A simple throughput bound for large closed queueing networks with finite capacities

van Dijk, N.M.

1989

SERIE RESEARCH MEMORANDA

A SIMPLE THROUGHPUT BOUND FOR LARGE
CLOSED QUEUEING NETWORKS WITH FINITE CAPACITIES

Nico M. van Dijk

Research Memorandum 1989-5
January 1989

VRIJE UNIVERSITEIT
FACULTEIT DER ECONOMISCHE WETENSCHAPPEN
EN ECONOMETRIE
AMSTERDAM

A SIMPLE THROUGHPUT BOUND FOR LARGE CLOSED QUEUEING NETWORKS WITH FINITE CAPACITIES

Nico M. van Dijk
Free University, Amsterdam, The Netherlands

Abstract Arbitrary closed exponential queueing networks are examined such as with finite capacities and blocking or dynamic routing. Conditions are provided from which a simple upper bound on the system throughput and an error bound for its accuracy can be concluded. These conditions are illustrated for a Jackson network with finite entrance buffer and overflow station. An explicit error bound for this application is derived of order M^{-1}, where M is the number of jobs.

Keywords Closed queueing network * throughput bound * error bound.

1 Introduction

Queueing networks have gained a wide popularity as a powerful modeling and evaluation tool in communication analysis, computer performance evaluation and flexible manufacturing. At first glance, most of these queueing network applications are to be seen as open, since jobs are usually generated exteriorly and depart the system upon completion of their required services. Typical present day applications, however, such as a multi-source computer system or a computerized assembly line feature an almost infinite input so that upon completion of a job or departure from the system a new job is instantly inserted. In such applications the number of jobs is usually large. Other systems of actual interest for which a closed queueing network modeling is appropriate are interconnection networks such as CSMA or BTMA broadcasting systems (cf. [15]). In this case the jobs of the queueing network modeling represent the transmitters which are fixed but usually large in number (cf. [9]).

When the closed queueing network exhibits a product form various techniques such as mean value analysis (cf. [10]) or statistical mechanics (cf. [9]) are available for computing performance measures of interest. With a large number of jobs, however, such methods become computationally expensive. To this end, also bounding methods for large product form networks have been developed (cf. [7], [25]). Unfortunately, practical features such as most notably finite capacity constraints generally destroy the celebrated product form expression (e.g. [4], [11]). Simple performance bounds for specific small non-product systems have recently been developed as based upon a product-form modification methodology (cf. [18], [21], [22]). For large scale non-product networks, however, this methodology is not appropriate and no simple general performance bounds seem to be available. As numerical computations easily grow astronomically, robust but simple guaranteed bounds would be useful for at least quick evaluation purposes.

This paper, therefore, will secure a simple upper bound as well as an error bound of its accuracy for the system throughput of large closed queueing networks, regardless of a product-form or not. This bound is based upon comparing the closed system with an appropriate open analog for which the throughput is trivially obtained. For concrete networks the order of accuracy will generally be reciprocal in the number of jobs and an explicit error bound can be derived.

Convergence results for approximating open systems by closed systems with a finite source input tending to infinity have been established (cf. [2], [8], [23]). Explicit error bounds, however, have not been reported other than for simple standard Erlang-type systems (cf. [23]). Error bounds for somewhat related state space truncation results have been proposed in [12], but these are just robust bounds which do not secure an order of accuracy.

The essential underlying condition to the results of this paper is a boundedness condition for so-called bias terms of total reward structures. The verification of this condition is the crucial part. In concrete situations an inductive Markov reward proof-technique can be succesful. This technique, which can be seen as a partial extension of monotonicity proof-techniques such as developed in [1], [13] and [14], has already been fruitful in simple network situations for slightly related problems (cf. [18], [22]). However, as complex technicalities are involved, it cannot be guaranteed generally to work well, especially not for multi-dimensional situations such as the queueing networks studied in this paper.

The main part of this paper, therefore, is concerned with illustrating how the necessary conditions can be verified for a particular non-product form queueing network. This concerns a Jackson network with a finite entrance station and overflow upon saturation, such as naturally arising in communication systems with alternate routing, packet switching or manufacturing.

A simple throughput approximation λ is shown to be an upper estimate with an explicit error bound

$$\lambda W M^{-1}$$

where W is a sojourn time as easily estimated from above by a standard product form network and where M is the number of jobs of the closed system. Extensions of this application such as to more capacity constraints are possible.

The organization is as follows. First, the general closed model is presented in section 2.1. Next, in section 2.2 an open analog is proposed which suggests a simple estimate (bound) λ for the throughput of the closed system. Conditions so as to guarantee that this estimate is an upper estimate as well as to obtain an explicit error bound are derived in section 2.3. These conditions are verified for a network application with overflow in section 3. An evaluation concludes the paper.

2 Result

2.1 General model

Consider an arbitrary closed exponential queueing network with M jobs and N+1 service stations numbered 0,1,...,N, such that for at least one station, say with number 0, the departure rate at which jobs actually leave this station is given by

$$\mu_0(n_0) \geq 0$$

when n_0 jobs are present at station 0, regardless of the number of jobs n_i at the other stations $i=1,\ldots,N$. As $n = n_1+\ldots+n_N$ uniquely determines the number $n_0 = M-n$, the system can thus be described by a continuous-time Markov chain with state vector $\bar{n} = (n_1\ldots,n_N)$. From now on we will refer to the stations 1,...,N as the "main system" and we write

$$\mu(n) = \mu_0(M-n).$$

Throughout we always let vectors have N-components and we use the notation $\bar{n} + e_i$ and $\bar{n} - e_i$ to denote the vector equal to $\bar{n}$ up to one job more respectively less at station i. The vector $\bar{n}-e_i+e_j$ thus denotes the state equal to $\bar{n}$ with one job moved from station i to j. Here, we allow i=0 or j=0 under the convention that $\bar{n} \pm e_0 = \bar{n}$ so as to model an arrival at (i=0) or departure from (j=0) the "main system". The transition rates $\bar{q}(\bar{n},\bar{n}-e_i+e_j)$ for a transition from a state $\bar{n}$ into $\bar{n}-e_i+e_j \neq \bar{n}$ can then be formulated as:

$$\bar{q}(\bar{n},\bar{n}+e_j) = \mu(n)\ \alpha_j(\bar{n}) \qquad (1 \le j \le N)$$

$$\bar{q}(\bar{n},\bar{n}-e_i+e_j) = \mu_{ij}(\bar{n}) \qquad (i,j \neq 0) \tag{2.1}$$

$$\bar{q}(\bar{n},\bar{n}-e_i) = \mu_i(\bar{n}) \qquad (i \neq 0)$$

where by assumption: $\alpha_1(\bar{n})+\ldots+\alpha_N(\bar{n}) = 1$. Without restriction of generality, also assume that the corresponding Markov chain has a unique stationary distribution $\bar{\pi}(\bar{n})$ restricted to $\bar{n} \in \bar{S}$ for some set $\bar{S}$ with $\bar{0} =(0,\ldots,0) \in \bar{S}$. Our objective is to evaluate the throughput $\bar{\lambda}$ of the main system, i.e.

$$\bar{\lambda} = \Sigma_{\bar{n}\in\bar{S}}\ \bar{\pi}(\bar{n})\ \mu_i(\bar{n}). \tag{2.2}$$

Remark 2.1 (Arrival blocking) Our formulation seems to exclude arrival blocking upon departures from station 0 as we assume $\mu_0(n_0) > 0$ for all n_0 and $\alpha_1(\bar{n})+\ldots+\alpha_N(\bar{n}) = 1$ for all $\bar{n}$. Clearly, blocking probabilities strictly less than 1 can hereby be modeled by scaling $\mu_0(n_0)$. Strict arrival blocking, however, that is with probability 1 such as due to a finite capacity constraint, can also be modeled by including an additional station, say *, to which blocked arrivals from station 0 are rerouted. For example, assuming that blocked arrivals are originally rerouted to station 0, this can be modeled by routing the blocked jobs from station 0 to station * which serves at a rate:

$$\mu_*(n_*) = \mu_0(n_0+n_*) - \mu_0(n_0).$$

Whenever n_0 jobs are still present at station 0 and n_* jobs are blocked jobs from station 0 that have still not accessed one of the stations 1,...,N. Departures from station * route as coming from station 0 and upon blocking return again to this station *.

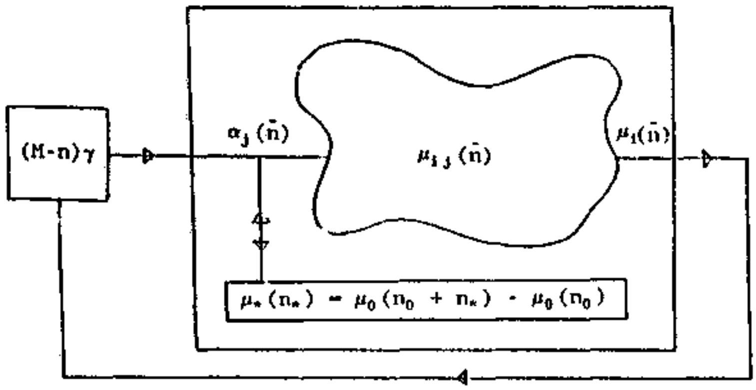

Figure 1

Clearly, various blocking protocols for routings inside the main system (i.e., stations 1,...,N) and departures from this main system are included as we allow a generally state dependent transition structure. For example, stations may have finite capacity constraints where upon saturation a newly arriving job may be returned to its source station (communication protocol) or rerouted to another station for possible access (overflow).

2.2 Open analog

Let

$$\lambda = \max_{\bar{n} \in \bar{S}} \mu(n) \tag{2.3}$$

and consider the open analog of the queueing network of section 2.1 with stations 1,...,N unchanged, but station 0 replaced by a Poisson input with parameter λ. More precisely, the corresponding transition rates $q(\bar{n},\bar{n}-e_i+e_j)$ for $\bar{n},\bar{n}-e_i+e_j \in \bar{S}$ are now given by

$$\begin{aligned} q(\bar{n},\bar{n}+e_j) &= \lambda\ \alpha_j(\bar{n}) && (1 \le j \le N) \\ q(\bar{n},\bar{n}-e_i+e_j) &= \mu_{ij}(\bar{n}) && (i,j \ne 0) \\ q(\bar{n},\bar{n}-e_i) &= \mu_i(\bar{n}) && (i \ne 0) \end{aligned} \tag{2.4}$$

As for transitions not restricted to $\bar{S}$ (thus with $\bar{n}\notin S$ and/or $\bar{n}-e_i+e_j\notin\bar{S}$) these transition rates can be arbitrarily chosen up to the conditions:

(i) The system is always fed by a Poisson arrival rate λ.

(ii) The corresponding Markov chain also has a unique stationary distribution $\pi(\bar{n})$restricted to some set S where $S\supset\bar{S}$.

(iii) The transition rates remain uniformly bounded, i.e. for some $Q < \infty$:

$$Q \geq \sup_{\bar{n}\in S} \Sigma_{i,j}\, q(\bar{n},\bar{n}-e_i+e_j).$$

As a consequence, since any Poisson arrival is accepted with probability 1 while the empty state $\bar{0}=(0,\ldots,0)\in\bar{S}\subset S$, so that any job will eventually leave the system, the system throughput of this open analog is equal to λ.

Remark 2.2 It is noted that this open analog needs to be formulated merely to prove our result of interest. The freedom of formulation outside $\bar{S}$ will play a role only in the verification of the necessary conditions and thereby the accuracy of the error bounds. In concrete situations, however, a natural extrapolation of the interior transition structure of $\bar{S}$ is likely to be a 'good' candidate, as will be illustrated in section 3.

2.3 Comparison results

Below we always denote an expression for the original closed system with an upper bar symbol "-" while for the open analog without upper bar. An upper bar symbol between parantheses (-) indicates that the expression is to be read for both the closed and open version. By virtue of the standard uniformization technique (e.g. [16], p.110) the stationary distributions $\bar{\pi}(.)$ and $\pi(\ )$ are equal to those of the discrete-time Markov chains with one-step transition probabilities $\bar{p}(\bar{n},\bar{n}-e_i+e_j)$ and $p(\bar{n},\bar{n}-e_i+e_j)$ respectively defined by

$$\overset{(-)}{p}(\bar{n},\bar{n}-e_i+e_j) = \begin{cases} \overset{(-)}{q}(\bar{n},\bar{n}-e_i+e_j)/Q & j\neq i \\ \\ [1 - \Sigma_{j\neq i} \overset{(-)}{q}(\bar{n},\bar{n}-e_i+e_j)/Q] & j=i \end{cases} \tag{2.5}$$

For evaluating the steady-state behaviour, we may thus restrict to these discrete-time Markov chains. To this end, let the one-step expectation operators ${}^{(\bar{~})}T_t$, t=0,1,2,... upon real-valued functions f(.) be defined by

$$\begin{cases} {}^{(\bar{~})}T_0 = I \\ {}^{(\bar{~})}T_{t+1} = {}^{(\bar{~})}T\,{}^{(\bar{~})}T_t \quad t=0,1,2,\ldots \\ {}^{(\bar{~})}T f(\bar{n}) = \Sigma_{i,j}\,{}^{(\bar{~})}p(\bar{n},\bar{n}-e_i+e_j)\; f(\bar{n}-e_i+e_j) \end{cases} \tag{2.6}$$

where these definitions are restricted to states $\bar{n} \in {}^{(\bar{~})}S$. Now in order to compute the system throughput, for s=0,1,2,... let

$${}^{(\bar{~})}V_s = \Sigma_{t=0}^{s-1}\,{}^{(\bar{~})}T_t r, \tag{2.7}$$

where the function r(.) is given by

$$r(\bar{n}) = \Sigma_{i=1}^{N} \mu_i(\bar{n})/Q. \tag{2.8}$$

Then, from the uniformization (2.5) and standard Markov reward theory, we conclude that for arbitrary $\bar{\ell} \in {}^{(\bar{~})}S$ the system throughputs are given by

$${}^{(\bar{~})}\lambda = \lim_{s\to\infty} \frac{Q}{s}\,{}^{(\bar{~})}V_s\;(\bar{\ell}) \tag{2.9}$$

The following simple key-theorem can now be proven. It gives a natural condition for the directly computable value λ from (2.3) to secure an upper bound for the system throughput $\bar{\lambda}$ of the original closed system and most importantly it provides a tool for guaranteeing an error bound.

Theorem 2.1

(i) We have

$$\lambda \geq \bar{\lambda}, \tag{2.10}$$

if for all $\bar{n},\bar{n}+e_j \in S$ and $t \geq 0$:

$$V_t(\bar{n}+e_j) \geq V_t(\bar{n}) \tag{2.11}$$

(ii) We have

$$|\lambda-\bar{\lambda}| \leq \varepsilon\ C, \tag{2.12}$$

if for some function $\beta(.)$, some initial state $\bar{\ell}$, some constants C and $\varepsilon>0$, and all $\bar{n}\in\bar{S}$:

$$\hat{T}_t\ \beta(\bar{\ell}) \leq C \quad (t \geq 0) \tag{2.13}$$

$$|[\lambda-\mu(n)]\ \Sigma_{j=1}^{N}\alpha_j(\bar{n})[V_t(\bar{n}+e_j)-V_t(\bar{n})]| \leq \varepsilon\ \beta(\bar{n}) \quad (t \geq 0) \tag{2.14}$$

Proof

(i) From (2.7) we conclude

$$^{(\bar{\ })}V_{t+1} = r + ^{(\bar{\ })}T\ ^{(\bar{\ })}V_t \tag{2.15}$$

As the transition probabilities $\bar{p}(.,.)$ remain restricted to $\bar{S}$ while $\bar{S}\subset S$, we may thus write for arbitrary $\bar{\ell}\in\bar{S}$:

$$(\bar{V}_s - V_s)(\bar{\ell}) = (\bar{T}\ V_{s-1} - TV_{s-1})(\bar{\ell})$$

$$= (\bar{T} - T)V_{s-1}(\bar{\ell}) + \bar{T}(\bar{V}_{s-1} - V_{s-1})(\bar{\ell})$$

$$= \Sigma_{t=0}^{s-1}\ \bar{T}_t(\bar{T}-T)V_{s-t-1}(\bar{\ell}) + \bar{T}_s(\bar{V}_0-V_0)(\bar{\ell}) \tag{2.16}$$

where the latter equality follows by iteration. The second term of the last right-hand side is equal to 0 as $^{(\bar{\ })}V_0(.)=0$ by definition. From comparing (2.1) and (2.4) and substituting (2.5), we find for any $\bar{n}\in\bar{S}$ and $t\geq0$:

$$(\bar{T}-T)V_t(\bar{n}) = [\lambda \Sigma_{j=1}^{N} \alpha_j(\bar{n}) V_t(\bar{n}+e_j) -$$

$$\mu(n) \Sigma_{j=1}^{N} \alpha_j(\bar{n}) V_t(\bar{n}+e_j)]/Q -$$

$$[(\lambda - \mu(n)) \Sigma_{j=1}^{N} \alpha_j(\bar{n}) V_t(\bar{n})]/Q$$

$$= (\lambda - \mu(n)) \Sigma_{j=1}^{N} \alpha_j(\bar{n})[V_t(\bar{n}+e_j)-V_t(\bar{n})]/Q \qquad (2.17)$$

Since the expectation operators $\bar{T}_t$ are monotone operators (i.e., $\bar{T}_t f \geq 0$ if $f \geq 0$ componentwise), inequality (2.10) now directly follows from (2.2), (2.3), (2.11), (2.16), (2.17) and (2.9).

(ii) From noting that also $\bar{T}_t f \leq \bar{T}_t g$ for any t and $f \leq g$ in the componentwise sense, taking absolute values in (2.16) and substituting (2.17), (2.13) and (2.14) we conclude:

$$|(\bar{V}_s - V_s)(\bar{\ell})| \leq \varepsilon\, Q^{-1} \Sigma_{t=0}^{s-1} \bar{T}_t\, \beta(\bar{\ell}) \leq \varepsilon\, s\, Q^{-1}\, C.$$

Applying (2.9) now proves (2.12). □

Remark 2.3 (Upper bound λ?) At first glance, inequality (2.10) seems trivial and generally valid. Counterintuitively, however, one can give counterexamples (see [1], [18], [22]) showing that the throughput of a system for a specific realization can be decreased by allowing more arrivals. Roughly speaking, the intuition seems to be correct for exponential services and monotone service rates. For specific situations, such as for assembly type networks without feedbacks, formal proofs for this latter type statement have recently been established in [1], [13], [22].

Remark 2.4 (Conditions 2.13 and 2.14) For the conditions (2.13) and (2.14) one must typically think of $\beta(.)$ to be some polynomial in n such as $\beta(\bar{n})=n$, and the so-called bias terms $V_t(\bar{n}+e_j)-V_t(\bar{n})$ to be bounded uniformly in $t\geq0$. This latter boundedness is standardly known in Markov reward theory as based upon mean first passage times. As these times, however, are extremely hard to obtain analytically for multi-dimensional processes (cf. [6] for one-dimensional situations), while no general bounds for them seem to be available, in the next section we will give an approach by which these bias-terms are estimated directly. Roughly speaking, condition (2.14) then requires that in any state either

(i) the difference $\lambda-\mu(n)$ is sufficiently small or
(ii) the marginal probability itself of being in states where this difference is large, is small. The example in the next section will include both aspects.

Remark 2.5 (Mixed open-closed networks). Clearly, the setting can be modified so as to include networks with both a fixed number of internal jobs that will never leave the system and external arrivals that will eventually depart. The throughput of the internal jobs can then be reformulated and estimated from above in a similar manner.

Remark 2.6 (Multiple job-types). For expository convenience the presentation has been restricted to a single job class. As general state dependent intensities are allowed of the form (2.1), we could just as well have included interdependencies of multiple-job types. Assuming that for at least one station jobs of any type can always depart regardless of the state of the other stations, (which does not exclude the routing probabilities $\alpha_j(.)$ to be state dependent), similar throughput estimates for multiple-job classes can be stated.

Remark 2.7 (Non-exponential case). The exponential assumption has been made to justify a Markovian analysis. It is standardly known, however, (e.g. [3]), that nonnegative distributions can be approximated arbitrarily closely, in the sense of weak convergence, by mixtures of Erlang distributions. As these mixtures still guarantee an exponential and thus Markovian structure while general state dependent intensities are allowed, generalizations to non-exponential networks can be expected. The technical details, however, will be highly complex particularly as for estimating the bias-terms in concrete situations (cf. [20]).

Remark 2.8 (Bounded intensities). The boundedness assumption (iii) in section 2.2 was made in order to apply the uniformization technique for obtaining a discrete-time formulation. This, however, can be avoided in a technical manner similarly to [19] such as to allow infinite server stations.

3 Application: A Jackson queueing network with a finite entrance buffer and overflow.

This section will illustrate how the conditions (2.11), (2.13) and (2.14) can be verified for a concrete network of practical interest. For expository convenience we restrict our illustration to a finite source Jackson network with a finite entrance station and overflow upon saturation of this station. Two essential features, a non-reversible finite capacity constraint and a dynamic routing, both of which generally destroy a product form expression, are hereby involved. With more finite capacity constraints the proofs become much more complicated (such as in [22]) but follow essentially the same approach. An overflow phenomenon has only been dealt with in the literature for simple Erlang type systems (cf. [17]) and is therefore included. Based upon the results obtained below and the above references, similar results can be expected along the same lines for systems with more complex finite capacity constraints.

3.1 Model

Consider a Jackson network with N-1 service stations, numbered 1,...,N-1, a separate overflow station N and a finite source input with M sources, such as visualized in figure 2 below.

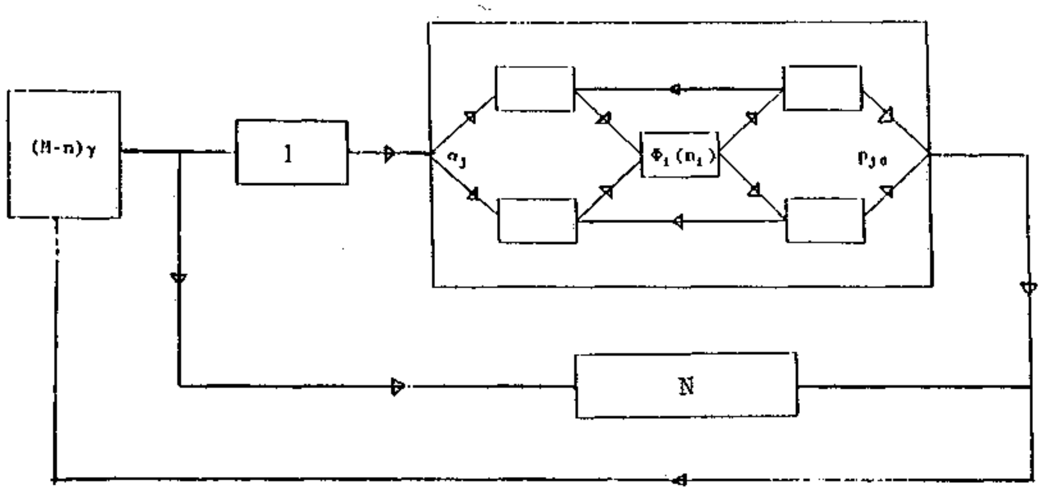

Figure 2

The service rate at station i when n_i customers are present is given by $\Phi_i(n_i)$. Upon service completion at station i a job routes to station j with probability p_{ij}, j=1,...,N-1 or leaves the system with probability $p_{i0} = 1 - \Sigma_{j\leq N-1} p_{ij}$, where $p_{N0} = 1$. Arrivals at the system are generated by a finite source input with M sources and exponential source times with parameter γ. That is, when n jobs are within the system the arrival rate is $(M-n)\gamma$. An arrival will enter the system at some entrance (buffer) station 1 which is to be seen as the bottleneck of the system as it has a finite capacity constraint of no more than B jobs. Upon saturation of this buffer-station, an arriving job is rerouted to a separate overflow station N. Switching from station N to the primary system at some later time is not allowed. Further, the matrix $(p_{ij})_{i,j=0,1,\ldots,N-1}$ is assumed to be irreducible. A unique steady state distribution and system throughput is thereby secured. Further, the service rates are assumed to be nondecreasing in the queue length, i.e. for any i=1,...,N and $n_i \geq 0$:

$$\Phi_i(n_i+1) \geq \Phi_i(n_i) \tag{3.1}$$

The system under consideration is not of product-form (cf. [4]). For the special case where the primary system is just an Erlang loss system, the overflow stream can be shown to be hyperexponential (cf. [17]), so that the overflow station in isolation can then be analyzed by a $GI|M|.$-system. Even then, however, an analytic expression for the system throughput $\tilde{\lambda}$ is not available as this depends on both the primary and overflow part. Therefore we wish to investigate the estimate

$$\lambda = \gamma M \tag{3.2}$$

3.2 Parametrizations

With $\bar{n} = (n_1, \ldots, n_N)$ denoting the number of jobs n_i at station i, i=1,...,N and $n=n_1+\ldots+n_N$, the above system is parametrized in the setting of section 2.1 by:

$$\mu(n) = (M-n)\gamma$$

$$\mu_{ij}(\bar{n}) = \Phi_i(n_i)p_{ij}$$

$$\mu_i(\bar{n}) = \Phi_i(n_i)p_{i0}$$

$$\alpha_j(\bar{n}) = 1 \quad \text{for} \begin{cases} j = 1 & \text{if } n_1 < B \\ j = N & \text{if } n_1 = B \end{cases} \tag{3.3}$$

Further, note that the assumption $\alpha_1(\bar{n})+\ldots+\alpha_N(\bar{n})=1$ is satisfied by formulation, while $\bar{0} = (0,0,\ldots,0) \in \tilde{S}$ where $\tilde{S}$ is the state space

$$\tilde{S} = \{\bar{n} | 0 \leq n_1 \leq B,\ n_1 + \ldots + n_N \leq M,\ n_i \geq 0,\ i = 2,\ldots,N\} \tag{3.4}$$

of admissible states. With λ given by (3.2) and

$$S = \{\bar{n} | 0 \leq n_1 \leq B,\ n_i \geq 0,\ \ i = 2,\ldots,N\}, \tag{3.5}$$

a natural open analog with Poisson input λ and state space $S \supset \tilde{S}$ in the setting of section 2.2 is then obtained by (2.4), as given by the parametrization (3.3) and the extrapolations:

$$\Phi_i(n_i) = \Phi_i(M) \qquad (n_i \geq M) \quad (i = 2,\ldots,N)$$

$$\Phi_1(n_1) = \Phi_1(B) \qquad (n_1 \geq B) \tag{3.6}$$

In words that is, the open analog is just the infinite version of the original system due to a Poisson input and service rates kept to their original maxima. The assumptions (i) and (iii) of section 2.2 are hereby guaranteed while assumption (ii) will be satisfied under the natural condition: $\lambda/\Phi_N(M) < 1$. We are thus able to apply the results of section 2.3.

3.3 Comparison result and error bound

We adopt all notation from section 2.3 based upon the above parametrizations. The following lemma is the most crucial step towards applying theorem 2.1. To this end, for arbitrary function g, we introduce the notation:

$$\Delta_j g(\bar{n}) = g(\bar{n}+e_j) - g(\bar{n}) \qquad (j = 1,\ldots,N) \tag{3.7}$$

Lemma 3.1 For all $t \geq 0$ and $\bar{n}+e_i \in S$:

$$0 \leq \Delta_i V_t(\bar{n}) \leq 1 \tag{3.8}$$

Proof This will be given by induction to t. For t=0, (3.8) trivially holds as $V_0(.) = 0$. Suppose that (3.8) holds for $t \leq m$ and write $h = 1/Q$, where Q is any number such that

$$Q \geq \sup_{\bar{n}\in S} [\lambda + \Sigma_i \Phi_i(n_i)] \tag{3.9}$$

Then by (2.7) (also see (2.15)), (2.8), (3.3) and (3.7) we have

$\Delta_i V_{m+1}(\bar{n}) =$

$\Big\{[\Phi_i(n_i+1)-\Phi_i(n_i)]h\ p_{i0}\ +$

$\Sigma_{j=1}^{N}\ \Phi_j(n_j)h\ p_{j0}\ +$

$\lambda h\ 1_{\{i\neq 1,n_1<B\}}\ V_m(\bar{n}+e_i+e_1)\ +$

$\lambda h\ 1_{\{i\neq 1,n_1=B\}}\ V_m(\bar{n}+e_i+e_N)\ +$

$\lambda h\ 1_{\{i=1,n_1+1<B\}}\ V_m(\bar{n}+e_1+e_1)\ +$

$\lambda h\ 1_{\{i=1,n_1+1=B\}}\ V_m(\bar{n}+e_1+e_N)\ +$

$[\Phi_i(n_i+1)\ -\ \Phi_i(n_i)]h\ p_{i0}\ V_m(\bar{n})\ +$

$[\Phi_i(n_i+1)\ -\ \Phi_i(n_i)]h\ \Sigma_{k=1}^{N}\ p_{ik}\ V_t(\bar{n}+e_k)\ +$

$\Sigma_{j=1}^{N}\ \Phi_j(n_j)h\ \Sigma_{k=0}^{N-1}\ p_{jk}\ V_m(\bar{n}+e_i-e_j+e_k)\ +$

$[1\ -\ \lambda h\ -\ [\Phi_i(n_i+1)\ -\ \Phi_i(n_i)]h\ -\ \Sigma_{j=1}^{N}\ \Phi_j(n_j)h]\ V_m(\bar{n}+e_i)\Big\}$

$-$

$\Big\{\Sigma_{j=1}^{N}\ \Phi_j(n_j)h\ p_{j0}\ +$

$\lambda h\ 1_{\{i\neq 1,n_1<B\}}\ V_m(\bar{n}+e_1)\ +$

$\lambda h\ 1_{\{i\neq 1,n_1=B\}}\ V_m(\bar{n}+e_N)\ +$

$\lambda h\ 1_{\{i=1,n_1+1<B\}}\ V_m(\bar{n}+e_1)\ +$

$\lambda h\ 1^{\{i=1,n_1+1=B\}}\ V_m(\bar{n}+e_1)\ +$

$[\Phi_i(n_i+1)\ -\ \Phi_i(n_i)]h\ p_{i0}\ V_m(\bar{n})\ +$

$[\Phi_i(n_i+1)\ -\ \Phi_i(n_i)]h\ \Sigma_{k=1}^{N}\ p_{ik}\ V_m(\bar{n})\ +$

$\Sigma_{j=1}^{N}\ \Phi_j(n_j)h\ \Sigma_{k=0}^{N-1}\ p_{jk}\ V_m(\bar{n}-e_j+e_k)\ +$

$[1\ -\ \lambda h\ -\ [\Phi_i(n_i+1)\ -\ \Phi_i(n_i)]h\ -\ \Sigma_{j=1}^{N}\Phi_j(n_j)h]\ V_m(\bar{n})\Big\}$

$[\Phi_i(n_i+1) - \Phi_i(n_i)]h\ p_{i0}$ +

$\lambda h\ 1_{\{i\neq 1, n_1<B\}}\ \Delta_i V_m(\bar{n}+e_1)$ +

$\lambda h\ 1_{\{i\neq 1, n_1=B\}}\ \Delta_i V_m(\bar{n}+e_N)$ +

$\lambda h\ 1_{\{i=1, n_1+1<B\}}\ \Delta_1 V_m(\bar{n}+e_1)$ +

$\lambda h\ 1_{\{i=1, n_1+1=B\}}\ \Delta_N V_m(\bar{n}+e_1)$ +

$[\Phi_i(n_i+1) - \Phi_i(n_i)]h\ \Sigma_{k=1}^{N-1}\ p_{ik}\Delta_k V_m(\bar{n})$ +

$[\Phi_i(n_i+1) - \Phi_i(n_i)]h\ p_{i0}\ [V_m(\bar{n}) - V_m(\bar{n})]$ +

$\Sigma_{j=1}^{N}\ \Phi_j(n_j)h\ \Sigma_{k=0}^{N-1}\ p_{jk}\ \Delta_i V_m(\bar{n}-e_j+e_k)$ +

$$\{1 - \lambda h - [\Phi_i(n_i+1) - \Phi_i(n_i)]h - \Sigma_{j=1}^{N}\Phi_j(n_j)h\}\Delta_i V_m(\bar{n}) \quad (3.10)$$

The lower estimate $\Delta_i V_{m+1}(\bar{n}) \geq 0$ then directly follows from substituting the induction hypothesis $\Delta_j V_m(\bar{n}) \geq 0$ for $j = i$ and $j = N$, noting that the seventh term is equal to 0, that $\Phi_N(n_N+1) - \Phi_N(n_N) \geq 0$ by assumption (3.1) and that the last term is nonnegative as $h^{-1} = Q \geq [\lambda + \Sigma_j\Phi_j(n_j) + \Phi_i(n_i+1) - \Phi_i(n_i)]$ for $\bar{n}+e_i \in S$.

The upper estimate $\Delta_i V_{m+1}(\bar{n}) \leq 1$ is concluded similarly by substituting the hypothesis $\Delta_j V_m(\bar{n}) \leq 1$ and specifically noting again that the seventh term is equal to 0, so that the negative part: $-[\Phi_i(n_i+1)-\Phi_i(n_i)]h$ from the last term compensates the first additional positive term: $[\Phi_i(n_i+1)-\Phi_i(n_i)]h\ p_{i0}$. □

Lemma 3.1 directly guarantees condition (2.11) and thus proves inequality (2.10). Furthermore, by (3.3) and (3.8) condition (2.14) becomes:

$$|[(M-n)\lambda M^{-1} - \lambda]\ \Sigma_{j=1}^{N}\ \alpha_j(\bar{n})[V_t(\bar{n}+e_j - V_t(\bar{n})]| =$$

$$|n\lambda M^{-1}\ [1_{\{n_1<B\}}\ \Delta_1\ V_t(\bar{n}) + 1_{\{n_1=B\}}\ \Delta_N\ V_t(\bar{n})]| \leq$$

$$n\,\lambda\,M^{-1}. \tag{3.11}$$

The following choice in order to verify (2.13) thus seems natural

$$\beta(\bar{n}) = n. \tag{3.12}$$

Lemma 3.2 below will guarantee (2.13) for this choice.

Lemma 3.2 Let W be the sojourn time of a job in the open version and denote by $\bar{0}=(0,\ldots,0)$ the empty state. Then for all $t\geq0$:

$$\bar{T}_t\ \beta(\bar{0}) \leq T_t\ \beta(\bar{0}) \leq \lambda\ W \tag{3.13}$$

Proof First we will prove that for all $t\geq0$:

$$\bar{T}_t\ f(\bar{0}) \leq T_t\ f(\bar{0}) \tag{3.14}$$

for any f such that for all j and $\bar{n},\bar{n}+e_j\in S$:

$$\Delta_j\ f(\bar{n}) \geq 0. \tag{3.15}$$

To this end, from (2.6) and the fact that $\bar{S}\subset S$, we obtain similarly to (2.16) or by direct telescoping:

$$(\bar{T}_s-T_s)\ f(\bar{0}) = \Sigma_{t=0}^{s-1}\ \bar{T}_t(\bar{T}-T)T_{s-t-1}\ f(\bar{0}) \tag{3.16}$$

As per (2.17) and (3.11), however, without taking absolute values and by substituting $V_t(\bar{n}) = V$, we also obtain for arbitrary function V and any $\bar{n}\in\bar{S}$:

$$(\bar{T}-T)V(\bar{n}) = -n\lambda M^{-1}[1_{\{n_1<B\}}\ \Delta_1 V(\bar{n})+1_{\{n_1=B\}}\ \Delta_N\ V(\bar{n})] \tag{3.17}$$

As the operators $\bar{T}_t$ remain restricted to $\bar{S}$ while $\bar{T}_t\psi(.)\leq 0$ whenever $\psi(.)\leq 0$ (i.e., in componentwise sense), from (3.16) and (3.17) inequality

(3.14) is concluded provided (3.15) holds with f replaced by $T_s f$ for all possible s and f satisfying (3.15). This will be proven by induction to s. For s=0, it is satisfied by definition. Suppose that $T_s f$ satisfies (3.15) for all s≤m and f satisfying (3.15). Then similarly to (3.10) we obtain:

$$\begin{aligned}
&\Delta_i(T_{m+1}f)(\bar{n}) = \\
&T(T_m f)(\bar{n}+e_i) - T(T_m f)(\bar{n}) = \\
&\lambda h\ 1_{\{i\neq 1,\, n_1<B\}}\ \Delta_i(T_m f)(\bar{n}+e_1) + \\
&\lambda h\ 1_{\{i\neq 1,\, n_1=B\}}\ \Delta_i(T_m f)(\bar{n}+e_N) + \\
&\lambda h\ 1_{\{i=1,\, n_1+1<B\}}\ \Delta_1(T_m f)(\bar{n}+e_1) + \\
&\lambda h\ 1_{\{i=1,\, n_1+1=B\}}\ \Delta_N(T_m f)(\bar{n}+e_1) + \\
&[\Phi_i(n_i+1)-\Phi_i(n_i)]h\ \Sigma_{k=1}^{N-1}\ \Delta_k\ (T_m f)(\bar{n}) + \\
&\Sigma_{j=1}^{N}\ \Phi_j(n_j)h\ \Sigma_{k=0}^{N-1}\ p_{jk}\ \Delta_i(T_m f)(\bar{n}-e_j+e_k) + \\
&\{1 - \lambda h - [\Phi_i(n_i+1) - \Phi_i(n_i)]h - \Sigma_{j=1}^{N}\ \Phi_j(n_j)h\}\ \Delta_i\ (T_m f)(\bar{n})
\end{aligned} \tag{3.18}$$

The induction hypotheses $\Delta_j(T_m f) \geq 0$ for all j now yield as in the proof of lemma 3.1: $\Delta_i(T_{m+1}f) \geq 0$. With i chosen arbitrarily, (3.14) is thus proven with f replaced by $(T_{m+1}f)$. As argued above this guarantees (3.14) for any f satisfying (3.15). As the function $\beta(\bar{n}) = n$ satisfies (3.15) the first inequality of (3.13) is concluded.

To conclude the second inequality of (3.13), we will now inductively prove that for f satisfying (3.15):

$$T_{t+1}f(\bar{0}) \geq T_t\ f(\bar{0}) \tag{3.19}$$

where it is noted that we are considering the open model. Then for t=0 we have with $h=Q^{-1}$:

$$T\, f(\bar{0}) = \lambda h\, f(\bar{0}+e_i) + [1-\lambda h]\, f(\bar{0}) \geq f(\bar{0}) \tag{3.20}$$

Suppose that (3.19) holds for t≤m and any f satisfying (3.15). Now recall that, as proven above, (3.15) holds with f replaced by $Tf=T_1f$ for any f satisfying (3.15). Inequality (3.19) for t=m+1 then follows by the induction hypothesis and

$$(T_{m+2}f - T_{m+1}f)(\bar{0}) = (T_{m+1} - T_m)(Tf)(\bar{0}) \geq 0 \tag{3.21}$$

Finally, with L the mean number of jobs in the open system, we conclude from (3.14), (3.19) and Little's result:

$$\bar{T}_t\ \beta(\bar{0}) \leq T_t\ \beta(\bar{0}) \leq \lim_{t\to\infty} T_t\ \beta(\bar{0}) = L = \lambda W \tag{3.22}$$

□

Combining theorem 2.1, lemma 3.1, inequality (3.11) and lemma 3.2 yields:

Theorem 3.3

$$\lambda \geq \bar{\lambda} \tag{3.23}$$

and with W the sojourn time of the open system:

$$\lambda - \bar{\lambda} \leq \lambda W\, M^{-1} \tag{3.24}$$

Example 3.4: Assembly line Let the stations 1,...,N be infinite server stations with service parameters μ_i at station i and assume a cyclic routing, i.e. $p_{i,i+1} = 1$ for i = 1,...,N where N+1 = 0, representing an assembly line. Then

$$0 \leq \lambda-\bar{\lambda} \leq \lambda M^{-1} \max\{\Sigma_{i=1}^{N-1}\ \mu_i^{-1},\ \mu_N^{-1}\} \tag{3.25}$$

Remark 3.5 (Other verification of (2.13)) As mentioned earlier, we could have allowed more finite capacity constraints such as in the assembly line example 3.4 for any station 1,...,N. The proof of lemma 3.1 would then become more complicated such as similarly to [22]. In that case, however, the number of jobs in the primary system is bounded say by some number F. By monotonicity arguments one can then prove

$$L \leq F + L_N \leq F + \Sigma_{k=1}^{\infty} \; k\lambda^k \left[\prod_{j=1}^{k} \mu_N(j)\right]^{-1} \left(\Sigma_{k=0}^{\infty} \; \lambda^k \left[\prod_{j=1}^{k} \mu_N(j)\right]^{-1}\right)^{-1} \quad (3.26)$$

as based upon estimating the queue length L_N of the overflow station by sending any job to this station. As per (3.22), the estimate λW in (3.12) could then be replaced by this estimate.

Evaluation The computation of the throughput of closed queueing networks with a large number of jobs is computationally most expensive. Especially for non-product form networks, as natural consequence of finite capacity constraints, no general efficient computational procedures are available. A robust but simple and general throughput estimate, therefore, might be of interest to practitioners. By comparing the closed system with an open analog such an estimate is proposed. Conditions are provided guaranteeing an error bound and order for the accuracy of this estimate. In concrete situations these conditions can be verified analytically based upon an inductive Markov reward proof technique. Explicit error bounds of order M^{-1} with M the number of jobs can so be obtained for concrete finite capacity networks. Extensions such as to multi-class and mixed open and closed system seem possible.

REFERENCES

[1] Adan, I.J.B.F. and Wal, J. van der, (1987), Monotonicity of the throughput of a closed queueing network in the number of jobs, Memorandum COSOR 87-03, Eindhoven University of Technology, Dep. of Math. and Comput. Sci., To appear in Opns. Res.

[2] Barbour, A.D. (1982), Generalized semi-Markov schemes and open queueing networks, *J. Appl. Prob.* 19, 469-474.

[3] Hordijk, A. and Schassberger, R. (1982), Weak coonvergence of generalized semi-Markov processes, *Stochastic Process. & Appl.* 2, 271-291.

[4] Hordijk, A. and Van Dijk, N.M. (1983), Networks of queues, *Lecture Notes in Control and Information Sciences*, Vol. 60, Springer, Berlin, 158-205.

[5] Jaiswal, N.K. (1968), *Priority Queues*, Academic Press, New York.

[6] Kemeny, J.G., Snell, J.L. and Knapp, A.W. (1966), *Denumerable Markov Chains*, Van Nostrand, Princeton NJ.

[7] Kerola, T. (1986), The composite bound method for computing throughput bounds in multiple class environments, *Performance Evaluation* 6, 1-9.

[8] Muntz, R.R. and Wong, J.W., Asymptotic Properties of closed queueing network models, Proc. 8th Princeton Conf. on Information Sciences and Systems. Princeton Univ. Press, Princeton NJ, 348-352.

[9] Pinsky, E. and Yemini, Y. (1984), A statistical mechanics of some interconnection networks, Performance '84, *North-Holland.*

[10] Reiser, M. and Lavenberg, S.S. (1980), Mean value analysis of closed multichain queueing networks, *JACM* 27, (2): 313-322.

[11] Schwartz, M. (1987), Telecommunication networks, *Addison Wesley.*

[12] Seneta, E. (1980), Non-negative matrices and Markov chains, Springer-Verlag.

[13] Shanthikumar, J.G. and Yao, D. (1987), Stochastic monotonicity of the queue lengths in closed queueing networks, *Oper. Res.* 35, 583-588.

[14] Stoyan, D. (1983), Comparison methods for queues and other stochastic models, Wiley, New York.

[15] Tobagi, F.A. and Shur, D.H. (1986), Performance evaluation of channel access schemes in multihop packet radio networks with regular structure by simulation, *Computer Networks and ISDN systems*, Vol.12, No.1, 39-60.

[16] Tijms, H.C. (1986), Stochastic Modelling and Analysis; A computational approach, *Wiley*, New York.

[17] Van Doorn, E.A. (1984), On the overflow process from a finite Markovian queue, *Performance Evaluation* 4, 233-240.

[18] Van Dijk, N.M. (1988), Simple bounds for queueing systems with breakdowns, *Performance Evaluation*, 8, 117-128.

[19] Van Dijk, N.M., (1988), On the finite horizon Bellman equation for controlled Markov jump models with unbounded characteristics: Existence and approximation, *Stochastic Procc. Appl.* 27, 261-277.

[20] Van Dijk, N.M., A formal proof for the insensitivity of simple bounds for finite multiserver non-exponential tandem queues based on monotonicity results, *Stochastic Procc. Appl.* 27, 261-277.

[21] Van Dijk, N.M. (1987), Simple and insensitive bounds for a grading and an overflow model, *Oper. Res. Letters* 6, 73-76.

[22] Van Dijk, N.M. and Van der Wal, J. (1987), Simple bounds and monotonicity results for multi-server exponential tandem queues, To appear: *Queueing Systems.*

[23] Whitt, W. (1984), Open and closed models for networks of queues, *AT&T Bell Laboratories Technical Journal* 63, 1911-1979.

[24] Yao, D.D. and Buzacott (1987), Modeling a class of flexible manufacturing systems with reversible routing. *Oper. Res.*, 35, 87-93.

[25] Zahorjan, J., Sevcik, K.C., Eager, D.L. and Galler, B., (1982), Balanced job bound analysis of queueing networks, Comm. ACM 25(2) 134-141.