, \text{ and}$$

$$\mu_{i-1} z_{i-1,i+1} - \mu_1 z_{i,i+1} + \mu_i z_{ii} - \mu_{i+1} z_{i+1,i} = E[\lambda w_0], \text{ for } 2 \leq i \leq L - 1.$$

However, $E[w_0 x_i] \leq E[x_i] = \bar{x}_i$, and $0 \leq E[w_0] \leq 1$, and so,

$$\lambda \bar{x}_1 - \mu_1 z_{11} \geq -\lambda, \tag{31}$$

$$\mu_{j-1}z_{j-1,j} - \mu_j z_{j,j} \geq -\lambda, \quad \text{for } 2 \leq j \leq L, \quad (32)$$

$$\lambda \bar{x}_2 - \mu_1 z_{12} + \mu_1 z_{11} - \mu_2 z_{21} \geq 0, \quad (33)$$

$$\lambda \bar{x}_j - \mu_1 z_{1j} - \mu_j z_{j1} + \mu_{j-1} z_{j-1,1} \geq 0 \quad \text{for } 3 \leq j \leq L, \text{ and} \quad (34)$$

$$\mu_{i-1} z_{i-1,i+1} - \mu_i z_{i,i+1} + \mu_i z_{ii} - \mu_{i+1} z_{i+1,i} \geq 0 \quad \text{for } 2 \leq i \leq L-1. \quad (35)$$

Let us call the linear program (12,9,31,32,33,34,35,8,10,13,14) the *Monotone LP*.

Letting the number of parts $N \rightarrow +\infty$, and invoking the weak convergence argument of Theorem 4, we see that if the Monotone LP is bounded, then the system has a finite first moment. Moreover, any feasible solution of the Monotone LP for a particular value of the arrival rate λ , is also a feasible solution for all higher values of λ . This is so because λ always multiplies a nonnegative quantity in an inequality going the right way. Thus the value of the LP is monotone increasing in λ . Hence its boundedness for a particular λ allows us to deduce the finiteness of the first moment for all lower arrivals rates too.

Theorem 14 (i) *Consider a system satisfying the auxiliary constraints (11), irrespective of the arrival process. Suppose the Monotone LP (12,9,31,32,33,34,35,8,10,13,14) is bounded for a particular value of λ . Then the system has a finite first moment for all arrival rates $\lambda' \leq \lambda$.*

(ii) *Moreover, under the assumptions of (i), for every arrival rate $\lambda' \leq \lambda$, the mean number of parts in the system is bounded above by the Performance LP (12,9,3,4,5,6,7,8,10,13,14), with λ replaced by λ' . It is bounded below by the same LP, except that “Max” is replaced by “Min.”*

11 Transient Performance Bounds and the Finite Time LP

We will now show how one may bound transient performance. We introduce a new LP called the *Finite Time LP*. It is a another modification of the Performance LP, where the “=”s in (3,4,5,6,7,8) are replaced by “ \geq ”s.

Consider the system started with an arbitrary initial condition $x(\tau_0)$. Then, for $j \neq 1$ and $i \geq j + 2$,

$$\begin{aligned} E[x_i(\tau_{n+1})x_j(\tau_{n+1}) \mid \mathcal{F}_{\tau_n}] &= x_i(\tau_n)x_j(\tau_n) + \mu_{i-1}w_{i-1}(\tau_n)x_j(\tau_n) \\ &\quad - \mu_i w_i(\tau_n)x_j(\tau_n) + \mu_{j-1}w_{j-1}(\tau_n)x_i(\tau_n) - \mu_j w_j(\tau_n)x_i(\tau_n). \end{aligned}$$

Hence, by telescoping,

$$\begin{aligned} 0 \leq \frac{1}{N}E[x_i(\tau_N)x_j(\tau_N)] &= \frac{1}{N}[x_i(\tau_0)x_j(\tau_0)] + \frac{1}{N} \sum_{n=0}^{N-1} E[\mu_{i-1}w_{i-1}(\tau_n)x_j(\tau_n) - \mu_i w_i(\tau_n)x_j(\tau_n) \\ &\quad + \mu_{j-1}w_{j-1}(\tau_n)x_i(\tau_n) - \mu_j w_j(\tau_n)x_i(\tau_n)]. \end{aligned}$$

Defining $z_{ij}^{(N)} := \frac{1}{N} \sum_{n=0}^{N-1} E[w_i(\tau_n)x_j(\tau_n)]$, we have,

$$\mu_{i-1}z_{i-1,j}^{(N)} - \mu_i z_{ij}^{(N)} + \mu_{j-1}z_{j-1,i}^{(N)} - \mu_j z_{ij}^{(N)} \geq -\frac{1}{N}x_i(\tau_0)x_j(\tau_0).$$

This however is just the constraint (8), except that “=” is replaced by “≥,” and the RHS is modified by taking into account the initial condition. Similarly, we can study the other values of i, j .

We thus deduce that the following constraints are satisfied by the $z_{ij}^{(N)}$'s and $\bar{x}_j^{(N)} := \sum_{i:\sigma(i)=\sigma(j)} z_{ij}^{(N)}$.

$$2\lambda\bar{x}_1 + 2\lambda - 2\mu_1 z_{11} \geq -\frac{1}{N}E[x_1^2(\tau_0)], \quad (36)$$

$$2\mu_{j-1}z_{j-1,j} + 2\lambda - 2\mu_j z_{j,j} \geq -\frac{1}{N}E[x_j^2(\tau_0)], \quad \text{for } 2 \leq j \leq L, \quad (37)$$

$$\lambda\bar{x}_2 - \lambda - \mu_1 z_{12} + \mu_1 z_{11} - \mu_2 z_{21} \geq -\frac{1}{N}E[x_1(\tau_0)x_2(\tau_0)], \quad (38)$$

$$\lambda\bar{x}_j - \mu_1 z_{1j} - \mu_j z_{j1} + \mu_{j-1}z_{j-1,1} \geq -\frac{1}{N}E[x_1(\tau_0)x_j(\tau_0)], \quad \text{for } 3 \leq j \leq L, \quad (39)$$

$$\mu_{i-1}z_{i-1,i+1} - \mu_i z_{i,i+1} - \lambda + \mu_i z_{ii} - \mu_{i+1}z_{i+1,i} \geq -\frac{1}{N}E[x_i^2(\tau_0)], \quad \text{for } 2 \leq i \leq L-1, \text{ and } (40)$$

$$\mu_{i-1}z_{i-1,j} - \mu_i z_{i,j} + \mu_{j-1}z_{j-1,i} - \mu_j z_{ji} \geq -\frac{1}{N}E[x_i(\tau_0)x_j(\tau_0)], \quad \text{for } 2 \leq i \leq L-2 \text{ and } i+2 \leq j \leq L. \quad (41)$$

Thus we have the following bounds on transient performance.

Theorem 15 *Consider a system with the auxiliary constraints (11). Let the initial system state be $x(\tau_0)$. Then the transient performance $\frac{1}{N}E[\sum_{n=0}^{N-1} c^T x(\tau_n)]$ is bounded above by the Finite Time LP with objective function $\sum_{j=1}^L c_j \bar{x}_j$, and subject to the constraints (9,36,37,38,39,41,10,13,14). It is bounded below by the same linear program with “Max” replaced by “Min.”*

It should be noted that the constraints of the dual of this Finite Time LP (with $c^T x := |x|$) are precisely those of the original stability LP proposed in Kumar and Meyn [11], which required all $q_{ij} \geq 0$.

12 Fluid Models

Recently, Rybko and Stolyar [12] have established the stability properties of certain queueing networks, by studying the properties of the associated fluid approximation, and Malyshev [30] has advocated the use of fluid models to analyze complex queueing networks. The results of [12] have been considerably generalized by Dai [13], and recently by Chen [14]. They show that one may establish the stability of a stochastic system by establishing the stability of its fluid approximation.

The stability tests given here carry over to such fluid models. Through them, one can thus establish the stability of networks without exponential assumptions on the service or interarrival distributions.

For notational convenience in this section, let us denote $x(\tau_n)$ by $x(n)$. Consider the collection of Markov chains $\{x(n)\}$ on the same measure space, but with different initial conditions x . We shall denote the measure corresponding to an initial condition x by P_x , and the corresponding expectation operator by E_x . For each initial condition $x \neq 0$, we construct a continuous time process $\phi^x(t)$ as follows. If $|x|t$ is an integer, we set

$$\phi^x(t) = \frac{1}{|x|} x(|x|t).$$

For all other $t \geq 0$, we define $\phi^x(t)$ by interpolation, so that it is continuous and piecewise linear in t . Note that $|\phi^x(0)| = 1$. Also, since $|x(n+1) - x(n)| \leq 2$, and $|x(n+1)| - |x(n)| \leq 1$,

it follows that $|\phi^x(t) - \phi^x(s)| \leq 2|t - s|$ and $|\phi^x(t)| \leq 1 + t$, for all x and $s, t \geq 0$. Taking the topology of uniform convergence on compact sets on $C[0, \infty)$, we see that the set of probability measures induced on $C[0, \infty)$ by $\{P_x : x \in Z_+^L, x \neq 0\}$, is tight; see Billingsley [31].

Denote the collection of all “fluid limits”

$$\Phi := \{ \phi : \phi^x \xrightarrow{w} \phi \text{ as } x \rightarrow \infty \text{ along some subsequence} \} = \bigcap_{n=1}^{\infty} \overline{\{ \phi^x : |x| > n \}}$$

where “ \xrightarrow{w} ” denotes weak convergence, and the overbar denotes weak closure. Note that $|\phi(t)| \leq 1 + t$, and $|\phi(t) - \phi(s)| \leq |t - s|$, for every $\phi \in \Phi$. It is shown in Dai [13] that any $\phi \in \Phi$ satisfies a certain integral equation. We will not require this property here though.

We will say that Φ is L_p -stable if

$$\lim_{t \rightarrow \infty} \sup_{\phi \in \Phi} \mathbb{E}[|\phi(t)|^p] = 0.$$

Now we show the equivalence of the L_2 -stability of the fluid model, and several notions of stability for the original system. The criterion (i) below is met when we have a copositive solution for the Drift LP; condition (ii) is a form of f -regularity (see [25]) that is implicit in Theorem 9 above; the bound (iii) is precisely that obtained from the Finite Time LP; while criterion (iv) is simply the L_2 -stability of the fluid limit processes.

Theorem 16 *The following stability criteria are equivalent.*

(i) *For some function V equivalent to a quadratic in the sense of (25), and some $c < \infty$,*

$$\mathbb{E}[V(x_{n+1}) \mid x_n = x] \leq V(x) - |x| + c, \text{ for all } x.$$

(ii) *For some quadratic function V and some $c < \infty$,*

$$\mathbb{E} \left[\sum_{n=1}^{\bar{\sigma}} |x(n)| \mid x(0) = x \right] \leq V(x) + c, \text{ for all } x,$$

where $\bar{\sigma}$ is the first entrance time to 0.

(iii) For some quadratic function V and some $c < \infty$,

$$\sum_{n=1}^N \mathbb{E}_x[|x(n)|] \leq V(x) + cN, \quad \text{for all } x \text{ and } N \geq 1.$$

(iv) The fluid model Φ is L_2 -stable.

Proof That (i) implies (iii) follows by telescoping using $V(x) \geq 0$. That (i) implies (ii) follows from Theorem 14.7.3 of [25]. To see the converse, let $V_0(x) := \mathbb{E}_x[\sum_{n=0}^{\sigma} |x(n)|]$, where $\sigma := \min\{n \geq 0 : x(n) = 0\}$ is the first *hitting* time to 0. Note that V_0 is equivalent to a quadratic, since it is bounded below by $1/2|x|^2$, and, since (ii) holds, also above by a quadratic plus a constant. Moreover, $\mathbb{E}[V_0(x(n+1)) \mid x(n) = x] = V_0(x) - |x| + b1(x=0)$, where $b := \mathbb{E}_0[\sum_{n=1}^{\bar{\sigma}} |x(n)|]$ is finite by (ii) with $x = 0$. This shows that (ii) \Rightarrow (i). To complete the proof we will show that (iii) \Rightarrow (iv) \Rightarrow (ii). This will be performed in several steps.

First we show that if (iii) holds then there exists a constant $d < \infty$ such that for all $s \geq 0$, and all fluid limits $\phi \in \Phi$,

$$\frac{1}{2}|\phi(s)|^2 \leq \mathbb{E}\left[\int_s^\infty |\phi(\tau)| d\tau \mid \phi(r) : r \leq s\right] \leq d|\phi(s)|^2. \quad (42)$$

The lower bound follows trivially from the Lipschitz property. To see the upper bound, let $g : \mathbb{R}_+^{mL} \rightarrow \mathbb{R}$ be any continuous function with bounded support, where $m \geq 1$ is any fixed positive integer. Let $0 \leq s_1 \leq s_2 \leq \dots \leq s_m < s$ be fixed times, and define $G^x = g(\phi^x(s_1), \dots, \phi^x(s_m))$. Let $\phi^x \xrightarrow{w} \phi$ as $x \rightarrow \infty$ along some sequence, and set $G = g(\phi(s_1), \dots, \phi(s_m))$. Since $\phi^x(t) \leq 1 + t$, the family of random variables $\{|\phi^x(t)|^p : x \neq 0, 0 \leq t \leq T\}$ is uniformly integrable for every $p > 0$ and $T < \infty$. Hence, from (iii), it follows that as $x \rightarrow \infty$ along this sequence,

$$\mathbb{E}\left[\int_s^t |\phi^x(\tau)| d\tau G^x\right] \rightarrow \mathbb{E}\left[\int_s^t |\phi(\tau)| d\tau G\right]. \quad (43)$$

The LHS of this expression can be computed as follows:

$$\int_s^t |\phi^x(\tau)| d\tau = \int_s^t \left| \frac{x(\lfloor |x|\tau \rfloor)}{|x|} \right| d\tau + o(1) = \frac{1}{|x|^2} \sum_{n=\lfloor |x|s+1 \rfloor}^{\lfloor |x|t \rfloor} |x(n)| + o(1),$$

where $[r]$ denotes the integer part of r , and the term $o(1) = O\left(\frac{1}{|x|}\right) \rightarrow 0$, as $x \rightarrow \infty$. Taking conditional expectations given $\mathcal{F}_{[|x|s]} = \{x(n) : n \leq [|x|s]\}$, we obtain from (iii),

$$\begin{aligned} \mathbb{E} \left[\int_s^t |\phi^x(\tau)| d\tau \mid G^x \mid \mathcal{F}_{[|x|s]} \right] &= \mathbb{E} \left[\frac{1}{|x|^2} \sum_{k=[|x|s+1]}^{[|x|t]} |x(k)| \mid \mathcal{F}_{[|x|s]} \right] G^x + o(1) \\ &\leq \frac{1}{|x|^2} [V(x([|x|s])) + c([|x|t] - [|x|s])] G^x + o(1). \end{aligned}$$

Choosing $d < \infty$ so that $V(x) \leq d|x|^2$ and taking expectations we have by (43), $\mathbb{E}[\int_s^t |\phi(\tau)| d\tau \mid G] \leq d\mathbb{E}[|\phi(s)|^2 G]$. This holds for g continuous with compact support, but then may be generalized to arbitrary bounded measurable g by dominated convergence. This completes the proof of (42).

Now we show that (iii) \Rightarrow (iv). Define the positive supermartingale $\{M(s), \sigma(\phi(r) : 0 \leq r \leq s)\}$ by $M(s) = \mathbb{E}[\int_s^\infty |\phi(\tau)| d\tau \mid \phi(r) : 0 \leq r \leq s]$. From (42) we see that $1/2|\phi(s)|^2 \leq M(s) \leq d|\phi(s)|^2$. Hence $\mathbb{E}[M(s)] \leq \mathbb{E}[M(0)] \leq d$. By the dominated convergence theorem applied to the random variables $\int_s^\infty |\phi(\tau)| d\tau$, $s \geq 0$, we have $\mathbb{E}[M(s)] \rightarrow 0$ as $s \rightarrow \infty$, and so

$$\mathbb{E}[|\phi(s)|^2] \rightarrow 0 \text{ as } s \rightarrow \infty. \quad (44)$$

To complete the proof we will show that the convergence in (44) is uniform over $\phi \in \Phi$. Suppose to the contrary that it is not. Then there exists a set of $\{\phi^n\} \subseteq \Phi$ and a sequence of times $t_n \rightarrow \infty$ such that $\mathbb{E}[|\phi^n(t_n)|^2] \geq \epsilon > 0$. Let $\{x_i^n : i, n \geq 0\}$ be the set of states such that $\phi^{x_i^n} \xrightarrow{w} \phi^n$ as $i \rightarrow \infty$, for each n . Denoting by M^n the supermartingale corresponding to ϕ^n , we thus have $\mathbb{E}[M^n(t_n)] \geq \frac{1}{2}\epsilon$, and since $\mathbb{E}[M^n(t)]$ is decreasing in t ,

$$\mathbb{E}[|\phi^n(s)|^2] \geq \frac{1}{d}\mathbb{E}[M^n(s)] \geq \frac{1}{2d}\epsilon, \quad 0 \leq s \leq t_n.$$

It follows that we may find a new sequence of states $\{y^n\}$ among the $\{x_i^n\}$ such that $\mathbb{E}[|\phi^{y^n}(t)|^2] \geq \frac{1}{4d}\epsilon$ for $0 \leq t \leq t_n$. Without loss of generality we may assume that $\phi^{y^n} \xrightarrow{w} \phi$ as $n \rightarrow \infty$, for some $\phi \in \Phi$. By weak convergence and uniform integrability, $\mathbb{E}[|\phi(t)|^2] \geq \frac{1}{4d}\epsilon$, for $t \geq 0$, which contradicts (44), and completes the proof that (iii) \Rightarrow (iv).

Now we complete the proof of the Theorem by showing that (iv) \Rightarrow (i). Under (iv), there exists $N < \infty$, $T < \infty$ such that $\mathbb{E}_x[|x([T|x|])|^2] \leq \frac{1}{2}|x|^2$ for $|x| \geq N$. Denoting by $p_{x,y}^n$ the

n -step transition probability, we have

$$\sum_y p_{x,y}^{n(x)} V(y) \leq V(x) - \frac{1}{2}|x|^2 \quad \text{for all } |x| \geq N, \quad (45)$$

where $V(x) = |x|^2$, and $n(x) = \lfloor T|x| \rfloor$.

Define the *sampled chain* \hat{x}_k , as in p. 468 of [25], as follows. The integer $n(x)$ is a (trivial) stopping time, and we define the iterates $\{s(k)\}$ by $s(0) = 0$, $s(1) = n(x)$, and $s(k+1) = s(k) + n(x(s(k)))$ for $k \geq 1$. Then $\hat{x}_k \triangleq x(s(k))$ is a Markov chain with one-step transition probability $p^{n(x)}$. It then follows from (45) and Theorem 14.2.2 of [25] that for some $d < \infty$, $\mathbb{E}_x[\sum_{k=0}^{\hat{\tau}} \frac{1}{2}|\hat{x}_k|^2] \leq V(x) + d$, where $\hat{\tau}$ is the first entrance time to some bounded set C for the sampled chain. Since $\mathbb{E}[\sum_{n=s(k)}^{s(k+1)} |x(n)| \mid x(m) : 0 \leq m \leq s(k)] \leq c'|\hat{x}_k|^2$, it then follows from the strong Markov property that $\mathbb{E}_x[\sum_{k=0}^{\hat{\tau}} \sum_{n=s(k)}^{s(k+1)} |x(n)|] \leq c'(V(x) + 1)$. Since $\hat{\tau}$ is the hitting time to C for the sampled chain, the hitting time τ_C to C for the unsampled chain satisfies $\tau_C \leq s(\hat{\tau})$, and thus $\mathbb{E}_x[\sum_{k=0}^{\tau_C} |x(k)|] \leq c'(V(x) + 1)$. This and Theorem 14.7.3 of [25] completes the proof. \square

Thus we obtain the following consequences of studying the fluid model; they extend Theorems 7 and 10. (It may be worth noting that the only properties of the Markov chain used in the above proof are that $x(n+1) - x(n)$ is bounded, and that 0 is reachable from every state).

Theorem 17 *Consider a system satisfying the auxiliary constraints (11). Suppose that there exists a non-copositive matrix Q such that the drift condition (22) is satisfied. Then Theorem 16(i)-(iv) are all violated, and hence the following are true.*

- (i) *There does not exist a Lipschitz function $W(x)$ satisfying the drift (23).*
- (ii) *The system is highly non-robust in that the mean of the emptying time $\bar{\sigma}$ is highly discontinuous in the initial condition,*

$$\sup_{\{(x,y):|x-y|\leq 2\}} |\mathbb{E}_x[\bar{\sigma}] - \mathbb{E}_y[\bar{\sigma}]| = \infty.$$

(iii) *The fluid model is unstable in the L_2 sense, as well as the almost sure sense of Dai [13] or Chen [14]. Moreover, the fluid model obtained in [13] with general (not necessarily exponential) service and interarrival times is also unstable in the L_2 -sense.*

Proof. To see (iii), note that the instability of the fluid model follows from Theorem 9, which shows that Theorem 16(ii) is violated, and hence that Theorem 16(iv) is also violated. In the general case where the arrival stream is not Poisson and the service times are not exponential, Dai [13] still obtains a fluid model which is virtually independent of these distributions. In particular, the exponential network when scaled as in [13] converges to a solution of the integral equations obtained in [13]. Hence, the solutions to these integral equations cannot converge to zero uniformly as required by [13] or [14].

To see (i), note that from Theorem 8, a Lipschitz solution $W(x)$ to (23) would imply that $W^2(x)$ would satisfy Theorem 16(i), implying the stability of the fluid model, a contradiction to (iii).

To see (ii), let $W(x) := E_x[\bar{\sigma}]$. Then $W(x)$ satisfies the drift in (23), and hence cannot be Lipschitz. \square

Hence if a non-copositive matrix Q is found satisfying the Drift LP, then given existing theory there is no hope in obtaining stability through the fluid model, and there is also no hope in obtaining a solution to Foster's condition (23), if one restricts attention to Lipschitz functions, such as piecewise linear functions. The following Theorem resolves the situation for copositive Q as well. Thus, from it and Theorem 17, one sees that a more detailed study of the fluid limit model for a particular problem, beyond that done above, is only needed when the Drift LP is infeasible, i.e., there is no feasible Q , and hence only when the Performance LP is unbounded.

Theorem 18 *Consider a system satisfying the auxiliary constraints (11). If a copositive matrix Q satisfying the drift condition (22) exists, then the fluid model of [13] is strongly stable in the sense of Chen [14].*

For any network with renewal type arrivals and services, with mean interarrival time $\frac{1}{\lambda}$, mean service times $\{\frac{1}{\mu_i}\}$, all with bounded second moments, the system is stable in the sense

that

$$\sup_N \frac{1}{N} \sum_{n=1}^N \mathbb{E}[|x(n)| \mid x(0) = x] < \infty.$$

If in addition, the interarrival times $\{v_i\}$ are unbounded, so that $P(v_i > t) > 0$ for all t , then the limit $\lim_{n \rightarrow \infty} \mathbb{E}[x(n) \mid x(0) = x]$ exists, and is independent of x .

Proof. The strong stability result is from [14], Theorem 3.6. The remaining conclusions are obtained by analyzing the sampled chain \hat{x}_k as in Theorem 16. \square

13 Concluding Remarks

We have exhibited some surprising and strong connections between the stability and performance problems for queueing networks and scheduling policies. Many interesting questions have arisen, that we have been unable to answer. For example, if the Performance LP is bounded, then is the system automatically stable? While we have answered it in the affirmative for the class of all stationary non-idling policies, we have not been able to resolve it for systems with auxiliary constraints arising from special scheduling classes or system structure. This is related to the issue of whether the dual Drift LP automatically generates a copositive matrix. In all the examples examined by us, the Drift LP has generated a non-negative matrix Q as a solution, whenever a solution exists. Resolving this would complete the circle of ideas tying together performance and stability in a duality framework. Second, it would be useful to resolve whether the existence of a non-copositive Q with negative drift actually implies the transience or null recurrence of the system. Third, whenever the Performance LP is unbounded, we are unable to assert anything regarding either performance or stability; the latter because there is no quadratic form with a negative drift. Thus, it becomes important to examine piecewise linear Lyapunov functions, as in Botvich and Zamyatin [23]. This raises many interesting questions regarding the construction of such functions, and their use in performance analysis.

References

- [1] D. Bertsekas and R. Gallager, *Data Networks*. Englewood Cliffs, NJ: Prentice Hall, 1987.
- [2] J. A. Buzacott and J. G. Shanthikumar, *Stochastic Models of Manufacturing Systems*. Prentice Hall, 1993.
- [3] S. B. Gershwin, *Manufacturing Systems Engineering*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [4] Stephen Lavenberg, Editor, *Computer Performance Modeling Handbook*. New York: Academic Press, 1983.
- [5] S. H. Lu and P. R. Kumar, “Distributed scheduling based on due dates and buffer priorities,” *IEEE Transactions on Automatic Control*, vol. 36, pp. 1406–1416, December 1991.
- [6] F. Baskett, K. M. Chandy, R. R. Muntz, and F. G. Palacios, “Open, closed and mixed networks of queues with different classes of customers,” *J. Assoc. Comput. Mach.*, vol. 22, pp. 248–260, April 1975.
- [7] F. P. Kelly, *Reversibility and Stochastic Networks*. New York, NY: John Wiley and Sons, 1979.
- [8] P. R. Kumar, “Re-entrant lines,” *Queueing Systems: Theory and Applications: Special Issue on Queueing Networks*, vol. 13, pp. 87–110, May 1993.
- [9] D. Bertsimas, I. Ch. Paschalidis and J. N. Tsitsiklis, “Optimization of multiclass queueing networks: Polyhedral and nonlinear characterizations of achievable performance.” Laboratory for Information and Decision Systems and Operations Research Center, M. I. T., December 1992.

- [10] S. Kumar and P. R. Kumar, “Performance bounds for queueing networks and scheduling policies,” technical report, Coordinated Science Laboratory, University of Illinois, Urbana, IL, 1992. To appear in *IEEE Transactions on Automatic Control*, 1994.
- [11] P. R. Kumar and S. P. Meyn, “Stability of queueing networks and scheduling policies,” tech. rep., C. S. L., University of Illinois, 1993.
- [12] A. N. Rybko and A. L. Stolyar, “On the ergodicity of stochastic processes describing open queueing networks,” *Problemy Peredachi Informatsii*, vol. 28, pp. 2–26, 1991.
- [13] J. G. Dai, “On the positive Harris recurrence for multiclass queueing networks: A unified approach via fluid models,” tech. rep., Georgia Institute of Technology, 1993.
- [14] H. Chen, “Fluid approximations and stability of multiclass queueing networks I: Work conserving disciplines,” tech. rep., University of British Columbia, 1993.
- [15] J. M. Harrison, “Brownian models of heterogeneous customer populations,” in *Stochastic Differential Systems, Stochastic Control Theory and Applications, IMA Volumes in Mathematics and its Applications* (W. Fleming and P. L. Lions, eds.), ch. 10, pp. 147–186, Springer-Verlag, 1988.
- [16] J. G. Dai and Y. Wang, “Nonexistence of Brownian models of certain multiclass queueing networks,” *Queueing Systems: Theory and Applications: Special Issue on Queueing Networks*, vol. 13, pp. 41–46, May 1993.
- [17] S. Lippman, “Applying a new device in the optimization of exponential queueing systems,” *Operations Research*, vol. 23, pp. 687–710, 1975.
- [18] D. Blackwell, “Positive dynamic programming,” in *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 415–418, 1965.
- [19] S. M. Ross, *Applied Probability Models with Optimization Applications*. San Francisco: Holden–Day, 1970.

- [20] H. Kushner, *Approximation and Weak Convergence Methods for Random Processes with Application to Stochastic Systems Theory*. Cambridge, MA: MIT Press, 1984.
- [21] G. Fayolle, “On random walks arising in queueing systems: Ergodicity and transience via quadratic forms as Lyapounov functions - part I,” *Queueing Systems*, vol. 5, pp. 167–184, 1989.
- [22] G. Fayolle, V. A. Malyshev, M. V. Mensikov and A. F. Sidorenko, “Lyapunov functions for Jackson networks,” *Rapports de Recherche No. 1380*, INRIA, Rocquencourt, France, 1991.
- [23] D. D. Botvich and A. A. Zamyatin, “Ergodicity of conservative communication networks,” *Rapports de Recherche No. 1772*, INRIA, Rocquencourt, France, 1992.
- [24] K. G. Murty, *Linear Programming*. New York, NY: John Wiley & Sons, 1983.
- [25] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*. Control and Communication in Engineering, London: Springer-Verlag, 1993.
- [26] J. L. Doob, *Stochastic Processes*. New York, NY: John Wiley & Sons, 1953.
- [27] R. W. Cottle, G. J. Habetler and C. E. Lemke, “On classes of copositive matrices,” *Linear Algebra and Its Applications*, vol. 3, pp. 295–310, 1970.
- [28] L.-E. Andersson, G. Chang, and T. Elfving, “Criteria for copositive matrices and non-negative Bezier patches,” *Tech. Rep. LiTH-MAT-R-93-27*, Linkoping University and University of Science and Technology, China, August 1993.
- [29] K. G. Murty and S. N. Kabadi, “Some NP-complete problems in quadratic and nonlinear programming,” *Mathematical Programming*, vol. 39, pp. 117–129, 1987.
- [30] V. A. Malyshev, “Networks and dynamical systems,” *Advances in Applied Probability*, vol. 25, pp. 140–175, 1993.

- [31] P. Billingsley, *Weak Convergence of Measures: Applications in Probability*. Philadelphia, PA: SIAM, 1971.