BOUNDS ON PERFORMANCE MEASURES FOR CLOSED
QUEUEING NETWORKS: NETWORKS WITH MULTISERVER NODES

M. M. Srinivasan
Dept. of Industrial and Operations Engineering
The University of Michigan
Ann Arbor, Michigan 48109 USA

Abstract

Recent years have witnessed a substantial amount of work on developing bounds on the throughput of product form closed queueing networks. The research has mostly been directed towards networks where each node is either a single server fixed-rate type, or a delay (infinite server type).

This paper presents some simple means of extending the above work. First, a new means of obtaining bounds on the throughput of networks with multiserver nodes is presented. Second, efficient bounds on the mean queue lengths forming at the nodes in such networks are obtained.

1. Introduction:

The performance of computer systems can be evaluated by modeling them as closed queueing networks. The evaluation is done by analyzing the queueing network model for the desired measures of performance, typically under equilibrium conditions. The Product Form (PF) assumption is often used here for reasons of tractability. Quite often, situations arise where exact solutions of these queueing networks may not be required. This is especially the case when many alternate configurations are to be evaluated. In such cases, one would like to obtain approximate solutions fairly quickly. This has motivated research on obtaining bounds on the performance measures of queueing networks [4,7,10,11,12,13,14,16]. Most of the research has been on obtaining bounds on throughput for networks where each node in the network is either a fixed-rate node or a delay (infinite server) node. Recently there has been some work on developing throughput bounds for networks where the nodes are allowed to exhibit load-dependent behaviour [10,12].

This paper addresses two problems related to bounding performance measures. First, a simple means of obtaining throughput bounds for networks with multiserver nodes is developed; these bounds appear to be tighter than those reported earlier. Second, a method for obtaining bounds on the mean queue lengths forming at the nodes in such networks is presented.

In Section 2, a few preliminaries are presented. Section 3 develops the bounds on the throughput for networks with multiserver nodes, and Section 4 presents a scheme for obtaining the bounds on mean queue lengths.

1

2. Preliminaries

Let M denote the number of nodes in a network, and N the customer population. Let $\mu_m(n)$ be the average service rate at node m, when there are n customers present at that node, including, if any, the one receiving service. For fixed-rate nodes we have $\mu_m(n) = \mu_m$ and for delay nodes, $\mu_m(n) = n \cdot \mu_m$.

Now consider a network where the nodes can have a service rate function of the form:

$$\mu_m(n) = n \cdot \mu_m / (a_m + b_m n), \quad a_m, b_m \geq 0. \tag{2.1}$$

Note that if the parameter $a_m = 0$, this equation characterizes the behaviour of a fixed-rate node, and if the parameter $b_m = 0$, this equation characterizes the behaviour of a delay node. Let v_m be the mean number of visits made by a typical customer to node m, and define

$$L_m = b_m \cdot v_m / \mu_m, \tag{2.2}$$

and

$$I_m = a_m \cdot v_m / \mu_m. \tag{2.3}$$

The values L_m and I_m are for convenience referred to as the 'fixed rate load' and the 'delay load', respectively, at node m. Let L (respectively, I) denote the total fixed rate load (respectively, delay load) over all nodes. The total load on the network is then L + I. Let ρ_m (respectively, σ_m) be the ratio of the fixed rate load (respectively, the ratio of the delay load) at node m, to the total load. In the following discussion, unless otherwise specified, the index for any summation is over the range 1 through M. Hence,

$$L = \sum_m L_m, \tag{2.4a}$$

$$I = \sum_m I_m, \tag{2.4b}$$

$$\rho_m = L_m / (L+I), \tag{2.4c}$$

$$\sigma_m = I_m / (L+I). \tag{2.4d}$$

Set

$$\theta_m = \rho_m + \sigma_m. \tag{2.4e}$$

Further, define the sequence of terms $\{S_i\}$, $i = 1,2,\ldots,N$, by

$$S_i = \sum_m \rho_m^{i-1} \cdot \theta_m. \tag{2.5}$$

A sequence of increasingly tighter upper and lower bounds on the throughput of networks in which nodes can exhibit load-dependent behaviour of the form considered in equation (2.1), can then be obtained using the terms S_i [12]. These are referred to as Successively Improving (SI) bounds of increasing 'levels'. The level one upper and lower bounds on the throughput use the terms S_1 through S_2; the level two bounds use the terms S_1 through S_3, and so on. These bounds are obtained as closed form expressions. For example, if we let u denote the node with the largest load, the level three upper SI bound on the throughput, $\overline{\lambda}^3(N)$, is obtained by the expression

$$\overline{\lambda}^3(N) = 2N/[(L+I)(1 + (N-1)\alpha_0 + SQRT(T1)], \tag{2.6}$$

where

$$T1 = ((N-1)\alpha_0+1)^2 + 4\cdot[(N-1)(N-2)(\alpha_1+(N-3)\alpha_2/\overline{D}_{N-3})], \tag{2.7a}$$

with

$$\alpha_0 = S_2, \tag{2.7b}$$

$$\alpha_1 = S_3 - \alpha_0 S_2, \tag{2.7c}$$

$$\alpha_2 = S_4 - \alpha_1 S_3 - \alpha_0 S_2, \tag{2.7d}$$

and

$$\overline{D}_{N-3} = 0.5\cdot[1 + (N-1)\rho_u + SQRT([(N-1)\rho_u-1]^2+4(N-1)S_2)]. \tag{2.7e}$$

3

3. Throughput Bounds for Networks with Multiserver nodes:

Consider a network where any node with load-dependent behaviour is either a delay node, a multiserver node, or a node with rate function as given by equation (2.1). When multiserver nodes are present, exact analysis of the network requires $O(MN^2)$ computation for obtaining the throughput. Two of the bounding techniques which have been proposed recently [10,12] can be used to obtain throughput bounds for networks with multiserver nodes in $O(M)$ time.

3.1 Previous work:

The technique proposed by Srinivasan [12] would apply, in general, to include in the network any node whose service rate is a concave non-decreasing function of its queue length such as, for example, multiserver nodes and flow equivalent centers [2]. Following this technique, in the case of networks with multiserver nodes, each multiserver node m would be replaced by a node having a parametrized rate function as given by (2.1), with a suitable choice of values for the parameters a_m and b_m. Using a result shown in [15] it is possible to show that the bounds obtained on the throughput of the resulting network properly bounds the throughput of the original network.

In a recent paper, Shanthikumar and Yao [10] obtain a lower bound on the throughput for networks with multiserver nodes. This uses some results developed by them on likelihood ratio ordering and its preservation under convolution. This bound is obtained as follows: suppose we are given a network where each load-dependent node m has a service rate $\mu_m(n)$ which is a non-decreasing function of the queue length. Let n_m be the smallest integer such that $\mu_m(n_m) = \max_{n \geq 1} \mu_m(n)$; let $a_m = n_m - 1$, and set $a = \sum_m a_m$. Now consider a new network where each load dependent (non-delay) node m with rate $\mu_m(n)$ is replaced by a single server fixed rate node with rate $\mu_m^u = \max_{n \geq 1} \mu_m(n)$.

If a < N, then Shanthikumar and Yao show that the throughput of this new network at population N-a is a lower bound on the throughput of the original network at population N. Since the new network consists only of fixed rate nodes, existing bounding techniques are used in turn to obtain a lower bound on the throughput of this network, which gives the desired lower bound. We shall refer to this lower bound on throughput as the SY bound. The method is effective where N-a is not very small.

3.2 A bounding scheme:

An alternate approach to obtain bounds on the throughput of networks with multiserver nodes is now presented. Consider a network of M nodes with nodes 1 through M-1 being fixed rate nodes, having mean number of visits v_m, and service rate μ_m, $m=1,\ldots,M-1$. Let node M be a multiserver node with K parallel servers, with mean number of visits v_M, and service rate $\mu_M(n)$ defined as follows:

$$\mu_M(n) = n\mu_M, \quad n \leq K; \tag{3.1a}$$

and

$$\mu_M(n) = K\mu_M, \quad n > K. \tag{3.1b}$$

Now suppose that node M is replaced by a set of J nodes indexed $M,\ldots,M+J-1$, each having a mean number of visits v_M. Set the service rate at node M to be equal to $K\mu_M$, and set the service rates at nodes M+1 through M+J-1 to all be equal to $k\mu_M$, where k is some value much larger than K and such that

$$\frac{1}{K\mu_M} + (J-1)\frac{1}{k\mu_M} = \frac{1}{\mu_M}. \tag{3.2}$$

Then it can easily be shown that a flow equivalent center [2] M1, obtained by aggregating these J nodes, has a service rate $\mu_{M1}(n)$ such that

$$\mu_{M1}(n) \leq \mu_M(n), \quad n \geq 1.$$

Hence, to obtain a lower bound on throughput we replace the Mth node by a set of J fixed rate nodes with service rates as specified above. For this network, the 'loads' are then given, in terms of the original network

5

parameters, by

$$L_m = v_m/\mu_m, \quad m=1,..,M-1,$$

$$L_M = Kv_M/\mu_M,$$

$$L_m = kv_M/\mu_M, \quad m=M+1,..,M+J-1.$$

The total load for this network is thus

$$L = \left(\sum_{m=1}^{M} L_m \right) + (J-1) \cdot (k/K) \cdot L_M.$$

The terms S_i, $i=1,..,t+1$, required to compute a level t bound are then given by

$$S_i = \left(\sum_{m=1}^{M} \rho_m^i \right) + (J-1) \cdot (k/K)^i \cdot \rho_M^i,$$

where ρ_m is as defined by equation (2.4c). Hence, although J-1 extra nodes were introduced, the increase in computational effort is just about five more arithmetic operations for each S_i term.

A result established in [15] shows that decreasing (respectively, increasing) the service rate of a node in networks of the type considered here decreases (respectively, increases) the throughput. Since the flow equivalent center M1, resulting from the aggregation of the J fixed rate nodes has a service rate bounded from above by the service rate of the multiserver node it replaces, it follows that the throughput of the resulting network with M+J-1 nodes is less than the throughput of the original network. The resulting network of fixed rate nodes is bounded for the throughput using the approach outlined in section 2, and the resulting lower bound on the throughput for this network is clearly a lower bound for the throughput of the original network. The lower bound so obtained is henceforth referred to as the lower SI bound.

In general, for a network with more than one multiserver node, a new network is constructed where each multiserver node is replaced by a set of

fixed rate nodes as outlined above. The computational effort required to calculate the S_i terms for this network is just a little more than that required to calculate S_i terms for a network with only M fixed rate nodes.

Example 3.1:

Consider a network with 4 multiserver nodes and with customer population 20. Let $v_m = 0.1$, m=1,..,4, and let $\mu_m(n) = n$, $n \leq 4$, and $\mu_m(n) = 4$, $n > 4$, for m=1,..,4. The actual throughput of this network is 32.660.

Using the technique of Shanthikumar and Yao, the SY bound is obtained by first constructing a network of fixed rate nodes, each with rate $\mu_m = 4$. Since all rates are the same, it is easy to obtain a lower bound on the throughput which in fact is the actual throughput of this network. This value is obtained at a customer population of 8, and is equal to 29.091.

Now using the approach outlined in this paper, a lower SI bound is obtained as follows: each multiserver node is replaced by a set of 50 nodes, with service rates in each set as follows: 1 node at rate 4, and 49 nodes with rate 65.333. The level 5 lower SI bound on this resulting network is 30.09.

An upper bound on the throughput for this network is given by considering all nodes to operate at their maximum rate. This upper bound is 34.783.

Table 3.1 summarizes the bounds obtained for this example.

Population	THROUGHPUTS			
	Actual	SY bound	SI lower bound	Upper bound
20	32.660	29.091	30.090	34.783

Table 3.1: Bounds obtained for example 3.1.

In some instances, a tighter upper bound for networks with multiserver nodes may be obtained along lines as suggested in [12]. For example, suppose

it can be estimated, apriori, that the mean queue length at a multiserver

node, m, with K parallel servers would be less than K. Then instead of

obtaining an upper bound by replacing the multiserver node by a fixed-rate

node operating at the maximum service rate of the multiserver node, a node

with a service rate function of the form given by equation (2.1) could be

considered with suitably chosen parameter values for a_m and b_m.

Example 3.2:

Consider a network with 11 nodes, of which the first seven are fixed rate

nodes. Nodes 8 and 9 are multiserver with 4 parallel servers each while nodes

10 and 11 are multiserver nodes with 5 parallel servers each. The mean number

of visits to each node is 1. The service rates μ_m, m = 1,..,7 are 1/21, 1/20,

1/20, 1/15, 1/10, 1/8, and 1/8. The service rates for the multiserver nodes 8

through 11 are given by the functions

(i) $\mu_m(n) = n \mu_m$, $n \leq 4$; $\mu_m(n) = 4 \mu_m$, $n > 4$, for m = 8,9.

(ii) $\mu_m(n) = n \mu_m$, $n \leq 5$; $\mu_m(n) = 5 \mu_m$, $n > 5$, for m = 10,11.

The values for μ_m, for m=8 through 11 are, respectively, 1/80, 1/40, 1/90, and

1/95.

The SY bound is obtained as before and is shown in Table 3.2, which also

indicates the actual throughput of this network along with the other bounds.

The lower SI bound on throughput is obtained here as follows: Node 8 is

replaced by 41 nodes, one with rate 1/20 and 40 with rate 1/1.5 each; node 9

is replaced by 41 nodes, one with rate 1/10 and 40 with rate 1/0.75 each; node

10 is replaced by 41 nodes, one with rate 1/18 and 40 with rate 1/1.8 each;

node 11 is replaced by 41 nodes, one with rate 1/19 and 40 with rate 1.9 each.

As before, a simple upper bound, termed UB1 in Table 3.2, is obtained by

letting nodes 8 through 11 work at their maximum rate throughout. An alternate

upper bound, termed the upper SI bound, is also shown. This is obtained by

letting nodes 8, 10, and 11 operate throughout at their maximum rate. Node 9

is replaced by a node with a rate function of the form given by equation (2.1). The values for a_9, and b_9 here were set at 0.888 and 0.028 respectively. These values ensure that the rate function for the multiserver node 9 is properly bounded from above by the rate function of the replacing node. From Table 3.2, it is seen that this results in an improved bound. The bounds are presented at population values of 20, 30, and 40.

| Population | THROUGHPUTS | | | | |
	Actual	SY bound	SI lower	UB1	SI upper
20	0.0344	0.0233	0.0318	0.0389	0.0385
30	0.0400	0.0359	0.0378	0.0427	0.0425
40	0.0428	0.0403	0.0409	0.0448	0.0448

Table 3.2: Bounds obtained for example 3.2.

4. Bounds on mean queue lengths:

Techniques now exist for obtaining reasonably tight bounds on throughput with relatively low computational effort compared to that required for exact analysis. This suggests the possibility of achieving good bounds on the mean queue lengths which form at the nodes. Needless to say, these bounds should require little computational effort compared to that required to obtain the exact values.

In section 4.1, some bounds are presented which are very easy to evaluate once bounds on the throughput have been obtained. These bounds appear adequate for nodes which do not experience very high utilizations (e.g., over 80%). For nodes with high utilizations, an alternate bounding scheme is developed in section 4.2. This latter method is based on an application of the convolution algorithm [1] and requires a relatively larger number of arithmetic operations to obtain the mean queue length bounds. The discussion below is restricted, for simplicity, to networks with fixed rate nodes. It is easily extended to networks with delay nodes, load dependent centers of the form given by (2.1), and networks with multiserver nodes.

4.1 A set of simple bounds on mean queue length:

For networks with only single server fixed-rate nodes nodes, the mean queue length at a fixed-rate node, m, is given as [3]:

$$Q_m(N) = \sum_{k=1}^{N} L_m{}^k \Lambda(N,k),$$ (4.1)

where

$$\Lambda(N,k) = \lambda(N) \ldots \lambda(N-k+1),$$ (4.1a)

and $\lambda(K)$ is the throughput of the network with population K. It is shown in [15] that for the kind of networks considered here,

$$\lambda(K) \geq \lambda(K-1); \quad K > 0.$$

Hence

$$Q_m(N) \leq \sum_{k=1}^{N} L_m{}^k [\lambda(N)]^k$$

$$= U_m(N) \; \frac{1 - (U_m(N))^N}{1 - U_m(N)} \qquad (4.2)$$

where

$$U_m(N) = L_m \lambda(N), \qquad (4.3)$$

is the utilization of node m. Let $\bar{\lambda}(N)$ denote an upper bound on $\lambda(N)$ and let $\bar{U}_m(N) = L_m \bar{\lambda}(N)$ denote the corresponding upper bound on the utilization at node m. Then it is easily seen that

$$Q_m(N) < \bar{U}_m(N) \; \frac{1 - (\bar{U}_m(N))^N}{1 - \bar{U}_m(N)}. \qquad (4.4)$$

Let L_u denote the load at the node with the highest load. As the network population increases, the network begins to behave like an open network with mean arrival rate $1/L_u$. For this corresponding open network, any node m with load $L_m < L_u$ will experience a mean queue length $Q_m = U_m/(1 - U_m)$ where $U_m = L_m/L_u$ would be the utilization of this node in the open network. Hence, as N becomes large the upper bound on mean queue lengths, as given by equation (4.4) would begin to get increasingly closer to the exact value for these nodes. For the node(s) with load L_u, for large N, an upper bound is given by

$$\bar{Q}_u(N) = 1/k \; (N - \sum_{m \neq u} \underline{Q}_m(N)) \qquad (4.5)$$

where k denotes the number of nodes with load L_u and $\underline{Q}_m(N)$ is a lower bound on $Q_m(N)$.

A lower bound on $Q_m(N)$ is given by Theorem 4.1:

Theorem 4.1:

In a closed PF network with only single server fixed rate nodes and delay nodes, the mean queue length, $Q_m(N)$, at a fixed rate node m with network population N, is bounded from below by $f(U_m(N))$, where

$$f\big(U_m(N)\big) = U_m(N) \cdot \frac{(N-1)\,\big(1 - U_m(N)\big) + 1}{(N-1)\,(1 - U_m(N))^2 + 1}$$ (4.6)

A proof of Theorem 4.1 is given in [12]. Hence, with appropriate bounds on the utilization $U_m(N)$, at node m, equation (4.6) yields a lower bound on $Q_m(N)$.

4.2 Obtaining tighter bounds on mean queue lengths:

Although the bounds presented in Section 4.1 are easy to evaluate, they can be quite loose for nodes which have high utilizations. It is, however, possible to obtain an exact closed form expression for the mean queue lengths in terms of a throughput with some additional computation. This is achieved by means of the Convolution algorithm [1]. For ease of presentation, only networks with fixed rate servers are considered here.

The Convolution algorithm for a network T with M fixed rate single server nodes gives an expression for the normalizing constant g(N,M) as

$$g(N,M) = g(N,M-1) + L_M g(N-1,M).$$ (4.7)

The throughput for this network, $\lambda(N)$, is then given as

$$\lambda(N) = g(N-1,M)/g(N,M).$$ (4.8)

Suppose it is desired to obtain bounds for the mean queue length at some designated node. Without loss of generality, let this be node M. The mean queue length at this node is given as [3]

$$Q_M(N) = \sum_{k=1}^{N} L_M{}^k \, g(N-k,M)/g(N,M) .$$ (4.9)

Now consider an augmented network $T^{(M)}$ with M+1 nodes where nodes 1 through M

have the same loads as in T and node M+1 has a load L_M. Let $\lambda^{(M)}(N)$ be the throughput of this augmented network. Lemma 4.1 then obtains an expression for $Q_M(N)$ in terms of $\lambda^{(M)}(N)$.

Lemma 4.1

Given a network, T, of M single server fixed rate nodes, the mean queue length at the Mth node, $Q_m(N)$ is given by

$$Q_M(N) = \frac{\lambda^{(M)}(N) \cdot L_M}{1 - \lambda^{(M)}(N) \cdot L_M} , \qquad (4.10)$$

where $\lambda^{(M)}(N)$ is the throughput of the network T augmented by one additional node, M+1, with load L_M.

Proof:

The normalization constant of the augmented network, $T^{(M)}$, is given by

$$g(N,M+1) = g(N,M) + L_M\, g(N-1,M-1), \qquad (4.11)$$

and the throughput of $T^{(M)}$ is

$$\lambda^{(M)}(N) = g(N-1,M+1)/g(N,M+1). \qquad (4.12)$$

By repeated application of (4.11) in equation (4.12), we get

$$\lambda^{(M)}(N) = \frac{g(N-1,M) + L_M\, g(N-2,M) + \ldots + L_M^{N-1}g(0,M)}{g(N,M) + L_M\, g(N-1,M) + \ldots + L_M^N g(0,M)}$$

$$= \lambda(N)\, \frac{1 + Q_M(N-1)}{1 + Q_M(N)} , \qquad (4.13)$$

where the last equality is obtained using equations (4.8) and (4.9).

From the Mean Value Analysis algorithm [9], an expression relating $Q_M(N)$ with $Q_M(N-1)$ can be obtained as:

$$Q_M(N) = \lambda(N)\, L_M + \lambda(N)\, L_M\, Q_M(N-1). \qquad (4.14)$$

Substituting the expression for $Q_M(N-1)$ resulting from (4.14) into equation (4.13), and simplifying, the desired result in obtained. □

13

In general, in order to use the approach outlined above, a new augmented network is to be constructed for each fixed rate node with a distinct load, for which mean queue length bounds are desired. This augmented network should then be analyzed, using a bounding technique, to obtain bounds on its throughput. Equation (4.10) should then be used to obtain the bounds of the mean queue length.

Suppose the SI bounding technique is used. Assume that level three bounds on the throughput have been obtained. This means that the terms S_i, for $i = 2, 3, 4$ must have been calculated, where S_i is defined by equation (2.5). Now, to obtain mean queue length bounds for a node, n, $n \leq M$, with load L_n, the augmented network with $M+1$ nodes is constructed. The values of the relative utilizations $\rho_m^{(n)}$, $m = 1, \ldots, M+1$, for this augmented network, are given by

$$\rho_m^{(n)} = \frac{\rho_m}{L + L_n} \cdot L_m, \quad m = 1, \ldots, M+1.$$

Let

$$S_i^{(n)} = \sum_{i=1}^{M+1} (\rho_m^{(n)})^i, \quad i = 2, \ldots, 4. \tag{4.15}$$

$$= (S_i + \rho_n^i)/(1 + \rho_n)^i. \tag{4.16}$$

Hence the terms $S_i^{(n)}$ are easy to evaluate, given the values for S_i and the relative utilizations. A level three upper bound on the throughput of the augmented network is then given by equation (2.7) where the terms α_0, α_1, and α_2 are replaced by $\alpha_0^{(n)}$, $\alpha_1^{(n)}$, and $\alpha_2^{(n)}$. The terms $\alpha_i^{(n)}$ are defined as

$$\alpha_i^{(n)} = S_{i+2}^{(n)} - \sum_{j=0}^{i-1} S_{i+1-j}^{(n)} \alpha_j, \quad i = 0,1,2.$$

Using this upper bound on $\lambda^{(M)}(N)$ in equation (4.10), an upper bound on the mean queue length at node n is obtained.

The use of the bounding techniques developed in this section is illustrated below with a few examples. In each case, the approach taken is to

14

use bounds developed in Section 4.1 for nodes with lower utilizations, and use the bounds from Section 4.2 for nodes with higher utilizations. The basis for determining the bounding scheme to be used was arbitrarily set as follows: from the upper bound on throughput, the upper bound on the utilizations at the nodes was determined using equation (4.3). Nodes with this upper bound value less than, or equal to, 0.80 were evaluated for their mean queue length bounds by the method of Section 4.1, while the other nodes used the method of Section 4.2. The SI bounding technique was used in all test cases, and level 5 bounds were used to calculate the throughput bounds.

For comparison, the exact values were also evaluated. The test cases were run on an AMDAHL 5860 running the MTS. The time taken by the exact analysis and the bounding technique were recorded in each case.

Example 4.1:

This example is taken from [4]. There are fixed rate 50 nodes with loads as follows: 1 node at 20, 2 nodes at 19, 5 nodes at 18, 5 nodes at 15, 5 nodes at 10, 8 nodes at 7, 8 nodes at 5, 8 nodes at 4, 8 nodes at 2, for a total load of 417. The bounds were evaluated at a network population of 50. The exact analysis required 9 milliseconds while the bounding technique required 2 milliseconds. Table 4.1 compares results for some of these nodes. For comparison purposes, bounds were evaluated for all 50 nodes, even though some nodes had identical loads.

Example 4.2:

The second example here is a very unbalanced network with fixed rate 20 nodes, and loads as follows: 2 nodes at 30, 1 node at 17, 3 nodes at 14, 1 node at 12, 2 nodes at 11, 1 node at 10, 2 nodes at 8, 1 node at 7, 2 nodes at 5, 3 nodes at 4, 1 node at 3, and 1 node at 1 for a total load of 212. Table

4.2 compares results for some of the nodes. The comparisons were made for a population of 45. The exact analysis required about 3.3 milliseconds while the bounding technique required about 0.7 millisecond of CP time.

Load at node	Mean queue lengths		
	Exact	Lower bound	Upper bound
20	5.181	4.484	5.691
19	4.027	3.531	4.329
18	3.207	2.464	3.749
15	1.781	1.544	1.923
10	0.753	0.701	0.781
7	0.431	0.409	0.443
5	0.274	0.262	0.281
4	0.208	0.200	0.213

Table 4.1: Mean queue length bounds for example 4.1

Load at node	Mean queue lengths		
	Exact	Lower bound	Upper bound
30	18.540	16.650	18.681
17	1.228	1.145	1.308
14	0.831	0.796	0.875
12	0.637	0.617	0.667
10	0.480	0.468	0.500
8	0.350	0.344	0.364
7	0.294	0.289	0.304
4	0.149	0.147	0.154

Table 4.2: Mean queue length bounds for example 4.2

It is important to note that in many instances, it may not be necessary to evaluate mean queue lengths at all nodes. (This is certainly true if many nodes have the same loads). It is especially in such cases that a bounding technique would have an edge over the exact analysis methods such as the MVA algorithm which would have to compute these mean queue lengths for all nodes, in any case, and for all intermediate population values. In the examples above, many nodes had the same load and it was not necessary to evaluate mean queue length bounds for more than one such node. This would have substantially improved on the time performance of the technique. This was not done so that a better comparison could be made, in general, between the times taken for the exact analysis and the bounding technique.

5. Conclusions:

A new method of obtaining bounds on the throughput of closed queueing networks with multiserver nodes has been presented. The bounds are expressed in terms of the bounds obtained for a network of fixed rate and delay nodes which is constructed by replacing every multiserver node in the original network by a set of fixed rate nodes. The bounds obtained are fairly tight, easy to evaluate, and compare very favorably to bounding techniques presented earlier for such networks. A simple, but effective means of obtaining bounds on the mean queue lengths that form at the nodes in such networks has also been presented.

REFERENCES

1. BUZEN, J.P., "Computational algorithms for closed queueing networks with exponential servers", C. ACM, v 16, no 9, Sept 1973, pp. 527-531.

2. CHANDY, K.M., HERZOG, U., and WOO, L., "Approximate analysis of general queueing networks", IBM J. Res. Dev., v 19, pp. 43-49.

3. DENNING, P.J., and BUZEN, J.P., "The operational analysis of queueing network models", Computing Surveys, v 10, 3, Sept 1978, pp. 225-262.

4. EAGER, D.L., and SEVCIK, K.L., "Performance bound hierarchies for queueing networks", ACM TOCS, vol 1, No 2, May 1983, pp. 99-115.

5. GORDON, W.J., and NEWELL, G.F., "Closed queueing systems with exponential servers", Operations Research, v 15, n 2, pp. 254-265.

6. KLEINROCK, L., Queueing systems, Volume 2: Computer applications, Wiley, New York, 1976.

7. KRIZ, J., "Throughput bounds for closed queueing networks", Performance evaluation, v 4, 1984, pp. 1-10.

8. LITTLE, J.D.C., "A proof of the queueing formula L = λW". Operations Research, 9, 1961, pp. 383-387.

9. REISER, M., and LAVENBERG, S.S., "Mean value analysis of closed multichain queueing networks", J. ACM, 27, 2, April 1980, pp. 313-322.

10. SHANTHIKUMAR, J.G., and YAO, D.D., "Throughput bounds for closed queueing networks with queue-dependent service rates", July 1985.

11. SRINIVASAN, M.M., "Successively improving bounds on performance measures for product form queueing networks", Tech. Report 85-2, Department of I & OE, University of Michigan, 1985.

12. SRINIVASAN, M.M., "On extending the scope of a bounding technique for closed queueing networks", T-ch. Repost 85-25, Department of I & OE, University of Michigan, 1985.

13. STEPHENS, L.E., and DOWDY, L.W., "Convolutional bound hierarchies", Performance Evaluation Review, Proceedings of the 1984 Sigmetrics Conference on Measurement and Modelling of Computer Systems, pp. 120-133.

14. SURI, R., "Generalized quick bounds for performance of queueing networks", Computer Performance, vol 5, No 2, June 1984, pp. 116-120.

15. SURI, R., "A concept of monotonicity and its characterization for closed queueing networks", Operations Research, v 33, 1985.

16. ZAHORJAN, J., SEVCIK, K.C., EAGER, D.L., and GALLER, B., "Balanced job bound analysis of queueing networks", C. ACM, 25, 2, Feb 1982, pp. 134-141.