Performance bounds for queues via generating functions

Arie Hordijk
Mathematical Institute
Leiden University
The Netherlands
Phone: 31 71 527 71 46
Fax: 31 71 5276985
Email: hordijk@math.leidenuniv.nl

Adam Shwartz*
Electrical Engineering
Technion 32000, Israel
(Corresponding author)
Phone: 972 4 829 4743
Fax: 972 4 832 3041
Email: adam@ee.technion.ac.il

December 13, 1998. Revised October 10, 1999 and May 8, 2000 .

Abstract

Modern applications, e.g. vlsi manufacturing, give rise to complicated queueing models, often of the re-entrant type. Their complexity, together with implications of their performance, renewed interest in their performance and the computation of good control (e.g. scheduling) policies. Recent work concentrated on computable (mostly linear) performance bounds. We show that the linear bounds can be obtained naturally, and under weaker assumptions, using generating function techniques. This approach gives rise to a new class of bounds, on performance over busy periods.

*Work performed in part while this author was visiting Leiden University, and later the Free University: their support and hospitality are greatly appreciated. This research was supported in part by the NWO-project "Stochastic Networks," in part by the fund for promotion of research at the Technion, in part by the fund for promotion of sponsored research at the Technion, and in part by the Israel Science Foundation, administered by the Israel Academy of Sciences and Humanities.

1 The models

Consider a basic open re-entrant line model, as in [7]. Parts arrive as a Poisson process with rate λ to buffer b_1. After being processed they proceed to buffer b_2, then to b_3 etc., and finally to buffer b_L. There are S processing machines, and machine $\sigma(j)$ is responsible for processing the parts in buffer j. Service time at b_i is exponential with mean $1/\mu_i$. Although this class of models is somewhat restricted, it suffices to illustrate our methods, which apply to a much wider class of models.

Let x_i denote the number of parts in buffer i, and $x = (x_1, x_2, \dots, x_L)'$. The control action $w_i \in \{0, 1\}$ encodes whether machine $\sigma(i)$ serves buffer b_i (in which case $w_i = 1$). We set $w_i = 0$ whenever $x_i = 0$. Assume that for some $0 \leq \underline{\rho} \leq \bar{\rho}$ and each machine σ,

$$\underline{\rho} \cdot \mathbf{1}\left[x_i \neq 0 \text{ for some } i : \sigma(i) = \sigma\right] \leq \sum_{i:\sigma(i)=\sigma} w_i \leq \bar{\rho} \leq 1 \; . \tag{1.1}$$

Then $\underline{\rho} = 1$ is the non-idling case, and by definition $\bar{\rho} \leq 1$. Unlike some previous work, we do not make a-priori assumptions of boundedness of moments, and we do not restrict to non-idling policies. We use uniformization, and normalize the rates so that $\lambda + \sum \mu_i = 1$. Denote by τ_n the time of the nth event and by $\mathcal{F}_n$ the σ-algebra of events up to τ_n. Denote by $x(n)$ $(w(n))$ the state (control action) vectors at τ_n.

This class of models arises in modern applications, especially in the area of vlsi manufacturing. The models are too complex for exact analysis, and it is not possible in general to obtain the form of optimal policies. However, improving the performance of these systems is of great practical importance. Consequently, there is renewed interest and much ongoing work researching their performance and searching for good control (e.g. scheduling) policies. Recent work concentrated on computable performance bounds, mostly linear, for classes of policies: see e.g. [1, 6, 7, 5, 8]. We show that the linear bounds of [1, 6] can be obtained in a natural way (and under weaker assumptions) using generating function techniques. The same approach also gives rise to a new class of bounds, on performance over busy periods. We show that generating function techniques give rise to linear relations between key variables, such as the total expected workload (total number served per buffer) until the system empties. Such linear relations lead to performance bounds, when performance is measured in terms of total workload.

Our techniques are not limited to this model, but can be used for a variety of complex queueing systems. We conclude this section with a derivation of the basic relation between generating functions that arise in the model. Note that the generating functions determine the distribution, and so these are in fact relations between the distributions of the queue sizes as well as the controls. In Section 2 we state results in the steady state limit, under minimal assumptions: see e.g. Remark 2.2. Technical proofs are deferred to Section 5.1. In Section 3 we develop and motivate the a new family of relations, over busy periods. We illustrate how the generating function technique gives rise to linear relations between key variables, such as the total expected workload (total number served per buffer) until the system empties. Such linear relations lead to performance bounds, when performance is measured in terms of total workload. Technical proofs for this Section are delayed to Section 5.2.

We use the notation $a \cdot b$ to denote the scalar product of two vectors: $a \cdot b = \sum_{i=1}^{L} a_i b_i$.

1.1 The basic generating functions relation

Relations involving consecutive event-times follow directly from the description of the model. Setting $\mu_0 = \lambda$, we have

$$x_i(n+1) = x_i(n) + \begin{cases} 1 & \text{if } i = 1, \text{ with probability } \lambda, \\ w_{i-1}(n) & \text{if } i > 1 \text{ and } x_{i-1}(n) > 0, \text{ with probability } \mu_{i-1} \\ -w_i(n) & \text{if } x_i(n) > 0, \text{ with probability } \mu_i \\ 0 & \text{otherwise.} \end{cases} \tag{1.2}$$

Since this is a continuous time model with exponential holding times, at most one event may occur at any given time τ_n. Defining $w_0(n) = x_0(n) = 1$, the first case in (1.2) can be omitted while the second case applies to all i (including $i = 1$).

Lemma 1.1 *Suppose* $c_i \le 0,\ i = 1, \dots, L$. *Then*

$$\mathbb{E}\left[e^{c\cdot x(n+1)} \,\middle|\, \mathcal{F}_n\right] - e^{c\cdot x(n)}$$
$$= \left[\sum_{i=1}^{L} \left[\mu_{i-1} w_{i-1}(n) \left[e^{c_i} - 1\right] + \mu_i w_i(n) \left[e^{-c_i} - 1\right]\right]\right] e^{c\cdot x(n)}. \quad (1.3)$$

Remark: here and below, expectations and conditional expectations are well defined since w is bounded, x is positive and $c_i \le 0$, so that we are dealing with bounded terms.

Proof. Fix $c < 0$ and $i > 1$. For simplicity, we only derive the relation for the case $c_i = c,\ c_j = 0,\ j \neq i$. The relation (1.3) is obtained using exactly the same arguments. From (1.2), since there can be at most one event at a time,

$$\mathbb{E}\left[e^{cx_i(n+1)} \mid \mathcal{F}_n\right] = \Big[\mu_{i-1} w_{i-1}(n)[e^{c} - 1]\mathbf{1}\left[x_{i-1}(n) \neq 0\right]$$
$$+\mu_i w_i(n)[e^{-c} - 1]\mathbf{1}\left[x_i(n) \neq 0\right] + 1\Big] e^{cx_i(n)}. \quad (1.4)$$

Since $w_i(n) = 0$ whenever $x_i(n) = 0$, we can drop the indicator functions so that

$$\mathbb{E}\left[e^{cx_i(n+1)} \mid \mathcal{F}_n\right] - e^{cx_i(n)} = \left[\mu_{i-1} w_{i-1}(n) \left(e^{c} - 1\right) + \mu_i w_i(n) \left(e^{-c} - 1\right)\right] e^{cx_i(n)}. \quad (1.5)$$

Since $w_0 = x_0(n) = 1$ and $\mu_0 = \lambda$, the same derivation holds also for $i = 1$. ∎

Note that no assumptions are needed to derive this relation.

2 The steady state limit

In this section we assume stability, in the sense that the random variables of interest converge in distribution, and denote the limits as (w_j, x_i). We make no assumptions on the finiteness of moments [2]. Recall our notational convention $w_0(n) = 1 = w_0$.

Theorem 2.1 *Choose* $c_i \le 0$, $i = 1, \dots, L$. *Assume that for all* $i = 0, 1, \dots, L$, $(w_i(n), x(n))$ *converge in distribution to* (w_i, x). *Then*

$$\mathbb{E}\left[\sum_{i=1}^{L}\left[\mu_{i-1}w_{i-1}\left[e^{c_i}-1\right]+\mu_i w_i\left[e^{-c_i}-1\right]\right]e^{c\cdot x}\right]=0. \tag{2.1}$$

In particular, if we choose $c_j = 0$ *for all* $j \neq i$ *and set* $c_i = c$, *then*

$$\mathbb{E}\left[\left[\mu_{i-1}w_{i-1}\left[e^{c}-1\right]+\mu_i w_i\left[e^{-c}-1\right]\right]e^{cx_i}\right]=0. \tag{2.2}$$

Proof. Follows immediately upon taking expectations and limits in (1.3) and (1.5): Lemma 2.7 shows that limits as $n \to \infty$ exist, and that the limits of the left hand sides of (1.3) and (1.5) equal 0. ∎

Remark 2.2 *If the control policy is stationary and the busy cycle* τ *(cf.* (3.1)*) has finite expectation, then by Lemmas 2.8 and 5.1 the conditions of the Theorem hold, and therefore* (2.1)–(2.2) *hold.*

2.1 Flow balance and consecutive buffers

The flow balance relations hold without any assumptions on moments, and (in contrast with previous work) it is not necessary to restrict to non-idling policies.

Theorem 2.3 *Assume* $(w_i(n), x(n))$ *converges in distribution for each* i, *and denote the limits by* (w_i, x). *Then for each* i,

$$\mu_i\,\mathbb{E}\,w_i = \lambda \quad \textit{and} \tag{2.3}$$

$$\mu_i\,\mathbb{E}\,w_i x_i = \lambda + \mu_{i-1}\,\mathbb{E}\,w_{i-1}x_i \tag{2.4}$$

in the sense that either both sides are finite and equal, or both are infinite.

It is possible that both sides are infinite: only convergence in distribution is required. Moreover, the non idling condition is not used. However, this result does not guarantee that $\mu_i \, \mathbb{E}\, w_i(n) x_i(n)$ converges: indeed, it is possible that it is infinite while the limit is finite (or the converse may hold): cf. [4].

2.2 Further linear relations

The relations between non neighboring stations are as follows.

Theorem 2.4 *Assume $(w(n), x(n)$ converges in distribution to (w, x). If $\mathbb{E}\, x_i < \infty$ for some $i \neq 1$, then for each $j \neq 1$ such that $|i - j| > 1$,*

$$\mu_{i-1} \, \mathbb{E}\, w_{i-1} x_j + \mu_{j-1} \, \mathbb{E}\, w_{j-1} x_i = \mu_i \, \mathbb{E}\, w_i x_j + \mu_j \, \mathbb{E}\, w_j x_i \ , \tag{2.5}$$

and in particular $\mathbb{E}\, w_{i-1} x_j < \infty$ if and only if $\mathbb{E}\, w_i x_j < \infty$.

The results hold even when the mean is infinite: in this case, both sides are infinite.

2.3 Performance bounds

The performance of re-entrant lines is often measured through a steady state measure $\mathbb{E}\, C(x)$ for some function C, were x is the steady state value of the (vector) queue sizes. In general, re-entrant line models are too complex for exact analysis, let alone optimization, since the state space is infinite and few general structural results are available. As in [6], we show that the performance can be bounded by the solution of a mathematical program. The number of variables in (hence the dimension of) this program is at most $S \times (L + 1)$. It should be noted, though, that the mathematical program is not equivalent to the optimization problem, and in particular the number of equality and inequality constraints is smaller than the number of variables.

We say that the variables $\{z_{ij},\ i=0,\dots,L,\ j=1,\dots L\}$ are feasible if they satisfy

$$z_{ij} \ge 0 \tag{2.6}$$

$$\sum_{i:\sigma(i)=\sigma(j)} \frac{z_{ij}}{\underline{\rho}} \ge z_{0j} \ge \sum_{i:\sigma(i)=s} \frac{z_{ij}}{\bar{\rho}} \qquad s=1,\dots,S \tag{2.7}$$

$$\mu_i z_{ii} = \lambda + \mu_{i-1} z_{(i-1)i} \qquad i \ge 1 \tag{2.8}$$

$$\mu_{(i-1)} z_{(i-1)j} + \mu_{j-1} z_{(j-1)i} = \mu_i z_{ij} + \mu_j z_{ji} \qquad i,j \ge 2,\ i \ne j\,. \tag{2.9}$$

Denote by z_0 the L-dimensional vector with components z_{0j}. Given a function C, define

$$C_* = \inf\{C(z_0) : z_0 \text{ feasible}\}\ , \tag{2.10}$$

$$C^* = \sup\{C(z_0) : z_0 \text{ feasible}\}\ . \tag{2.11}$$

We can now state a lower as well as upper bound for linear performance functions, in terms of a finite linear program.

Theorem 2.5 *Assume $(w_i(n), x(n))$ converge in distribution and denote the limits by (w_i, x). If C is linear then $C_* \le \mathbb{E}\, C(x) = C(\mathbb{E}\, x) \le C^*$. If the control policy is non-idling, we can restrict the feasible region by*

$$z_{0j} = \sum_{i:\sigma(i)=\sigma(j)} z_{ij}\,. \tag{2.12}$$

Proof. Given the limit (w_i, x) define $z_{ij} = \mathbb{E}\,(w_i x_j)$. These obviously satisfy (2.6). By definition, $w_0 = 1$ so that $z_{0j} = \mathbb{E}\, x_j$, and since C is linear, $\mathbb{E}\, C(x) = C(\mathbb{E}\, x) = C(z_0)$. The relation (2.7) follows from (1.1). The relations (2.8) and (2.9) are exactly the relations (2.4) and (2.5). Thus any policy for which convergence in distribution holds gives rise to a feasible solution, and the upper and lower bounds follow. Finally, for a non-idling policy $\underline{\rho} = 1$ and (2.12) follows from (1.1). ∎

For convex functions, the bounds arise from a convex mathematical program.

Corollary 2.6 *Assume the conditions of Theorem 2.5. If C is convex then $C_* \leq \mathbb{E}\, C(x)$. If C is concave then $\mathbb{E}\, C(x) \leq C^*$. For non-idling policies,* (2.12) *holds.*

Proof. By Jenesen inequality, for C convex $\mathbb{E}\, C(x) \geq C(\mathbb{E}\, x)$. The arguments of Theorem 2.5 now give the result. The proof for the concave case is symmetric. ∎

We can incoporate assumptions on the policies into the mathematical program. For example, if we restrict to priority policies and station j has higher priority than i with $s = \sigma(i) = \sigma(j)$, then $w_i(n) = 1$ implies $x_j(n) = 0$ so $\mathbb{E}\, w_i(n) x_j(n) e^{c x_j(n)} = 0$. But

$$\lim_{c \uparrow 0} \mathbb{E}\, w_i(n) x_j(n) e^{c x_j(n)} = \mathbb{E}\, w_i x_j = z_{ij} \,, \tag{2.13}$$

so that $z_{ij} = 0$. This reduces the number of unknowns considerably: if server s serves K_s stations, a priority policy for server s will imply that $\frac{(K_s - 1) K_s}{2}$ variables are zero.

The same technique can be applied in order to derive further relations to provide tighter bounds. By similar manipulations of the basic relation (1.3) we can obtain new linear relations between the variables $z_{ijk} = \mathbb{E}\, w_i x_i x_j$. This results in a new linear program, with higher dimension, but one that provides tighter bounds.

2.4 Some technical results

Lemma 2.7 *Fix $c_i \leq 0$, $i = 1, \ldots, L$. Assume that $(w_i(n), x(n))$ converges in distribution, and denote the limit by (w_i, x). Then $\mathbb{E}\, w_i(n) e^{c \cdot x(n)} \to \mathbb{E}\, w_i e^{c \cdot x}$.*

Proof. Since $c_i \leq 0$, $x_i \geq 0$, and $0 \leq w \leq 1$, the function $w e^{c \cdot x}$ is bounded and continuous. The result follows since convergence in distribution implies convergence of expectations of bounded continuous functions (cf. [2]). ∎

Lemma 2.8 *Assume that $x(n)$ converges in distribution. If the control policy is stationary, then $(w_i(n), x(n))$ converge in distribution for all i.*

Proof. If the control policy is stationary, then $w_j(n)$ is a (continuous—since the state space is discrete) function of $x(n)$ and possibly of some additional independent randomization, where the distribution of this randomization may depend only on the value of x. This guarantees (cf. [2]) the joint convergence in distribution. ∎

3 Busy period relations

In this section we show that the relation (1.3) can be used to obtain relations of a new type, over busy periods. Our purpose here is to illustrate how the technique can be applied: it is clear, however, that using this technique it is possible to obtain performance bounds in the spirit of Sub-section 2.3, but in terms of busy periods. Such relations may be of interest in plants where there are large set-up costs, and the set-up may be changed only at the end of a busy periods (when all parts have cleared the machines). We note that stationarity is not assumed and is not relevant here: we are interested in the total values of the variables, summed over a busy period. No assumptions on the state or control are required, except "stability" in the sense that the empty state is reached.

Suppose all queues are empty at $n = 0$, that is $x(0) = 0$. Define

$$\tau = \min\{n > 0 : x(n) = 0\}. \tag{3.1}$$

The basic relation over a busy period is given in the following theorem.

Theorem 3.1 *Assume* $\mathbb{E}\,\tau < \infty$. *Then*

$$\mathbb{E}\left[\sum_{n=0}^{\tau-1}\left[\sum_{i=1}^{L}\left[\mu_{i-1}w_{i-1}(n)\left[e^{c_i}-1\right]+\mu_i w_i(n)\left[e^{-c_i}-1\right]\right]\right]e^{c\cdot x(n)}\right]=0. \tag{3.2}$$

Proof. Summing the left-hand-side of (1.3) from $n = 0$ to $\tau - 1$, taking expectation and using $x(0) = x(\tau) = 0$ we obtain

$$\begin{aligned}
&\mathbb{E}\sum_{n=0}^{\tau-1}\left[\mathbb{E}\left[e^{c\cdot x(n+1)} \mid \mathcal{F}_n\right] - e^{c\cdot x(n)}\right] \\
&\qquad = \mathbb{E}\sum_{n=0}^{\tau-1}\left[\mathbb{E}\left[e^{c\cdot x(n+1)} \mid \mathcal{F}_n\right] - e^{c\cdot x(n+1)}\right] \\
&\qquad = \mathbb{E}\sum_{n=0}^{\infty}\left[\mathbb{E}\left[\mathbf{1}\left[n < \tau\right] e^{c\cdot x(n+1)} \mid \mathcal{F}_n\right] - \mathbf{1}\left[n < \tau\right] e^{c\cdot x(n+1)}\right]
\end{aligned} \tag{3.3}$$

since $\{n < \tau\}$ is in $\mathcal{F}_n$. Since the last sum is bounded (in absolute value) by $2\,\mathbb{E}\,\tau$, we can take the expectation inside the sum. This shows that both sides of (3.3) equal 0. Now substitute (1.3) into the left-hand-side of (3.3) to obtain (3.2). ■

Remark: we can obtain an analogue of Theorem 3.1 over a busy period of one queue: if we set $\tau_i = \min\{n > 0 : x_i(n) = 0\}$, we can repeat the derivation of Theorem 3.1 as well as other results below to obtain relations under assumptions on τ_i.

3.1 Flow balance and consecutive buffers

As before, the basic Theorem 3.1 gives rise to linear relations, this time over busy period. Define $\tau^N = \min\{\tau, N\}$.

Theorem 3.2 *If $\tau < \infty$ w.p.1 then for all i,*

$$\mu_{i-1}\,\mathbb{E}\left[\sum_{n=0}^{\tau-1} w_{i-1}(n)\right] = \mu_i\,\mathbb{E}\left[\sum_{n=0}^{\tau-1} w_i(n)\right] \tag{3.4}$$

$$= \mu_i \lim_{N\to\infty} \mathbb{E}\left[\sum_{n=0}^{\tau^N-1} w_i(n)\right] \tag{3.5}$$

$$= \lambda\,\mathbb{E}\,\tau \tag{3.6}$$

so that if $\mathbb{E}\,\tau = \infty$*, all terms are infinite.*

Thus the total expected work brought in over a busy period equals the total expected work done on a queue, provided only the period is finite. This work is infinite unless the period has finite expectation.

Finally, we obtain a relation between products $(w_i(n)x_i(n))$ over a busy period.

Theorem 3.3 *If* $\mathbb{E}\,\tau < \infty$ *then*

$$\mu_i\,\mathbb{E}\left[\sum_{n=0}^{\tau-1} w_i(n)x_i(n)\right] = \mu_{i-1}\,\mathbb{E}\left[\sum_{n=0}^{\tau-1} w_{i-1}(n)x_i(n)\right] + \lambda\,\mathbb{E}\,\tau\ . \tag{3.7}$$

Additional relations can be obtained using the same methods, leading to performance bounds as in Subsection 2.3, but now for $\mathbb{E}\left[\sum_{n=0}^{\tau-1} x_1(n), \dots, \sum_{n=0}^{\tau-1} x_1(n)\right]$.

3.2 Some technical results

Corollary 3.4 *Define* $\tau^N = \min\{\tau, N\}$*. Then*

$$\mathbb{E}\left[\sum_{n=0}^{\tau^N-1}\left[\sum_{i=1}^{L}\left[\mu_{i-1}w_{i-1}(n)\left[e^{c_i}-1\right] + \mu_i w_i(n)\left[e^{-c_i}-1\right]\right]\right] e^{c\cdot x(n)}\right]$$
$$= 1 - \mathbb{E}\,e^{c\cdot x(\tau^N)}. \tag{3.8}$$

If $\tau < \infty$ *w.p.1 then, in addition,* $\lim_{N\to\infty}\left(1 - \mathbb{E}\,e^{c\cdot x(\tau^N)}\right) = 0$.

Proof. The first claim follows from (3.3) where now $e^{c\cdot x(0)} = 1 \neq e^{c\cdot x(\tau^N)}$. Here the change of order of summation and expectation is justified since the sum is finite (N terms at most). The second claim follows from the bounded convergence theorem since by assumption $\lim_{N\to\infty} x(\tau^N) = 0$, and $e^{c\cdot x(\tau^N)} \le 1$. ∎

Remark: The corollary gives information about the limit of the expectation, as $N \to \infty$. If $\mathbb{E}\,\tau = \infty$ then it does not provide information about the expectation of the limit, that is, about the expectation of the left hand side of (3.8), where τ^N is replaced with τ. Indeed, if $\mathbb{E}\,\tau = \infty$, if we define $\underline{c} = \min_i c_i$ then

$$\mathbb{E}\,\tau = \lim_{\underline{c}\uparrow 0} e^{\underline{c}}\,\mathbb{E}\,\tau = \lim_{\underline{c}\uparrow 0} \mathbb{E}\left[\sum_{n=0}^{\tau-1} e^{\underline{c}}\right] \tag{3.9}$$

$$\geq \lim_{\underline{c}\uparrow 0} \mathbb{E}\left[\sum_{n=0}^{\tau-1} e^{c\cdot x(n)}\right] = \mathbb{E}\,\tau \tag{3.10}$$

by monotone convergence and the definitions of τ and $\underline{c}$. Thus the expectation of the limit may be ill defined, being the difference of two infinite quantities.

4 Concluding remarks

We illustrated our approach by re-deriving some existing bounds, and obtaining new busy-period bounds. The approach gives stronger results in that weaker assumptions are needed, and typical results establish equalities even when the terms are not finite. The standard linear bounds have been used to obtain stability and performance bounds in [7, 8] and later papers. We intend to continue our investigation of the new approach, and in particular use our busy-period results to relate the bounds to state action frequencies, so as to obtain characterizations of useful policies.

5 Appendix

5.1 Proofs for Section 2

Proof of Theorem 2.3. We start with (2.2). Rearranging terms,

$$\mu_i\left[e^{-c} - 1\right]\mathbb{E}\,w_i e^{cx_i} = \mu_{i-1}\left[1 - e^{c}\right]\mathbb{E}\,w_{i-1}e^{cx_i} \tag{5.1}$$

$$\begin{aligned} \mu_i\,\mathbb{E}\,w_i e^{cx_i} &= \frac{1-e^{c}}{e^{-c}-1}\mu_{i-1}\,\mathbb{E}\,w_{i-1}e^{cx_i} \\ &= e^{c}\mu_{i-1}\,\mathbb{E}\,w_{i-1}e^{cx_i} \qquad \text{all } c<0\ . \end{aligned} \tag{5.2}$$

The functions $w_j e^{cx_i}$ are positive (meaning greater or equal to 0) and increase monotonically to the limit w_j as $c \uparrow 0$. Therefore, by the monotone convergence theorem,

$$\mu_i \, \mathbb{E} \, w_i = \lim_{c \uparrow 0} \mu_i \, \mathbb{E} \, w_i e^{cx_i} \tag{5.3}$$

$$= \lim_{c \uparrow 0} [e^c \mu_{i-1} \, \mathbb{E} \, w_{i-1} e^{cx_i}] \tag{5.4}$$

$$= \mu_{i-1} \, \mathbb{E} \, w_{i-1} = \lambda \tag{5.5}$$

where the last equality is just the case $i = 1$. Returning to the basic equation (5.2), by Lemma 5.2 (see Appendix) we can differentiate both sides with respect to c, and interchange derivatives and expectation. This yields

$$\mu_i \, \mathbb{E} \, w_i x_i e^{cx_i} = e^c \left(\mu_{i-1} \, \mathbb{E} \, w_{i-1} e^{cx_i} + \mu_{i-1} \, \mathbb{E} \, w_{i-1} x_i e^{cx_i} \right) \; . \tag{5.6}$$

Therefore, we obtain using the monotone convergence theorem as before

$$\mu_i \, \mathbb{E} \, w_i x_i = \lim_{c \uparrow 0} \mu_i \, \mathbb{E} \, w_i x_i e^{cx_i} \tag{5.7}$$

$$= \lim_{c \uparrow 0} \mu_{i-1} \, \mathbb{E} \, w_{i-1} e^{cx_i} + \lim_{c \uparrow 0} \mu_{i-1} \, \mathbb{E} \, w_{i-1} x_i e^{cx_i} \tag{5.8}$$

$$= \mu_{i-1} \, \mathbb{E} \, w_{i-1} + \mu_{i-1} \, \mathbb{E} \, w_{i-1} x_i \tag{5.9}$$

$$= \lambda + \mu_{i-1} \, \mathbb{E} \, w_{i-1} x_i \tag{5.10}$$

where either both sides are finite, or both are infinite. ∎

Lemma 5.1 *Assume the control policy is stationary. Then* $\{x(n)\}$ *converges in distribution if and only if* $\mathbb{E}_0 \, \tau < \infty$ *(cf.* (3.1)*).*

Proof. Due to the uniformization, the queueing process is a-periodic. The claim then follows from standard results on Markov chains [3]. ∎

Proof of Theorem 2.4. Fix $c_i < 0$ and $c_j < 0$, and set $c_k = 0$ for $k \notin \{i, j\}$. Starting with (2.1), after rearranging terms, we obtain

$$\begin{aligned}0 &= \mu_{i-1}\left(e^{c_i}-1\right)\mathbb{E}\, w_{i-1}e^{c_ix_i+c_jx_j} + \mu_{j-1}\left(e^{c_j}-1\right)\mathbb{E}\, w_{j-1}e^{c_ix_i+c_jx_j} \\ &\quad + \mu_i\left(e^{-c_i}-1\right)\mathbb{E}\, w_ie^{c_ix_i+c_jx_j} + \mu_j\left(e^{-c_j}-1\right)\mathbb{E}\, w_je^{c_ix_i+c_jx_j}\ .\end{aligned} \tag{5.11}$$

As before, we can differentiate and interchange differentiation with expectation. Taking partial derivatives with respect to c_i we obtain

$$\begin{aligned}0 &= \mu_{i-1}\left[e^{c_i}\,\mathbb{E}\, w_{i-1}e^{c_ix_i+c_jx_j} + \left(e^{c_i}-1\right)\mathbb{E}\, w_{i-1}x_ie^{c_ix_i+c_jx_j}\right] \\ &\quad + \mu_{j-1}\left(e^{c_j}-1\right)\mathbb{E}\, w_{j-1}x_ie^{c_ix_i+c_jx_j} \\ &\quad + \mu_i\left[-e^{-c_i}\,\mathbb{E}\, w_ie^{c_ix_i+c_jx_j} + \left(e^{-c_i}-1\right)\mathbb{E}\, w_ix_ie^{c_ix_i+c_jx_j}\right] \\ &\quad + \mu_j\left(e^{-c_j}-1\right)\mathbb{E}\, w_jx_ie^{c_ix_i+c_jx_j}\ .\end{aligned} \tag{5.12}$$

Taking partial derivatives now with respect to c_j we obtain

$$\begin{aligned}0 &= \mu_{i-1}\left[e^{c_i}\,\mathbb{E}\, w_{i-1}x_je^{c_ix_i+c_jx_j} + \left(e^{c_i}-1\right)\mathbb{E}\, w_{i-1}x_ix_je^{c_ix_i+c_jx_j}\right] \\ &\quad + \mu_{j-1}\left[e^{c_j}\,\mathbb{E}\, w_{j-1}x_ie^{c_ix_i+c_jx_j} + \left(e^{c_j}-1\right)\mathbb{E}\, w_{j-1}x_ix_je^{c_ix_i+c_jx_j}\right] \\ &\quad + \mu_i\left[-e^{-c_i}\,\mathbb{E}\, w_ix_je^{c_ix_i+c_jx_j} + \left(e^{-c_i}-1\right)\mathbb{E}\, w_ix_ix_je^{c_ix_i+c_jx_j}\right] \\ &\quad + \mu_j\left[-e^{-c_j}\,\mathbb{E}\, w_jx_ie^{c_ix_i+c_jx_j} + \left(e^{-c_j}-1\right)\mathbb{E}\, w_jx_ix_je^{c_ix_i+c_jx_j}\right]\ .\end{aligned} \tag{5.13}$$

Now set $c_i = c_j = c$. Due to Lemma 5.3, if x_i has finite expectation, then

$$\lim_{c\uparrow 0}\mathbb{E}\, w_lx_ix_je^{cx_i+cx_j}\left(e^c-1\right) = 0$$

with l taking the value $i-1,\ i,\ j-1$ or j. Moreover, by dominated convergence,

$$\lim_{c\uparrow 0} e^c\,\mathbb{E}\, w_{j-1}x_ie^{cx_i+cx_j} = \mathbb{E}\, w_{j-1}x_i \tag{5.14}$$

$$\lim_{c\uparrow 0} e^c\,\mathbb{E}\, w_jx_ie^{cx_i+cx_j} = \mathbb{E}\, w_jx_i\ , \tag{5.15}$$

and both are finite. Rearranging (5.13) we therefore have

$$\lim_{c\uparrow 0} \mu_{i-1} e^{c}\, \mathbb{E}\, w_{i-1} x_j e^{cx_i + cx_j} + \mu_{j-1}\, \mathbb{E}\, w_{j-1} x_i$$

$$= \lim_{c\uparrow 0} \mu_i e^{-c}\, \mathbb{E}\, w_i x_j e^{cx_i + cx_j} + \mu_j\, \mathbb{E}\, w_j x_i \quad (5.16)$$

in the sense that either both sides are finite or both are infinite. The result now follows by an application of the dominated (or of the monotone) convergence theorem. ∎

5.2 Proofs for Section 3

Proof of Theorem 3.2. Note that if $\mathbb{E}\,\tau$ is finite, we can start with (3.2), set $c_j = 0$ for $j \neq i$, divide by $e^{-c_i} - 1$ and then take the limit as $c_i \uparrow 0$ (using monotone convergence). The last equality is just the case $i = 1$.

More generally, set $c_j = 0$ for $j \neq i$ and rewrite (3.8) as

$$\mathbb{E}\left[\sum_{n=0}^{\tau^N - 1} \left[\mu_i w_i(n) \left[e^{-c_i} - 1\right] e^{c_i x_i(n)}\right]\right]$$

$$= \mathbb{E}\left[\sum_{n=0}^{\tau^N - 1} \left[\mu_{i-1} w_{i-1}(n) \left[1 - e^{c_i}\right] e^{c_i x_i(n)}\right]\right] + 1 - \mathbb{E}\, e^{c_i x_i(\tau^N)}. \quad (5.17)$$

Dividing by $e^{-c_i} - 1$ we obtain

$$\mu_i\, \mathbb{E}\left[\sum_{n=0}^{\tau^N - 1} w_i(n) e^{c_i x_i(n)}\right] = \mu_{i-1}\, \mathbb{E}\left[\sum_{n=0}^{\tau^N - 1} \left[w_{i-1}(n) e^{c_i (x_i(n) + 1)}\right]\right] + \frac{1 - \mathbb{E}\, e^{c_i x_i(\tau^N)}}{e^{-c_i} - 1}. \quad (5.18)$$

Now since all terms below are positive and increasing, all limits below exist and

$$\lim_{c_i \uparrow 0} \lim_{N \to \infty} \mu_i \, \mathbb{E}\left[\sum_{n=0}^{\tau^N - 1} w_i(n) e^{c_i x_i(n)}\right] = \lim_{c_i \uparrow 0} \mu_i \, \mathbb{E}\left[\sum_{n=0}^{\tau - 1} w_i(n) e^{c_i x_i(n)}\right] \tag{5.19}$$

$$= \mu_i \, \mathbb{E}\left[\sum_{n=0}^{\tau-1} w_i(n)\right] = \lim_{N \to \infty} \mu_i \, \mathbb{E}\left[\sum_{n=0}^{\tau^N - 1} w_i(n)\right] \tag{5.20}$$

$$= \lim_{N \to \infty} \lim_{c_i \uparrow 0} \mu_i \, \mathbb{E}\left[\sum_{n=0}^{\tau^N - 1} w_i(n) e^{c_i x_i(n)}\right] \tag{5.21}$$

by monotone convergence, and similarly for the first term on the right of (5.18), where these are possibly infinite. However, by bounded convergence,

$$\lim_{c_i \uparrow 0} \lim_{N \to \infty} \frac{1 - \mathbb{E}\, e^{c_i x_i(\tau^N)}}{e^{-c_i} - 1} = \lim_{c_i \uparrow 0} \frac{1 - \mathbb{E}\, e^{c_i x_i(\tau)}}{e^{-c_i} - 1} = 0 \tag{5.22}$$

since $x_i(\tau) = 0$, so that the nominator vanishes for all $c_i < 0$. ∎

Proof of Theorem 3.3. Starting with Equations (5.18) we let $N \to \infty$ and obtain by monotone convergence and (5.22)

$$\mu_i \, \mathbb{E}\left[\sum_{n=0}^{\tau-1} w_i(n) e^{c_i x_i(n)}\right] = \mu_{i-1} \, \mathbb{E}\left[\sum_{n=0}^{\tau-1} w_{i-1}(n) e^{c_i (x_i(n)+1)}\right] . \tag{5.23}$$

As in Lemma 5.2 we can show that the slope (with respect to c_i) is bounded by a constant times τ, so that we can interchange differentiation and expectation, to obtain

$$\frac{d}{dc}\left[\mu_i \, \mathbb{E}\left[\sum_{n=0}^{\tau-1} w_i(n) e^{c_i x_i(n)}\right]\right] = \mu_i \, \mathbb{E}\left[\frac{d}{dc}\left[\sum_{n=0}^{\tau-1} w_i(n) e^{c_i x_i(n)}\right]\right] \tag{5.24}$$

$$= \mu_i \, \mathbb{E}\left[\sum_{n=0}^{\tau-1} w_i(n) x_i(n) e^{c_i x_i(n)}\right] \tag{5.25}$$

and similarly

$$\frac{d}{dc}\left[\mu_{i-1}\,\mathbb{E}\left[\sum_{n=0}^{\tau-1}\left[w_{i-1}(n)e^{c_i(x_i(n)+1)}\right]\right]\right]$$

$$= \mu_{i-1}\,\mathbb{E}\left[\sum_{n=0}^{\tau-1}\left[w_{i-1}(n)(x_i(n)+1)e^{c_i(x_i(n)+1)}\right]\right] \quad (5.26)$$

Taking $c_i \uparrow 0$, using monotone convergence and Theorem 3.2 we obtain

$$\mu_i\,\mathbb{E}\left[\sum_{n=0}^{\tau-1} w_i(n)x_i(n)\right] = \lim_{c_i\uparrow 0}\mu_i\,\mathbb{E}\left[\sum_{n=0}^{\tau-1} w_i(n)x_i(n)e^{c_i x_i(n)}\right] \quad (5.27)$$

$$= \lim_{c_i\uparrow 0}\mu_{i-1}\,\mathbb{E}\left[\sum_{n=0}^{\tau-1}\left[w_{i-1}(n)x_i(n)e^{c_i(x_i(n)+1)}\right]\right] \quad (5.28)$$

$$+ \lim_{c_i\uparrow 0}\mu_{i-1}\,\mathbb{E}\left[\sum_{n=0}^{\tau-1}\left[w_{i-1}e^{c_i(x_i(n)+1)}\right]\right] \quad (5.29)$$

$$= \mu_{i-1}\,\mathbb{E}\left[\sum_{n=0}^{\tau-1}\left[w_{i-1}(n)x_i(n)\right]\right] + \lambda\,\mathbb{E}\,\tau \quad (5.30)$$

and the result is established. ∎

5.3 Technical estimates

Lemma 5.2 *Let x be a positive random variable and fix $c < 0$. For any $n \geq -1$,*

$$\frac{d}{dc}\,\mathbb{E}\left[x^n e^{cx}\right] = \mathbb{E}\left[\frac{d}{dc}\left(x^n e^{cx}\right)\right] = \mathbb{E}\left[x^{n+1}e^{cx}\right] \quad (5.31)$$

Proof.

$$\frac{d}{dc}\,\mathbb{E}\left[x^n e^{cx}\right] = \lim_{a\to c}\mathbb{E}\left[\frac{x^n e^{cx} - x^n e^{ax}}{c-a}\right] \quad (5.32)$$

but $$\left|\frac{x^n e^{cx} - x^n e^{ax}}{c-a}\right| \leq x^n \cdot x \cdot e^{cx/2} \quad (5.33)$$

for all a close enough to c. Since $x^{n+1}e^{cx/2}$ has finite expectation, the dominated convergence theorem implies that we can take the limit inside the expectation, that is, differentiate under the expectation. ∎

Lemma 5.3 *Let x and y be positive random variables. Then $\mathbb{E}\,x < \infty$ implies*

$$\lim_{c \uparrow 0} \mathbb{E}\, xy\,(e^c - 1)\, e^{cx+cy} = 0 \ .$$

Proof. Define $f(t,c) = te^{ct}$ for $t \geq 0,\ 0 > c$. Then for each fixed $c < 0$, the function f is positive, bounded in $t \geq 0$ and $f(0,c) = \lim_{t\to\infty} f(t,c) = 0$. Therefore, to find its maximum compute

$$0 = \frac{d}{dt} f(t,c) = e^{ct} + cte^{ct} \tag{5.34}$$

$$t_{\max} = \frac{-1}{c} \tag{5.35}$$

$$f(t_{\max}, c) = \frac{-1}{ce} \ . \tag{5.36}$$

Since $$\frac{d^2}{dt^2} f(t_{\max}, c) = ce^{-1} < 0 \ , \tag{5.37}$$

this is indeed the maximum. Since $e^c - 1$ is convex and $e^0 - 1 = 0$, we have that $c \leq e^c - 1$ for all c, so that $0 \leq 1 - e^c \leq -c$ for $c < 0$ and hence

$$\sup_{t\geq 0,\ 0>c\geq -1} f(t,c)\,(1 - e^c) = \sup_{0>c\geq -1} \frac{e^c - 1}{ce} \leq \frac{1}{e} < 1 \ . \tag{5.38}$$

Therefore

$$\left|\, xy\,(e^c - 1)\, e^{cx+cy} \,\right| = |\, xe^{cx} y e^{cy}\,(e^c - 1)\,| \tag{5.39}$$

$$\leq xe^{cx} < x \ . \tag{5.40}$$

Since by assumption $\mathbb{E}\, x < \infty$ we can invoke the dominated convergence theorem to conclude that

$$\lim_{c \uparrow 0} \mathbb{E}\, xy\,(e^c - 1)\, e^{cx+cy} = \mathbb{E} \lim_{c \uparrow 0} xy\,(e^c - 1)\, e^{cx+cy} = 0\ . \tag{5.41}$$

■

References

[1] D. Bertsimas, I.C. Paschalidis and J.N. Tsitsiklis, "Optimization of multiclass queueing networks: Polyhedral and nonlinear characterization of achievable performance," *Ann. Applied Prob.* **4** pp. 43–75, 1994.

[2] P. Billingsley, *Convergence of probability measures*, Wiley, NY, 1968.

[3] K.L. Chung, *Markov chains with stationary transition probabilities*, Springer Verlag, New York, 1967.

[4] P.J. Holewijn and A. Hordijk, "On the convergence of moments in stationary Markov chains," *Stochastic Processes and Their Applications* **3** pp. 55-64, 1975.

[5] Z. Liu, "Performance analysis of stochastic timed Petri nets using linear programming," *IEEE Trans. Software Eng.* **24** pp. 1014–1030, 1998.

[6] S. Kumar and P.R. Kumar, "Performance bounds for queueing networks and scheduling policies," *IEEE Transactions on Automatic Control* **39** pp. 1600-1611, 1994.

[7] P.R. Kumar and S.P. Meyn, "Duality and Linear Programs for Stability and Performance Analysis of Queueing Networks and Scheduling Policies," *IEEE Transactions on Automatic Control* **41** pp. 4–17, 1996.

[8] J.R. Morrison and P.R. Kumar, "New Linear Program Performance Bounds for Queueing Networks," *Journal of Optimization Theory and Applications* **100**, pp. 575–597, 1999.