SERIE RESEARCH MEMORANDA

ANALYTIC ERROR BOUNDS FOR APPROXIMATIONS
OF QUEUEING NETWORKS WITH AN APPLICATION
TO ALTERNATE ROUTING

Nico M. van Dijk

Research Memorandum 1989-6
January 1989

VRIJE UNIVERSITEIT
FACULTEIT DER ECONOMISCHE WETENSCHAPPEN
EN ECONOMETRIE
AMSTERDAM

ANALYTIC ERROR BOUNDS FOR APPROXIMATIONS OF QUEUEING NETWORKS WITH AN APPLICATION TO ALTERNATE ROUTING.

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

Abstract A general condition is provided from which an error bound can be concluded for approximations of queueing networks which are based on modifications of the transition and state space structure. This condition relies upon Markov reward theory and can be verified inductively in concrete situations. The results are illustrated by estimating the accuracy of a simple throughput bound for a closed queueing network with alternate routing and a large finite source input. An explicit error bound for this example is derived, of order M^{-1}, where M is the number of sources.

Keywords Queueing network * throughput * finite source input * alternate routing * approximation * error bound.

1 Introduction

Ever since Erlang's and Engset's classical results in the early twenties, queueing theory has been extensively involved in teletraffic and communication theory Particularly, motivated, by Jackson's celebrated product form results in the late fifties (cf. [10]), the theory of queueing networks has gained a wide popularity in telecommunication and computer performance evaluation. Part of this success can be attributed to the various product form extensions and their robustness with respect to the underlying distributional assumptions (insensitivity properties), (e.g. [2-9], [12], [14], [18], [29], [32]. Most unfortunately, typical practical features such as blocking phenomena, dynamic routing, overflow, breakdowns and job-priorities usually destroy the appealing product form from an analytical point of view, (e.g. [8], [19], [23], [33]).

Another part of their success, however, can be explained by the fact that simplifying assumptions (such as infinite and independent stations) which guarantee product forms, tend to give "reasonable" approximate results in various practical situations, especially when the system is large, (cf. [19], [33]). These simplications can often be seen as minor though critical modifications of the underlying transition structure such as by adding or deleting particular transitions. Despite numerical support, however, analytic a priori error bounds for the accuracy of such "product form" approximations do not seem to be available.

Also other types of approximate modeling issues are typically concerned with networks of queues. One of these is the issue of a closed (finite source input) or open (Poissonian input) description (cf.[31]), with advantages (e.g computational, finiteness) and disadvantages (e.g. complexity, station dependence) for either of them. Convergence results for closed approximations of open systems have been established (cf. [20],[31]). But (error) bounds of this form are limited to simple Erlang type systems (cf [31]) or robust bounds for state space truncations which do not secure an order of accuracy (cf. [20]).

Another approximate or modeling issue is the exactness of system input parameters such as the mean arrival and service rates, as in practice these are usually subject to randomness (e.g. resulting from confidence intervals for statistical estimates or external fluctuations). To this end, perturbation results with error bounds have recently been developed in [27], with one dimensional queueing applications.

All of the above "approximations" come down to some kind of modification or perturbation of the transition structure and/or a truncation or an extension of the state space. This paper, therefore, aims to provide a general tool for concluding error bounds for such approximations. It thereby extends the perturbation error bound results from [27] in that it

(i) allows modifications of the state space such as a truncation for closed or an infinite extension for open modeling and
(ii) particularizes to networks of queues rather than one-dimensional queueing applications.

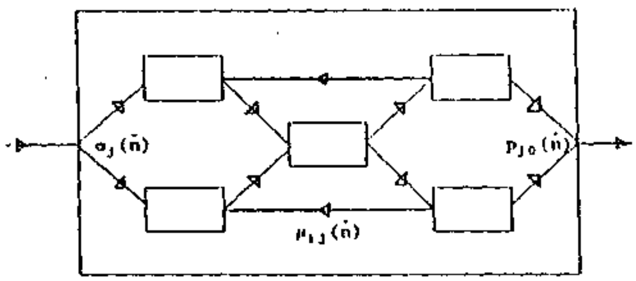
A pair of simple conditions is provided from which error bounds can be concluded. The actual verification of these conditions, however, is the crucial part for practical application To this end, an inductive verification technique based on Markov reward equations will be presented. This technique has already proven to be succesful in somewhat related situations (cf. [24] [26], [28]), but cannot be guaranteed in generality as complex technicalities are involved. The main part of this paper, therefore, is concerned with illustrating how the necessary conditions can be verified for a particular non-product form system of practical interest (cf. [1], [15]). This concerns a queueing network with alternate routing upon saturation of a primary access station and a large finite source input. A simple throughput is proposed and an explicit error bound is derived of order M^{-1}, with M the number of sources.

Though the example is to be seen as mainly generic as it is relatively simple from a practical point of view, it includes the essential phenomenon of a finite capacity constraint (or blocking) and a state dependent (thus dynamic) routing. For more complex situations, such as with more capacity constraints and blocking phenomena, similar though more complicated proofs can be expected along essentially the same lines.

First, the general models are presented in section 2. Next, the corresponding error bound result is developed in section 3. Finally, an illustrative network with alternate routing is analyzed in section 4. An evaluation concludes the paper.

2 Comparative Models

Consider an arbitrary open or closed single class exponential queueing network with N service stations (hereafter called the original model), such as illustrated below.



The state of the network is described by $\bar{n} = (n_1, \ldots, n_N)$ where n_i is the number of jobs at station i, i=1,...,N. By $\bar{n}+e_i$ or $\bar{n}-e_i$ we denote the state equal to $\bar{n}$ except for one job more respectively less at station i, where $\bar{n}-e_i=\bar{n}$ for $n_i=0$, i=1,...,N and where we also allow i=0 with the convention that $\bar{n}\pm e_0=\bar{n}$. Consequently, by $\bar{n}-e_i+e_j$ we denote the state equal to $\bar{n}$ with one job moved from station i into station j, where i=0 corresponds to an external arrival at station j and j=0 to a departure from the system at station i. Let $q(\bar{n}, \bar{n}-e_i+e_j)$ for i,j = 0,1,...,N be the transition rate for a change from state $\bar{n}$ into state $\bar{n}-e_i+e_j$, while transition rates for changes not of this form are assumed to be 0. For example, for a standard Jackson network we have

$$q(\bar{n}, \bar{n}-e_i+e_j) = \mu_i \; p_{ij}$$

with μ_i the service rate at station i and p_{ij} the routing probability from station i to j, while an additional capacity constraint N_j yielding a reflective blocking (communication protocol) is parametrized by

$$q(\bar{n}, \bar{n}-e_i+e_j) = \mu_i \ p_{ij} \ 1(n_j<N_j)$$

where 1(A) or $1_{\{A\}}$ denotes an indicator of event A, i.e. $1(A)=1_{\{A\}}=1$ if event A is satisfied and $1(A)=1_{\{A\}}=0$ otherwise. Without restriction of generality, the following assumptions are made:

1. The underlying Markov jump process is irreducible at some set S of admissible states $\bar{n}$, with a unique stationary distribution $\pi(.)$.
2. The transition rates are uniformly bounded. That is, we can choose a finite Q such that

$$Q \geq \sup_{\bar{n}\in S} \Sigma_{i,j} \ q(\bar{n}, \bar{n}-e_i+e_j) \tag{2.1}$$

3. For some given reward rate $r(\bar{n})$ the value g is finite and well-defined by

$$g = \Sigma_{\bar{n}} \ \pi(\bar{n}) \ r(\bar{n}) \tag{2.2}$$

The value g then represents some performance measure of interest, such as the throughput of a particular station j by

$$r(\bar{n}) = \mu_j(n_j)$$

or the steady state probability of a particular subset B by

$$r(\bar{n}) = 1(\bar{n}\in B)$$

Comparative model. Now consider a modified version of the single class exponential queueing network (hereafter called the modified model) with a description as above, but with $q(\bar{n},\bar{n}-e_i+e_j)$ replaced by $\bar{q}(\bar{n},\bar{n}-e_i+e_j)$, the assumptions 1, 2 and 3 adopted with S, π, r and g replaced by $\bar{S}$, $\bar{\pi}$, $\bar{r}$ and $\bar{g}$, but Q kept the same, and most essentially

$$\bar{S} \subset S \tag{2.3}$$

3 Comparison Result

We now wish to evaluate the difference $|\bar{g}-g|$, that is the difference of the performance measure for he original and modified queueing network, without having to compute the stationalry probabilities $\pi(.)$ and $\bar{\pi}(.)$.

To this end, as justified by the boundedness assumption (2.2), we first apply the standard uniformization technique (e.g. [23], p. 110) in order to transform the continuous-time description in a discrete-time formulation. More precisely, let Q be any arbitrary finite number satisfying (2.2) and define one-step transition probabilities $p(\bar{n},\bar{n}-e_i+e_j)$ and $\bar{p}(\bar{n},\bar{n}-e_i+e_j)$ by

$$
\begin{aligned}
p(\bar{n}, \bar{n}-e_i+e_j) &= q(\bar{n}, \bar{n}-e_i+e_j)/Q \\
\bar{p}(\bar{n}, \bar{n}-e_i+e_j) &= \bar{q}(\bar{n}, \bar{n}-e_i+e_j)/Q \\
p(\bar{n}, \bar{n}) &= 1-[\Sigma_{i,j=0}^{N} q(\bar{n}, \bar{n}-e_i+e_j)/Q] \\
\bar{p}(\bar{n}, \bar{n}) &= 1-[\Sigma_{i,j=0}^{N} \bar{q}(\bar{n}, \bar{n}-e_i+e_j)/Q]
\end{aligned}
\tag{3.1}
$$

while transition probabilities $p(.,.)$ and $\bar{p}(.,.)$ for any other type transition are assumed to be 0. From now on, we always use an upper bar "-" symbol to indicate an expression for the modified system and the symbol "(-)" to indicate that the expression is to be read for both the original and modified system. Further, let operators $^{(\bar{\ })}T$ and $\{^{(\bar{\ })}T_t \mid t=0,1,2,\ldots\}$ upon arbitrary functions $f: {}^{(\bar{\ })}S \to R$ be defined by

$$
\begin{aligned}
{}^{(\bar{\ })}T f(\bar{n}) &= \Sigma_{i,j=0}^{N} {}^{(\bar{\ })}p(\bar{n},\bar{n}-e_i+e_j)\ f(\bar{n}-e_i+e_j) \\
{}^{(\bar{\ })}T_{t+1} f &= {}^{(\bar{\ })}T\, {}^{(\bar{\ })}T_t f \quad (t \ge 0), \\
{}^{(\bar{\ })}T_0 &= I
\end{aligned}
\tag{3.2}
$$

And define the reward functions $\{\bar{V}_t \mid t=0,1,2,\ldots\}$ at $\bar{S}$ by

$$\bar{V}_N = Q^{-1} \sum_{t=0}^{N-1} \bar{T}_t \, r \tag{3.3}$$

Then by virtue of the uniformization technique (cf. [23], p. 110) and the irreducibility assumptions of $\bar{S}$, by standard Markov reward theory (cf. [16]) we conclude

$$\bar{g} = \lim_{N\to\infty} \frac{Q}{N} \bar{V}_N \, (\bar{\ell}) \tag{3.4}$$

for arbitrary $\bar{\ell} \in \bar{S}$. This leads to the following key-theorem which guarantees an error bound for the difference $|\bar{g}-g|$. Its conditions will be discussed later on. Herein, we use the abbreviation

$$\Delta(\bar{n}, \bar{n}-e_i+e_j) = [\bar{q}(\bar{n}, \bar{n}-e_i+e_j) - q(\bar{n}, \bar{n}-e_i+e_j)]$$

Theorem 3.1 (General Conditions) Suppose that for some constants β, δ, $\varepsilon > 0$, some state $\bar{\ell}\in\bar{S}$, some nonnegative function $\Phi(.)$ and all $t\geq 0$, $\bar{n}\in\bar{S}$:

$$\bar{T}_t \Phi(\bar{\ell}) \leq \beta \tag{3.5}$$

$$|\bar{r}(\bar{n})-r(\bar{n})| \leq \delta \; \Phi(\bar{n}) \tag{3.6}$$

$$|\Sigma_{i,j=0}^{N} \Delta(\bar{n},\bar{n}-e_i+e_j) \; [V_t(\bar{n}-e_i+e_j)-V_t(\bar{n})]| \leq \varepsilon \;\; \Phi(\bar{n}) \tag{3.7}$$

Then

$$|(\bar{V}_N-V_N)(\bar{\ell})| \leq \beta[\delta+\varepsilon] \; N/Q \tag{3.8}$$

and

$$|\bar{g}-g| \leq \beta[\delta+\varepsilon] \tag{3.9}$$

Proof Clearly, (3.9) immediately follows from (3.4) and (3.8). To prove (3.8) first conclude from (3.2) and (3.3) that for any $t \geq 0$:

$$^{(\bar{}}\bar{V}_{t+1}^{)} = {}^{(\bar{}}\bar{r}^{)}/Q + {}^{(\bar{}}\bar{T}_t^{)}\,{}^{(\bar{}}\bar{V}_t^{)} \tag{3.10}$$

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

$$(\bar{V}_N - V_N)(\bar{n}) = (\bar{r}-r)(\bar{n})/Q + (\bar{T}\bar{V}_{N-1} - TV_{N-1})(\bar{n})$$

$$= (\bar{r}-r)(\bar{n})/Q + (\bar{T}-T)V_{N-1}(\bar{n}) + \bar{T}(\bar{V}_{N-1} - V_{N-1})(\bar{n})$$

$$= \Sigma_{t=0}^{N-1} \bar{T}_t([\bar{r}-r]/Q + [(\bar{T}-T)V_{N-t-1}])(\bar{n}) + \bar{T}_N(\bar{V}_0 - V_0)(\bar{n}), \tag{3.11}$$

where the latter equality follows by iteration. Now note that the last term in the last right hand side is equal to 0 as $\bar{V}_0(.) - V_0(.) = 0$ by definition. Further, from (3.1) and (3.2) we find for any s and $\bar{n} \in \bar{S}$:

$$(\bar{T}-T)V_s(\bar{n}) = \Sigma_{i,j=0}^{N} \bar{q}(\bar{n},\bar{n}-e_i+e_j)[V_s(\bar{n}-e_i+e_j) - V_s(\bar{n})]/Q$$

$$- \Sigma_{i,j=0}^{N} q(\bar{n},\bar{n}-e_i+e_j)[V_s(\bar{n}-e_i+e_j) - V_s(\bar{n})]/Q$$

$$= \Sigma_{i,j=0}^{N} \Delta(\bar{n},\bar{n}-e_i+e_j)[V_s(\bar{n}-e_i+e_j) - V_s(\bar{n})]/Q \tag{3.12}$$

Further, note that $\bar{T}_t f_1 \leq \bar{T}_t f_2$ for any $f_1 \leq f_2$ in component-wise sense as $\bar{T}_t$ is an expectation operation. As a result, by substituting (3.12) in (3.11), substituting $\bar{n} = \bar{\ell}$, taking absolute values and applying (3.5) (3.7), we obtain:

$$|(\bar{V}_N - V_N)(\bar{\ell})| \leq [\delta+\varepsilon]Q^{-1}\, \Sigma_{t=0}^{N-1} \bar{T}_t \Phi(\bar{\ell}) \leq \beta[\delta+\varepsilon]N/Q. \tag{3.13}$$

Remark 3.2 (Discussion of the theorem). In the above theorem one must typically think of β and/or $[\delta+\varepsilon]$ to be small. To this end, several steps are involved, as will be discussed below.

Step 1 (Bounded bias-terms) As first and most essential step one has to estimate (bound) the so-called bias terms $V_t(\bar{n}-e_i+e_j) - V_t(\bar{n})$ as:

$$|V_t(\bar{n}-e_i+e_j)-V_t(\bar{n})| \leq B_{i,j} \qquad (3.14)$$

uniformly in t. From Markov reward theory it is standardly known that such terms are bounded uniformly in t for any given i and j as based upon mean first passage time results (cf. [16], [27]) and assuming r(.) to be bounded. For finite networks a bound B uniformly in t and i,j can thus be concluded. The actual computation of such bounds by means of mean first passage times, however, becomes practically impossible for multi-dimensional applications such as considered in this paper (See [11] or [27] for simple one-dimensional cases). In the next section, therefore, we will illustrate how estimates for these bias-terms can be derived analytically.

Step 2 (Transition differences) Secondly, one has to find out whether the differences in the transition structure $\Delta(.,.)$ are small or just bounded up to a state dependent scaling function $\Phi(.)$. For illustration, think of $\Phi(.)=1$ and consider the following examples.

Example 1 Consider a standard single-server queue with arrival rate λ and service rate μ as original model and the same model with arrival rate $\lambda \pm \tau$ (perturbation),resulting from a statistical confidence interval, as modified model, where τ is small. Then

$$|\Delta(.,.)| \leq \tau$$

Example 2 Consider the same original model as in the example above but now with rejection of arrivals (state space truncation) if upon arrival the number n of jobs present is equal to some limit L. Then

$$|\Delta(.,.)| \leq \lambda 1(n=L)$$

Step 3 (Bounding function Φ) By comparing the transition structures, candidates for an appropriate bounding function function $\Phi(.)$ come up naturally. Here one may typically think of polynomial type functions, for example, $\Phi(\bar{n}) = n$ with n the total number of jobs present. One may thus have various options. As illustration, in example 2 above, condition (3.7) will be satisfied with some constant B resulting from (3.14) and

$$\begin{cases} \varepsilon = \lambda B/L \text{ if } & \Phi(.) = n \\ \varepsilon = \lambda B \quad \text{if} & \Phi(.) = 1(n{=}L) \end{cases}$$

Step 4 (Stability) Which option of $\Phi(.)$ is appropriate will eventually depend on whether we can easily verify (3.5), requiring that its expected value over time remains bounded (stability) by either a small or just a finite number. As illustration, again for example 2 from above, we have

$$\begin{cases} \beta = \beta_1 = 1 & \text{if } \Phi(.) = 1 \\ \beta = \beta_2 = (\lambda/\mu)^L / \Sigma_{k=0}^{L} (\lambda/\mu)^k & \text{if } \Phi(.) = 1(n{=}L) \\ \beta = \beta_3 = \Sigma_{k=0}^{L} k(\lambda/\mu)^k \ / \ \Sigma_{k=0}^{L} (\lambda/\mu)^k & \text{if } \Phi(.) = n \end{cases}$$

Roughly speaking, theorem 3.1 can thus be applicable in a twofold manner given that the bias-terms can be sufficiently estimated:

(i) By showing that the impact of the difference Δ in the transition structures upon the state-dependent estimates for the bias terms is sufficiently small, such as for example 1 with $\varepsilon=\tau B$ and $\beta=1$ by using $\Phi(.)=1$, or example 2 with $\varepsilon=\tau B/L$ and $\beta=\beta_3$ by using $\Phi(.)=n$.

(ii) By showing that the expected value of the scaling function or the probability of being in states where this difference is significant, is sufficiently small, such as for example 2 with $\varepsilon=\tau B$ and $\beta=\beta_2$ by using $\Phi(.)=1(n=L)$.

Remark 3.3 (Unbounded rewards) Note that no assumption has been made as to any boundedness of the reward rate. For example, we can have $r(\bar{n})=n_j$ so as to calculate the mean queue length at a particular infinite station j.

Remark 3.4 (Unbounded intensities) The boundedness assumption (2.1) is made in order to apply the uniformatization technique (3.1) yielding a recursive formulation. This, however, can be avoided in a technical manner similarly to [25] so as to allow unbounded intensities, such as from infinite server stations.

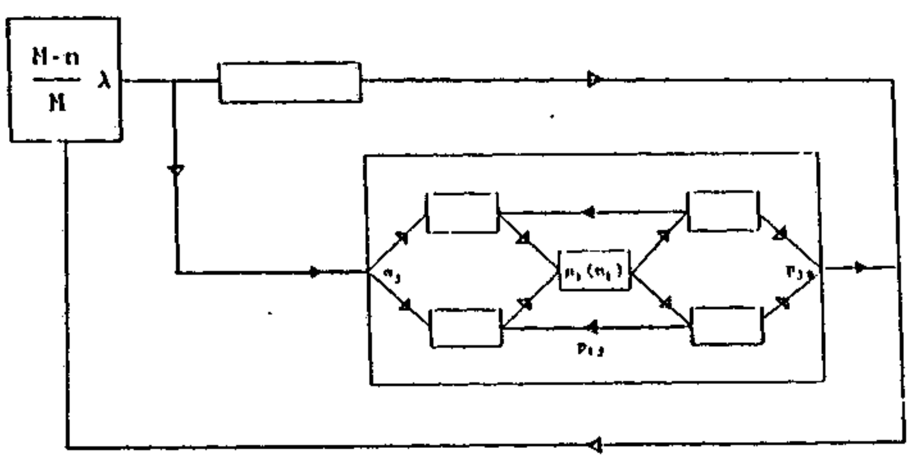
4 Application: A simple throughput and explicit error bound for a closed queueing network with alternate routing

4.1 Model

This section investigates an application of the preceeding results which contains both a perturbation of the transition structure and a state space truncation. Moreover, the performance measure of interest, the system throughput, involves an unbounded reward structure. The application concerns a generic example of a practical phenomenon in teletraffic analysis: blocking with alternate routing.

Consider a queueing network with N service stations, as illustrated below, of which the first station is a primary entrance station which allows no more than some finite number L of jobs and where upon saturation of this station jobs have to take an alternate route according to routing probabilities p_{ij}, $i,j=0,2,\ldots,N$, starting at station j with probability $p_{0j}=\alpha_j$. Upon service completion at station i a job leaves the system with probability $p_{i0}=1-\Sigma_{j=2}^{N} p_{ij}$ for $i\neq 1$ and $p_{10}=1$, where the transition matrix (p_{ij}) for $i,j=0,2,\ldots,N$ is assumed to be irreducible.

The service rate at station i is $\mu_i(n_i)$ when n_i jobs are present where $\mu_i(n_i)$ is assumed to be nondecreasing in n_i. Jobs arrive at the system according to a finite source exponential input with M sources and exponential idle times with parameter γ. That is, if n jobs are present in the system the arrival rate is $(M-n)\gamma$.



The system under consideration is not of product-form due to the dynamic routing feature upon saturation. This feature naturally arises in teletraffic applications for which various alternate routing schemes are of actual interest (cf. [1],[15]). Here a large finite source input is most realistic, so that often a Poissonian input approximation is used to simplify analysis or avoid complex computations of a performance measure such as the throughput. Below we will investigate the accuracy of such an approximation, or more precisely, of the throughput bound (it can be shown to be indeed an upper bound):

$$\lambda = \gamma M \tag{4.1}$$

4.2 Parametrization

As we require $\hat{S} \subset S$, we consider the open Poisson input case as the original model and the closed finite source case as the modified model. Then with

$$Q \geq \lambda + \Sigma_i \mu_i(M), \tag{4.2}$$

and choosing

$$\mu_i(n_i) = \mu_i(M)$$

for $n_i \geq M$, as well as

$$S = \{\bar{n} | n_1 \leq L\}$$

$$\hat{S} = \{\bar{n} | n_1 \leq L,\ n_1 + \ldots + n_N = M\}, \tag{4.3}$$

the assumptions 1, 2 and 3 of section 2 are guaranteed for both the closed and open version witl respective transition rates $\bar{q}(.,.)$ and $q(.,.)$ given by

$$\stackrel{(-)}{q}(\bar{n}, \bar{n}-e_i) \quad = \mu_.(n_i) p_{i0} \qquad (i=1,\ldots,N)$$

$$\stackrel{(-)}{q}(\bar{n}, \bar{n}-e_i+e_j) = \mu_.(n_i) p_{ij} \qquad (i,j=2,\ldots,N)$$

but

$$\begin{cases} q(\bar{n},\bar{n}+e_1) & = \lambda\ 1(n_1<L) \\ q(\bar{n},\bar{n}+e_j) & = \lambda\ 1(n_1=L)\ \alpha_j \qquad (j=2,\dots,N) \end{cases}$$

and

$$\begin{cases} \bar{q}(\bar{n},\bar{n}+e_1) & = (M-n)\gamma\ 1(n_1<L) \\ \bar{q}(\bar{n},\bar{n}+e_j) & = (M-n)\gamma\ 1(n_1=L)\ \alpha_j \qquad (j=2,\dots,N) \end{cases} \qquad (4.4)$$

The uniformization (3.1) is thus justified and with

$$^{(\bar{\ })}r(\bar{n}) = \Sigma_{i=1}^{N}\ \mu_i(\bar{n}_i)p_{i0} \qquad (\bar{n}\in S) \qquad (4.5)$$

the values $\bar{g}$ and g, as per (2.2) or equivalently (3.4), represent the throughput of the closed and open system. Since, however, $g=\lambda-\gamma M$ while $\bar{g}$ cannot be computed easily, it is of interest to investigate theorem 3.1 so as to estimate the difference $|\bar{g}-\lambda|$.

4.3 Comparison result

We adopt all notation from section 3. As per the discussion in remark 3.2, the following lemma is the most crucial step. Herein, for arbitrary functions $f:S\to R$ and $j=1,\dots,N$ we use the notation:

$$\Delta_j f(\bar{n}) = f(\bar{n}+e_j)-f(\bar{n}) \qquad (4.6)$$

Lemma 4.1 For all $t\geq 0$ and j and $\bar{n}$ such that $\bar{n}+e_j\in S$:

$$0 \leq \Delta_j V_t(\bar{n}) \leq 1 \qquad (4.7)$$

Proof This will be given by induction to t. For t=0, (4.7) trivially holds as $V_0(.)=0$. Suppose that (4.7) holds for $t\leq m$ and for convenience write

$h = Q^{-1}$. In advance it is noted that in the derivation below terms are added artificially and splitted for appropriate comparison of corresponding terms. Also, some terms that are actually equal to 0 will be written out for clarity. Further, we note that this lemma concerns the open case with a Poisson intensity λ as according to (4.4) for the transition rates $q(.,.)$. Then from (3.10), (4.4), (4.5) and (4.6) we find for $t=m+1$ and $i=1,\ldots,N$:

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

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

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

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

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

$$\lambda h \; 1_{\{i\neq 1, n_1=L\}} \Sigma_{j=2}^{N} \; \alpha_j \; V_m(\bar{n}+e_i+e_j) \; +$$

$$\lambda h \; 1_{\{i=1, n_1+1=L\}} \Sigma_{j=2}^{N} \; \alpha_j \; V_m(\bar{n}+e_1+e_j) \; +$$

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

$$[\mu_i(n_i+1)-\mu_i(n_i)] \; h \; \Sigma_{k=2}^{N} \; p_{ik} \; V_m(\bar{n}+e_k) \; +$$

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

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

$$-$$

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

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

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

$$\lambda h\ 1_{\{i\neq 1, n_1=L\}}\ \Sigma_{j=2}^{N}\ \alpha_j V_m(\bar{n}+e_j)\ +$$

$$\lambda h\ 1_{\{i=1, n_1+1=L\}}\ \Sigma_{j=2}^{N}\ \alpha_j V_m(\bar{n}+e_1)\ +$$

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

$$[\mu_i(n_i+1)-\mu_i(n_i)]h\ \Sigma_{k=2}^{N}\ p_{ik}\ V_m(\bar{n})\ +$$

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

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

=

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

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

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

$$\lambda h\ 1_{\{i\neq 1, n_1=L\}}\ \Sigma_{j=2}^{N}\ \alpha_j \Delta_i V_m(\bar{n}+e_j)\ +$$

$$\lambda h\ 1_{\{i=1, n_1+1=L\}}\ \Sigma_{j=2}^{N}\ \alpha_j \Delta_j V_m(\bar{n}+e_1)\ +$$

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

$$[\mu_i(n_i+1)-\mu_i(n_i)]h\ \Sigma_{k=2}^{N} p_{ik} \Delta_k V_m(\bar{n})\ +$$

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

$$[1-\lambda h-[\mu_i(n_i+1)-\mu_i(n_i)h-\Sigma_{j=1}^{N}\ \mu_j(n_j)h]\ \Delta_i V_m(\bar{n}) \qquad (4.8)$$

The lower estimate $\Delta_i V_{m+1}(\bar{n})\geq 0$ now directly follows from substituting the induction hypothesis $\Delta_j V_m(.)\geq 0$ for all j, noting that the one but last term is equal to 0, recalling that the service rate $\mu_i(.)$ is nondecreasing by assumption and observing that the last term is nonnegative by virtue of $h=Q^{-1}$ satisfying (4.3).

The upper estimate $\Delta_i V_{m+1}(\bar{n})\leq 1$ is concluded similarly by substituting the induction hypothesis $\Delta_i V_m(\bar{n})\leq 1$ for all j, again noting that the one but last term is equal to 0, for which the first term $[\mu_i(n_i+1)-\mu_i(n_i)]h\ p_{i0}$ can be substituted, and observing that all coefficients together than sum up to 1.

We are now able to verify condition (3.8). To this end, recall that the transition structures q(.,.)and $\bar{q}$(.,.) as according to (4.4) differ only in their arrival rates. With (4.7) and $\gamma=\lambda M^{-1}$ as per (4.1), we then find

$$|\Sigma^{N}_{i,j=0}\ \Delta(\bar{n},\bar{n}-e_i+e_j)\ [V_t(\bar{n}-e_i+e_j)-V_t(\bar{n})]| =$$

$$|[(M-n)\lambda M^{-1}-\lambda]\ \{1_{\{n_1<L\}}[V_t(\bar{n}+e_1)-V_t(\bar{n})] +$$

$$1_{\{n_1=L\}}\ \Sigma^{N}_{j=2}\alpha_j[V_t(\bar{n}+e_j)-V_t(\bar{n})]\ \}| \leq$$

$$n\lambda/M \tag{4.9}$$

The following choice thus seems appropriate

$$\Phi(\bar{n}) = n \tag{4.10}$$

Lemma 3.2 below investigates whether (3.5) can then be verified.

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

$$\bar{T}_t\Phi(\bar{0}) \leq T_t\Phi(\hat{0}) \leq \lambda W \tag{4.11}$$

Proof First we will prove that for all $t\geq 0$:

$$\bar{T}_t f(\bar{0}) - T_t f(\bar{0}) \leq 0 \tag{4.12}$$

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

$$f(\bar{n}+e_j)-f(\bar{n})\geq 0 \qquad (j=1,\ldots,N) \tag{4.13}$$

To this end, from (3.10) and the fact that $\bar{S}\subset S$ we obtain similarly to (3.11) or by direct telescoping:

$$(\bar{T}_t - T_t)f(\bar{0}) = \Sigma_{s=0}^{t-1}\ \bar{T}_s(\bar{T}-T)T_{t-s-1}f(\bar{0}) \tag{4.14}$$

As per (3.12) and (4.9) however we also have for any $\bar{n}\in\bar{S}$ and function V

$$(\bar{T}-T)V(\bar{n}) = -[\lambda/M]\{1_{\{n_1<L\}}[V(\bar{n}+e_1)-V(\bar{n})] +$$

$$+ 1_{\{n_1=L\}}[\Sigma_{j=2}^{N}\alpha_j[V(\bar{n}+e_j)-V(\bar{n})]\} \tag{4.15}$$

Since the operators $\bar{T}_s$ remain restricted to $\bar{S}$ while $\bar{T}_s\psi \geq 0$ whenever $\psi \geq 0$ componentwise, from (4.14) and (4.15) inequality (4.12) is concluded, provided (4.13) holds with f replaced by $T_s f$ for any s, where f itself also satisfies (4.13).

This will be proven by induction to s. For s=0 it is satisfied by definition. Suppose that T_s satisfies (4.13) for $s\leq m$, then similarly to (4.8):

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

$$T(T_m f)(\bar{n}+e_i)-T(T_m f)(\bar{n}) =$$

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

$$\lambda h\ 1_{\{i=1, n_1+1<L\}}\Delta_{.}(T_m f)(\bar{n}+e_1) +$$

$$\lambda h\ 1_{\{i\neq 1,\, n_1 = L\}}\ \Sigma_{j=2}^{N}\ \alpha_j \Delta_i (T_m f)(\bar{n}+e_j)\ +$$

$$\lambda h\ 1_{\{i=1,\, n_1+1=L\}}\ \Sigma_{j=2}^{N}\ \alpha_j \Delta_j (T_m f)(\bar{n}+e_1)\ +$$

$$\Sigma_{j=1}^{N}\ \mu_j(n_j) h\ \Sigma_{k=0}^{N} p_{j\,k}\ \Delta_i (T_m f)(n^{-} - e_j + e_k)\ +$$

$$[\mu_i(n_i+1) - \mu_i(n_i)] h\ \Sigma_{k=2}^{N} p_{ik}\ \Delta_k (T_m f)(\bar{n})\ +$$

$$[1 - \lambda h - [\mu_i(n_i+1) - \mu_i(n_i)] h - \Sigma_{j=1}^{N} \mu_j(n_j) h]\ \Delta_i (T_m f)(\bar{n}) \qquad (4.16)$$

The induction hypotheses $\Delta_j(T_m f) \geq 0$ for all j, now yield as in the proof of lemma 4.1: $\Delta_i(T_{m+1} f) \geq 0$.

Inequality (4.12) is hereby proven and particularly, since $\Phi(\bar{n}) = n$ satisfies (4.13), also the first inequality of (4.11). To prove the second, we will now inductively prove that, again for f satisfying (4.13), for all $t \geq 0$:

$$T_t f(\bar{0}) \leq T_{t+1} f(\bar{0}) \qquad (4.17)$$

For t=0, we have with $h = Q^{-1}$:

$$Tf(\bar{0}) = \lambda h\ f(\bar{0}+e_1) + [1-\lambda h] f(\bar{0}) \geq f(\bar{0}) \qquad (4.18)$$

Assume that (4.17) holds for $t \leq m$ for any f satisfying (4.13). Then from this induction hypothesis and, as proven above, the fact that (4.13) also holds with f replaced by $Tf - T_1 f$ when f satisfies (4.13), inequality (4.17) is proven for t=m+1 by:

$$(T_{m+1} f - T_{m+2} f)(\bar{0}) = (T_m - T_{m+1})(Tf)(\bar{0}) \leq 0 \qquad (4.19)$$

With L the mean number of jobs in the open system, finally, we conclude from (4.12) and (4.17) with $f(\bar{n}) = \Phi(\bar{n}) = n$ and Little's result:

$$\bar{T}_t \Phi(\bar{0}) \leq T_t \Phi(\bar{0})\ \lim_{t\to\infty} T_t \Phi(\bar{0}) = L = \lambda W \qquad (4.20)$$

From $\bar{r}(.)=r(.)$ as per (4.5), lemma 4.1, inequality (4.9) and lemma 4.2., we now directly obtain by applying theorem 3.1:

Theorem 4.2 Throughput error bound With $\bar{\lambda}$ the throughput of the (γ,M)-finite source system, W the sojourn time of a job in the open version and $\lambda = \gamma M$:

$$|\bar{\lambda}-\lambda| \leq \lambda W/M \tag{4.21}$$

Example 4.3 (Deterministic alternate routing) Let all stations be infinite server stations with service parameters μ_i at station i, and assume that $\alpha_2=1$, $p_{i,i+1}=1$ for $i=2,\ldots,N-1$ and $p_{N,0}=1$, then

$$|\bar{\lambda}-\lambda| \leq \lambda M^{-1} \max\,(\mu_1^{-1},\ \Sigma_{i=2}^{N}\mu_i^{-1}). \tag{4.22}$$

Remark 4.4 ($\bar{\lambda}\leq\lambda$) By using the lower estimates $\Delta_j V_m \geq 0$ from lemma 4.1 in (3.11) and (4.9), rather than upper estimates after taking absolute values, as in (3.13) and noting that $(\bar{r}-r)(.)=0$, from (3.11), (3.12), (4.9) and (3.4) we can also conclude: $\bar{\lambda}\leq\lambda$. Intuitively, this may seem trivial. Counter-intuitively, however, as per counterexamples of related situations in [14] and [20], such monotonicity results will not generally hold.

Evaluation Approximations for queueing networks are often based on modifications of the original transition structure and/or the set of admissible states. An analytical tool is provided in order to estimate the accuracy of such approximations. Particularly, as scaling functions such as polynomials are allowed, the results do not require the modifications themselves to be small. The necessary conditions are generally verifiable by inductive Markov reward arguments. A typical application is an open approximation of a large closed system. Extensions to multi-class and non-exponential networks seem possible.

References

[1] Ash G.R., Kafker A.H., and Krishnan K.R., "Intercity Dynamic Routing Architecture and Feasibility", Proceedings of the 10th International Teletraffic Congress, Paper 3.2.2., Montreal, June.

[2] Barbour, A. (1976). "Networks of queues and the method of stages." *Advances in Applied Probability* 8: 584-591.

[3] Baskett, F., Chandy, M., Muntz, R., & Palacios, J. (1975). "open, closed and mixed networks of queues with different classes of customers." *J.A.C.M.* 22.

[4] Burman, D.Y., Lehoczky, J.P., & Lim, Y. (1982). "Insensititivy of blocking probabilities in an circuit switching network." *Journal of Applied Probability* 21: 850-859.

[5] Chandy, K.M., Howard, J.H., & Towsley, D.F. (1977). "Product form and local balance in queueing networks." *J.A.C.M.* 24: 250-263.

[6] Chandy, K.M. & Martin, A.J. (1983). "A characterization of product-form queueing networks." *J.A.C.M.* 30: 286-299.

[7] Hordijk, A. & van Dijk, N.M. (1983). "Networks of queues. Part I: Job-local-balance and the adjoint process. Part II: General routing and service characteristics." *Lecture Notes in Control and Informational Sciences*, Springer Verlag, 60: 158-205.

[8] Hordijk, A. & van Dijk, N.M. (1981). "Networks of queues with blocking." In K.J. Kylstra (ed.), *Performance '81*, North Holland, pp. 51-65.

[9] Hordijk, A. & van Dijk, N.M. (1983). "Adjoint process, job-local-balance and insensitivity of stochastic networks." *Bulletin of the 44th Session of the International Statistical Institute* 50: 776-788.

[10] Jackson, J.R. (1963). "Jobshop-like queueing systems." *Management Science* 10: 131-142.

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

[12] Kelly, F.P., (1979), "Reversibility and Stochastic Networks", *Wiley*.

[13] Kohlas, J., (1982), "Stochastic Methods of Operations Research", *Cambridge University Press*.

[14] Noetzel, A.S., "A Generalized Queueing Discipline for Product Form Network Solutions. *Journal of the ACM* 26(4): 779-793, October 1979.

[15] Ott, T.J. and Krishnan, K.R., (1985). "State dependent routing of telephone traffic and the use of separable routing schemes", Proceedings of the 11th International Teletraffic Congres, Tokyo, Japan.

[16] Ross, S.M. (1970). "Applied probability models with optimization application", *Holden-Day*, San Francisco.

[17] Rohlicek, J.R. and Willsky, A.S. (1988). "The reduction of perturbed Markov generators: an algorithm exposing the role of transient states", *J.A.C.M.* 35: 675-696.

[18] Schassberger, R. (1978). "Insensitivity of steady state distributions of eneralised semi-Markov processes with speeds", *Adv. Appl. Prob.*10: 836-851.

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

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

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

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

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

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

[25] 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.*, 28: 141-157.

[26] Van Dijk, N.M. "A formal proof for the insensitivity of simple bounds for finite multi-server non-exponential tandem queues based on monotonicity results", *Stochastic procc. Appl.*, 27: 261-277.

[27] Van Dijk, N.M. and Puterman, M.L. (1988), "Perturbation theory for Markov reward processes with applications to queueing systems", *Adv. Appl. Prob.* 20: 79-98.

[28] 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.*

[29] Walrand J. (1988). "Introduction to queueing networks.", *Prentice-*