PERFORMANCE '87
P.-J. Courtois and G. Latouche (editors)
Elsevier Science Publishers B.V. (North-Holland), 1988

General Closed Queueing Networks with Blocking

I. F. Akyildiz

School of Information and Computer Science
Georgia Institute of Technology
Atlanta, Georgia 30332
U. S. A.

ABSTRACT

Approximate solution techniques are given for closed networks with transfer blocking, general service time distributions and FCFS scheduling disciplines. In the first part of this work two station networks with blocking are analyzed. It is shown that the results are exact if one of the stations possesses negative exponential service time distribution. For networks with more than two stations the throughput analysis technique is introduced. The idea is based on the concept that a nonblocking network with certain total number of jobs can be found of which the average number of active stations equals the average number of active stations in the blocking network. The throughput of the nonblocking network is then approximately equal to the throughput of the blocking network. Since the total number of jobs in the nonblocking network can be noninteger number, the α-MVA is used for the computation of throughput values in case of exponential service time distributions. For the computation of the mean number of jobs in each station the finiteness of station capacities is ignored and the Method of Marie in its original form is executed and the service rates are adjusted, i.e., the service rates are exponentialized. The network is then transformed into a Markovian network with blocking which is then analyzed using the State Normalization Technique. Analytical results are validated by simulation of several numerical examples.

Key Words: Closed Queueing Networks, Finite Station Capacities, Method of Marie, Performance Measures.

1. Introduction

In recent years there has been a growing interest in the development of computational methods to analyze queueing networks with blocking. These are networks where the stations have finite capacities, hence blocking can occur if the station is full to capacity. A job which wants to enter the full station must reside in its source station and block the source station until a place is available in the destination station. The interest in networks with blocking comes primarily from the realization that these models are useful in the study of the behavior of subsystems of computers and communication networks, in addition to detailed descriptions of several computer-related applications such as flexible manufacturing systems.

This work is supported in part by the Air Force Office of the Scientific Research (AFOSR) under Grant AFOSR 87-0160

Blocking networks are classified into 3 different types [17].

a) **Transfer Blocking.** In this case, the blocking event occurs when a job completing service at station i cannot proceed to station j because station j is full. The job resides in station i's server, which stops processing until station j releases a job. This type of blocking has been used to model systems such as production systems and disk I/O subsystems [1,2,3,4,6,7,10,18,21,22,27,28].

b) **Service Blocking.** In this case, blocking occurs when a job in front of queue at station i declares its destination station j before it starts its service in station i's server. If the destination station j is full, the i-th server becomes blocked, i.e., it cannot serve jobs. When a departure occurs from destination station j, the i-th server becomes unblocked and the job begins receiving service. This blocking type has been used to model systems such as telecommunication and production systems [9,13,15,26].

c) **"Rejection Blocking".** In this case, blocking occurs when a job completes service at station i's server and wants to join station j, whose capacity is full. The job is rejected by station j. That job goes back to station i's server and receives a new service with the same mean service time. This activity is repeated until station j releases a job, and a place becomes available. The "rejection blocking" type has been used to model systems such as communication networks, computer systems with limited multiprogramming, production lines and flexible manufacturing systems [5,8,14,23].

Formal comparisons between these distinct classes of blocking have been carried out extensively by Onvural/Perros [17]. Several investigators in recent years have published results on queueing networks with transfer, service and rejection blocking. A bibliography concerning queueing network models with blocking is given by Perros [19].

2. Model Description

In this work we investigate closed queueing networks with transfer blocking and with the following assumptions. There are N stations and the number of jobs in the system is fixed at K. All jobs belong to the same class. Each station has a single server and a generally distributed service time with mean value $1/\mu_i$ $(i=1,2,...,N)$ and coefficient of variation c_i for $i = 1,2,...,N$ having rational Laplace transform. Each station has a finite fixed capacity B_i $(i=1,2,...,N)$ where $(B_i = \textit{queue capacity} + 1)$. The cases where the stations can have infinite capacity are also allowed. Any station whose capacity reaches the total number of jobs in the network can be considered to have infinite capacity. A job which is serviced by the i-th station proceeds to the j-th station with probability p_{ij}, if the number of jobs in the j-th station has not reached the capacity B_j, for $i,j=1,2,...,N$. Otherwise, the job is blocked in the i-th station's server until the destination station has an available place. The service discipline in each station is First-Come First Served. If there are several stations all linked to station j then it is possible that any time there might be blocked jobs from more than one station waiting to enter station j. The maximum number of blocked jobs will be equal to the number of stations which are directly connected to station j. We assume that these blocked jobs enter station j on a first-blocked-first-deblocked basis. Furthermore, we assume that the network must be deadlock-free. Finite station capacities and blocking can introduce the deadlock situation. In a simple example, deadlock may occur if a job which has finished its service at station i's server wants to join station j, whose capacity is full. That job is blocked in station i. Another job which has finished its service at j-th station now wants to proceed to the i-th station, whose capacity is also full. It blocks the j-th station. Both jobs are waiting for each other. As a result, a deadlock situation arises. The possibility of deadlock in a network increases with the ratio of the number of jobs in the network to the total capacity of the network.

Necessary and sufficient condition for a closed queueing network with transfer blocking to be deadlock free is given in the following. First we define a cycle C as a sequence of stations $(y_1, y_2, ..., y_j)$ each pair of consecutive stations is joined by an arc (y_i, y_{i+1}), including the arc (y_j, y_1).

A closed queueing network of transfer type blocking is deadlock free [3], if and only if for each cycle C in the network the following condition (1) holds. Simply stated, the total number of jobs in the network must be smaller than the sum of station capacities in each cycle.

$$K < \sum_{j \in C} B_j \tag{1}$$

A necessary and sufficient condition for a tandem network to be deadlock free is given by:

$$K < \sum_{i=1}^{N} B_i \tag{2}$$

The Method of Marie [16], in short form MM, is one of the most accurate approximations for closed single class queueing networks with non-exponential servers. Therefore we use the Method of Marie in our investigation of transfer blocking networks with general service time distributions.

3. Method of Marie

The algorithm steps are outlined below:

1. For any mean μ and any squared coefficient of variation $c^2 \geq 0.5$, it is possible to represent a station's server by a Cox-model with 2 phases [16]. The parameters for the Cox distribution is determined by:

$$\mu_{i1} = 2\mu_i \tag{3a}$$

$$\mu_{i2} = \frac{\mu_i}{c_i^2} \tag{3b}$$

$$a_i = \frac{1}{2\,c_i^2} \tag{3c}$$

$$b_i = 1 - a_i \tag{3d}$$

μ_{ij} denote the service rate at the $j-th$ phase $(j = 1,2)$ of the i-th station $(i = 1,2,...,N)$ and a_j be the probability that a job upon completion of its service at the j-th phase will proceed to the $(j + 1)$-th phase and b_j denotes the probability that a job upon completion of its service at phase j will depart from the station.

2. Iterative Part

 i) The arrival rates to the i-th station are denoted by $\lambda_i(k)$. To determine the load-dependent arrival rates $\lambda_i(k)$ (for $k = 0,1,2,...,K$) of a station i (for $i = 1,...,N$), the $i-th$ station is shorted, i.e., its service time is set to zero. The subnetwork containing all stations but i is assumed to have product form solution which is then analyzed and the throughput values $\lambda'_c(k)$ are obtained by any product form network algorithm such as mean value analysis or convolution algorithm [25]. The load-dependent arrival rates $\lambda_i(k)$ to the $i-th$ station are then:

$$\lambda_i(k) = \lambda'_c(K-k) \quad \text{for } k = 0,1,...,K-1 \tag{4}$$

It is clear that if k jobs are present at the $i-th$ station, then $(K-k)$ jobs remain in the subnetwork. Thus the throughput of the subnetwork with $(K-k)$ jobs is equal to the arrival rate of the i-th station with k jobs. Note that

$$\lambda_i(K) = 0 \tag{5}$$

since no job is in the subnetwork and consequently the throughput will be zero, $\lambda'_c(0) = 0$. In this way each station is shorted and the throughput λ'_c of the corresponding subnetwork is computed and assigned to the according arrival rates λ_i.

ii) Compute the conditional throughputs $\nu_i(k)$ using [16]:

$$\nu_i(k) = \frac{b_i\ \lambda_i(k)\ \mu_{i1} + \mu_{i1}\ \mu_{i2}}{\lambda_i(k) + \mu_{i1} + \mu_{i2} - \nu_i(k-1)} \tag{6}$$

where

$$\nu_i(1) = \frac{b_i\ \lambda_i(1)\ \mu_{i1} + \mu_{i1}\ \mu_{i2}}{\lambda_i + \mu_{i2} + a_i\ \mu_{i1}} \tag{7}$$

iii) Compute the equilibrium state probabilities for stations $i = 1,2,\dots,N$ using:

$$p'_i(k) = p_i(0) \prod_{n=0}^{K-1} \frac{\lambda_i(n)}{\nu_i(n+1)} \qquad \text{for } k = 1, \cdots, K. \tag{8}$$

where

$$p'_i(0) = \left[1 + \sum_{k=1}^{K} \prod_{n=0}^{k-1} \frac{\lambda_i(n)}{\nu_i(n+1)} \right]^{-1} \tag{9}$$

iv) After each iteration check to see if the sum of the mean number of jobs is equal to the total number of jobs in the given network within a tolerance level ε:

$$\frac{\left| K - \sum_{i=1}^{N} \sum_{k=1}^{K} k \cdot p_i(k) \right|}{K} < \varepsilon \tag{10}$$

If the test is not successful, then adjust the service rates:

$$\mu_i(k) := \nu_i(k) \qquad \text{for } k = 1, \dots, K. \tag{11}$$

Informally equation (11) states that the conditional throughputs, equations (6 and 7) are assumed to be the new service rates and the next iteration is carried out with these new service rates. Iterations continue until acceptable tolerances are obtained. ε is a tolerance level. Usual value is $\varepsilon = 10^{-4}$.

Note that in the first iteration the service rates of each station are the originally given values μ_i. In future iterations the conditional throughputs $\nu_i(k)$, are used as the adjusted service rates for the stations.

4. Two Station Networks with Blocking

A state of a station is denoted by a pair:

$$\underline{k} = (k, j)$$

where

k is the number of jobs, $k = 0,1,2,..,K$

j is the phase number, $j = 1,2$.

A transition from one state to another takes place either when a new job arrives or when a job leaves the station through either phase. We will represent arrival rates by $\lambda_i(k)$, where the arrival rates are dependent on the number of jobs in the station. Since we are dealing with two-station networks, it is clear that the arrival rate $\lambda_i(k)$ of the i-th station is equal to the service rate of the j-th station for $i,j = 1,2 ; i \neq j$. A job leaving the station after phase one is denoted by $b_i\, \mu_{i1}(k)$. $\mu_{i2}(k)$ is the departure rate of a job leaving the station after phase two. A job enters into phase two at a rate $a_i\, \mu_{i1}(k)$ and may do so only after receiving service in phase one. The state (0) denotes that no job is in the i-th station.

The feasible states for the blocking network are obtained by considering the fact that the number k_i of jobs in the i-th station may not exceed the capacity of the station B_i. On the other hand, at least $(K - B_j)$ jobs can be accomodated by the i-th station due to the capacity restriction of the j-th station. The number of jobs in the i-th station has the following bounds:

$$(K - B_j) \leq k_i \leq B_i \tag{12}$$

The immediate neighbors of a feasible state represent the blocking states in the network, i.e., whenever a transition occurs from one state to another state where the capacity limit of a station would be violated, we assume that the transition causes a blocking of a server of that particular station. The immediate neighbors of the feasible states are included in the state space of the blocking network. Except for the blocking states denoted in Figure 1 by a "*", all the other states violating the station capacities are non-feasible and are cancelled. The bounds, equation (12), for the number of jobs in the i-th station are extended as follows:

$$(K - B_j - 1) \leq k_i \leq (B_i + 1) \tag{13}$$

We obtain then the complete state space for the blocking network as follows:

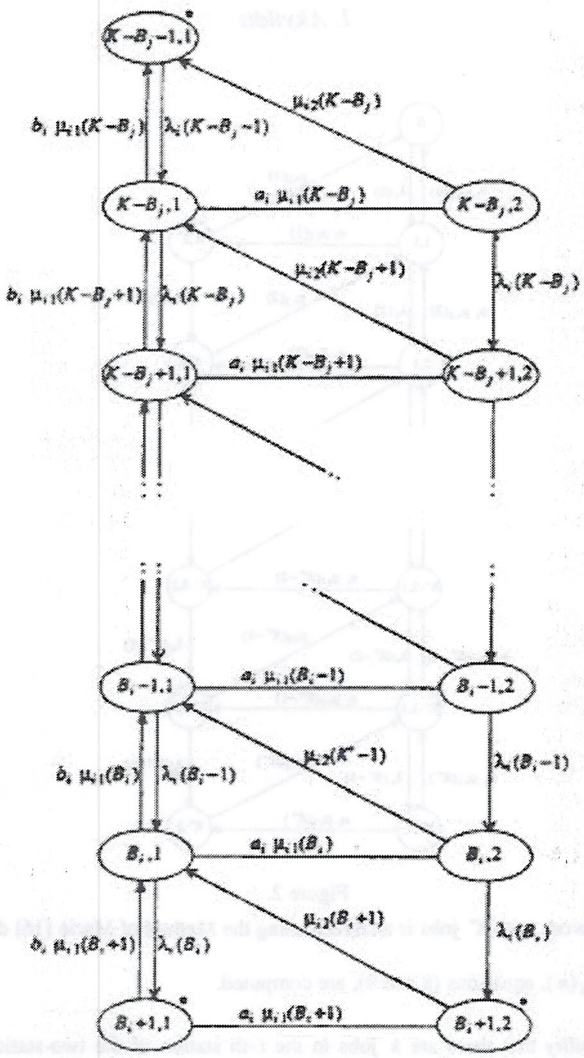

Figure 1.

As in case of Markovian networks [1], also here we can find an equivalent network without limitation of station capacities with appropriate total number of jobs K' which provides exactly the same state space structure as the blocking network given in Figure 1. The equivalent non-blocking network has the same number of stations; the same generally distributed service times; the same transition probabilities. The only difference is that the station capacities are unlimited, hence no blocking can occur and the total number of jobs in both systems is not equal, $K \neq K'$, where K' is computed by [1]:

$$K' = \min\{K, B_1 + 1\} + \min\{K, B_2 + 1\} - K \tag{14}$$

The state space for the non-blocking network with K' jobs is given in Figure 2.

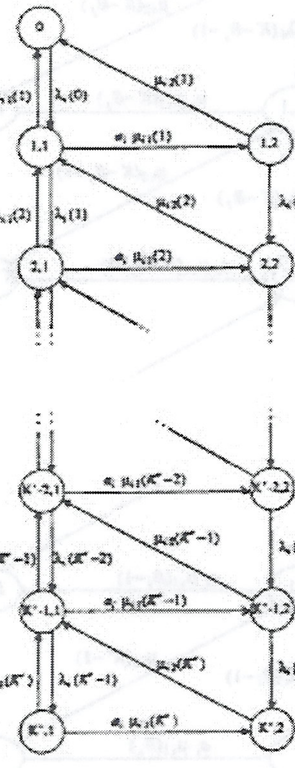

Figure 2.

The non-blocking network with K' jobs is analyzed using the Method of Marie [16] described in section 3 and the marginal probabilities $p'_i(n)$, equations (8 and 9), are computed.

The *marginal probability* that there are k jobs in the i-th station of the two-station network with transfer blocking is obtained from the marginal probabilities of the equivalent non-blocking network:

$$p_i(k+s) = p'_i(s) \quad \text{for } s = 0,1,2,\ldots,K' \; ; \; i = 1,2 \tag{15}$$

where k is defined in the range given by equation (13).

The *mean number of jobs* is computed by:

$$\bar{k}_i(K) = \sum_{n=K-B_i}^{B_i} n \cdot p_i(n) + B_i \, p_i(B_i+1) + (K-B_j) \cdot p_i(K-B_j-1) \tag{16}$$

for $i, j = 1, 2 \; ; \; i \neq j$.

The *throughput* of the blocking network is obtained by:

$$\lambda_i(K) = [\, 1 - p'_i(0) \,] \cdot \mu_i \quad \text{for } i = 1,2 \tag{17}$$

The *mean residence time* (time spent by a job in queue, in service and in the blocking phase) is calculated by Little's law.

$$\bar{t}_i(K) = \frac{\bar{k}_i(K)}{\lambda_i(K)} \quad \text{for } 1,\ldots,N. \tag{18}$$

REMARK

As generally known and also as mentioned above the Method of Marie [16], *MM*, is an approximation method which is one of the most accurate techniques for closed single class queueing networks with nonexponential servers. Even though the state space transformation concept is exact for two-station Markovian networks [1] because of the *MM* the results obtained so far are approximate. However, *MM* provides exact solution, if one station has negative exponential service time distribution and FCFS service discipline. Consequently, two-station blocking networks with one station having negative exponential service time distribution can be solved exactly.

5. Approximate Analysis of Multiple Station Networks with Blocking

As mentioned before, the load dependent arrival rates $\lambda_i(k)$ of a station i are needed in the *MM* for further computations. In order to determine these arrival rates, the i-th station is shorted, i.e., its service time is set to zero. The subnetwork containing all remaining stations but the i-th, is assumed to have negative exponential service time distribution. In other words, the subnetwork is a Markovian network with finite capacities. Using the throughput algorithm for blocking networks [2] the Markovian subnetwork could be analyzed and the throughput values could be obtained which could then be assigned to the according arrival rates of the i-th station. However, a major problem arises in the selection of the entire capacity of the subnetwork during the analysis of the i-th station. If the capacity of the subnetwork is selected as

$$B_\sigma = \sum_{j \in \sigma} B_j$$

(where σ represents the subsystem).

our study showed that this overestimates throughput, since the shorted station can be blocked in the actual network with less than B_j jobs in the subsystem σ.

Another possibility is using a capacity weighted by the transition probabilities from the shorted station i to the stations in the subsystem σ:

$$B_\sigma = \sum_{j \in \sigma} B_j \; p_{ij}$$

However, we discovered that this underestimates the throughput.

We omit in our approach the problem of finding an appropriate capacity for the subnetwork. Our approach is described in the following chapter.

5.1. Throughput Analysis

In case of blocking networks with $N > 2$ stations, there exists no equivalent non-blocking network with an isomorphic state space as in the case of two-station networks with blocking as shown in section 4.

However, we can find a nonblocking network with appropriate number of jobs K' which has approximately the same throughput as the blocking network. The nonblocking network has the same number of stations, the same generally distributed service times and the same transition probabilities as the blocking network. The only difference is the total number of jobs K' in the nonblocking network, where K' is smaller than K. K' is determined in a way that the average number of active stations in the nonblocking network equals the average number of active stations in the blocking network. A station is active, if it contains at least one job and it is not blocked.

Note that the algorithm provides non-integer value for K'. In case of general service time distributions we round up K' to the next integer number and analyze the non-blocking network by the *MM*, described in section 3, which is applicable only to networks with integer number of jobs. This will be explained in the first part of the following section.

If all stations of the network have exponentially distributed service times, the nonblocking network is analyzed with the α-mean value analysis [11], in short form (α-mva), which is a modification of the well known mean value analysis [24] for networks with noninteger total number of jobs. This will be explained in the second part of the following section.

5.1.1. Non-exponential Service Times

The correct determination of the number of jobs K' in the nonblocking network strongly influences the accuracy of the method. We compute K' as follows. First we define $\underline{k} = (k_1,...,k_N)$ as a state of the network where k_i is the number of jobs in the $i-th$ station. A state is feasible, if the number of jobs in each station i satisfies the following inequality:

$$k_i \le B_i$$

Additionally, we have to consider the blocking events in the network. In order to obtain the blocking states we determine the number of stations which can be blocked in each feasible state m. The station i can be blocked in a state m, if it has at least one job ($k_i > 0$) and has a positive transition probability ($p_{ij} > 0$) to a station j, ($k_j = B_j$) which is full at its capacity, i.e., ($k_j = B_j$). Considering the fact that no station, one station or several stations can be blocked in m we compute the possible total number of siblings of the feasible state m as follows:

$$z_m = 2^{b_m} \tag{19}$$

b_m denotes the number of stations which can be blocked in a feasible state m. Note that the feasible state m is also counted in z_m. The total number of states Z in the blocking network is then obtained from

$$Z = \sum_{m=1}^{n} z_m = \sum_{m=1}^{n} 2^{b_m} \tag{20}$$

where n is the maximum number of feasible states.

Next we determine the number A of stations which are active in all states Z.

$$A = \sum_{m=1}^{n} z_m \left[a_m - \frac{b_m}{2} \right] \tag{21}$$

where a_m is the number of stations with at least one job in the m-th feasible state.

From this we obtain the average number r of the active stations over all states:

$$r = \frac{A}{Z} \tag{22}$$

Our objective is to compute the number of jobs K' in the corresponding non-blocking network such that the average number r' of active stations in the nonblocking network equals r, equation (22).

The total number of states Z' is computed from the binomial coefficient formula:

$$Z' = \begin{bmatrix} N + K' - 1 \\ N - 1 \end{bmatrix} \tag{23}$$

The number of active stations A' in a nonblocking network with K' jobs are obtained from:

$$A' = \sum_{m=1}^{Z'} a_m \tag{24}$$

where a_m is the number of stations with at least one job in a state m.

r' is computed from:

$$r' = \frac{A'}{Z'} \tag{25}$$

Note that r' is strictly monoton increasing with K'

$$r'(K') < r'(K' + 1) \tag{26}$$

and is bounded by:

$$1 \le r'(K') < N \tag{27}$$

where N is the number of stations in the network.

The value for K' is non-integer and is between two integer values L and $L+1$

$$r'(L) < r < r'(L+1) \tag{28}$$

From equations (22, 25 and 28) we obtain

$$K' = L + \frac{r - r'(L)}{r'(L+1) - r'(L-1)} \tag{29}$$

The corresponding nonblocking network with K' jobs has approximately the same throughput as the blocking network with K jobs [2]:

$$\lambda(K) \approx \lambda'(K') \tag{30}$$

In case of general service time distributions we analyze the nonblocking network with the MM, given in section 3. As mentioned before K' has non-integer value which must be rounded to the closest integer value. If K' is exactly in the middle of two integer numbers, the higher number is selected. Note that there exists an integer K' only in the following case:

$$r'(K') = r \tag{31}$$

The throughput of each station in the blocking network are computed from:

$$\lambda_i(K) = e_i\, \lambda(K) \tag{32}$$

5.1.2. Exponential Service Times

As mentioned in the beginning of this section the non-blocking network can be analyzed with the non-integer number of jobs K', if all stations have exponentially distributed service times. For this we use the α-mva [11], where

$$K' = L + \alpha$$

and L is the next smaller integer number of K'. In the following we outline the α-mva:

α-mean value analysis:

- Initialization:

$$\bar{k}_i(0) = 0 \qquad \text{for } i = 1,2,\ldots,N.$$

- Iteration: $k = 0,1,2,...,L$
  - Mean Residence Time of Jobs in station i:

$$\bar{T}_i(k+\alpha) = \frac{1}{\mu_i}\left[\alpha + \bar{k}_i(k-1-\alpha)\right] \tag{33}$$

  - Total Throughput:

$$\lambda(k+\alpha) = \frac{k+\alpha}{\sum_{i=1}^{N} e_i\ \bar{T}_i(k+\alpha)} \tag{34}$$

  - Mean Number of Jobs:

$$\bar{k}_i(k+\alpha) = \lambda(k+\alpha)\, e_i\, \bar{T}_i(k+\alpha) \qquad \text{for } i = 1,2,...,N \tag{35}$$

The α-MVA computes the throughput of a closed queueing network with exponentially distributed load-independent service times and non-integer number of jobs [11].

5.2. Computation of the Mean Number of Jobs

The given queueing network with finite station capacities and generally distributed service times with mean value $1/\mu_i$ and squared coefficient of variation $c_i^{\,2}$ is assumed to have infinite station capacities. The *MM* given in section 3, in its original form is applied in order to obtain the conditional throughputs, equations (6 and 7) which are then so-called *exponentiated (adjusted)* service rates. The network is then assumed to have finite capacities but with the adjusted new exponentially distributed service rates $\nu_i(k)$. This Markovian network with blocking can then be analyzed by the Method suggested by Akyildiz [3]. In the next section we briefly outline the computation of the mean number of jobs. The steps of the approach are outlined below:

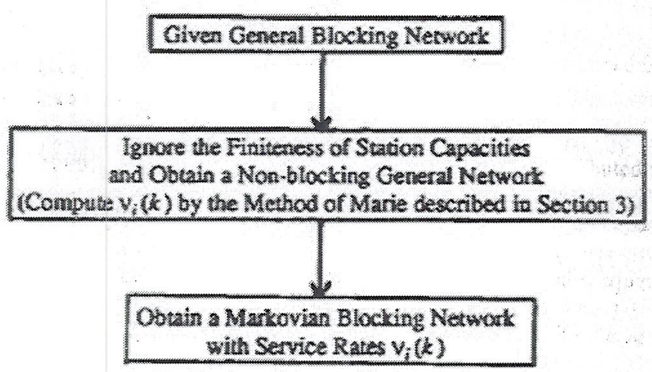

Figure 3.

The concept is based on the idea that the non-feasible states violating station capacities are normalized, i.e., the states are adapted to the allowed capacities of the stations. The normalization of the states exceeding station capacities is executed by using the following formula which indicates that if the number of jobs in a state of a station exceeds its capacity, $k_i^* > M_i$, then the job distribution is normalized until no violation of the capacity limits exists [3]:

$$k_j^* := \begin{cases} M_j & \text{if} \quad j = i \\ k_j^* + (k_i^* - M_i) \dfrac{e_j \, p_{ji}}{e_i \, (1 - p_{ii})} & \text{if} \quad j \neq i \end{cases} \quad \text{for } all \quad j = 1,2,\ldots,N. \tag{36}$$

The informal interpretation of the formula is: if the capacity of a station is reached in a state, the number of jobs in that station is set equal to its capacity M_j (first term) and distribute the remaining number of jobs in the predecessor stations according to the transition probabilities (second term).

Note that in case of networks with arbitrarily linked stations the transition probabilities p_{ij} in the normalization procedure, equation (36) causes the number of jobs k_i in a state to have non-integer values. All non-feasible states $\underline{k}^*$ of the queueing network are normalized by equation (36). The job distribution in each state is adapted to the capacity limit of each station in the blocking network:

$$h(\underline{k}^*) = (\underline{k})$$

where $\underline{k}$ is the normalized state for the blocking network. The function h transforms the non-feasible state ($\underline{k}^*$) to the feasible state ($\underline{k}$).

By the normalization procedure an equivalent state space structure is obtained for the blocking network. The only difference between the two state space structures is that the jobs are distributed according to the station capacities in the state space of the blocking network.

The *equilibrium probability* for the feasible states in blocking networks is computed by [3]:

$$p(\underline{k}) = \sum_{\underline{k}^* \text{ where } h(\underline{k}^*)=\underline{k}} p^*(\underline{k}^*) \tag{37}$$

where

$p^*(\underline{k}^*)$ represents the equilibrium state probability distribution of the network without station capacity restrictions computed by [12]:

$$p^*(\underline{k}^*) = \frac{1}{G(K)} \prod_{i=1}^{N} \left[\frac{e_i}{\prod_{j=1}^{k_i} \nu_i(k)} \right]^{k_i} \tag{38}$$

where

$$e_i = \sum_{j=1}^{N} e_j \, p_{ji}$$

is the mean number of visits that a job makes to station i and $G(K)$ is the normalization constant such that the sum of all probabilities sum up to unity.

$\nu_i(k)$ are the computed service rates from the *MM* given in section 3.

Performance measures such as the mean number of jobs $\bar{k}_i$ at the i-th station are computed by [3]:

$$\bar{k}_i(K) = \sum_{\sum_{j=1}^{N} k_j = K \ \& \ \underline{k}} h_i(\underline{k}^*) \quad p(\underline{k}) \quad \text{for } i = 1, \cdots, N \tag{39}$$

where the function $h_i(\underline{k}^*)$ is the i-th component of the function $h(\underline{k}^*)$.

The mean residence time are computed by Little's law, equation (18) where $\lambda(k)$'s are determined from the previous section.

6. Evaluation

All algorithms presented in this paper are implemented on a VAX 11/780 system. Several examples are executed for the validation of the algorithms. Each network model is analyzed by varying the number of jobs or the coefficient of variation values. The number of jobs is varied from 5 to 100, the number of stations from 2 to 6. The analytical results are compared with simulation executed by RESQ package [25].

As can easily be seen in the following chart, the throughput values of the 60 examples are in the range of 0 to 5% deviation.

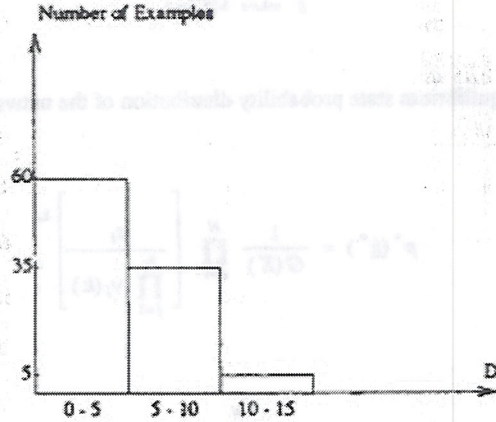

Figure 4.

For the computation of the mean number of jobs $\bar{k}_i$ the nonfeasible states are normalized. The normalization procedure takes fairly long run time for large networks. Compared to the throughput values, the errors in results for the mean number of jobs are higher. The reason for the high deviations is based on the combination of two approximate techniques for the analysis of blocking networks with general service time distributions. The method of Marie in its original form has its own deviations. Additionally, the state normalization technique [3] for blocking networks with exponential service time distributions has its own deviation. The accumulation of these deviations provide large errors (20 or 25%) for $\bar{k}_i$ as shown in the following chart. However, as it can also be observed in the chart, the 2/3 of the examples have deviations in the range of 0 to 10%.

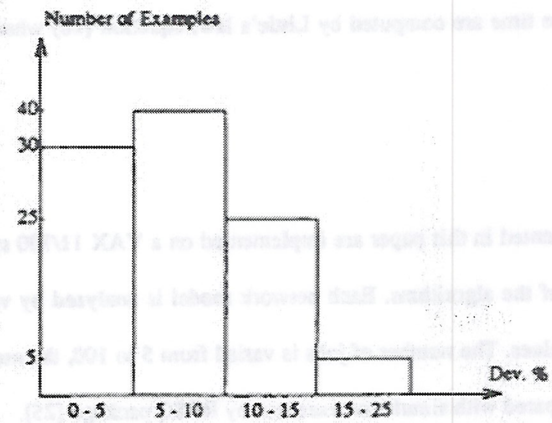

Figure 5.

Note that both methods (computation of throughput and computation of $\bar{k}_i$) showed a stable behavior. In the following six examples we give our approximate results and compare them with exact (in case of two station networks) as well as simulation results.

Example 1.

In the following example we vary the values for squared coefficient of variations and obtain the performance measures for two station blocking networks.

Station	B_i	μ_i
1	6	2
2	7	3

- $K = 10$ Jobs.

		Exact			Approximate					
c_1^2	c_2^2	λ	$\bar{k}_1$	$\bar{k}_2$	λ	$\delta(\%)$	$\bar{k}_1$	$\delta(\%)$	$\bar{k}_2$	$\delta(\%)$
0.5	0.5	1.9790	5.5198	4.4802	1.9761	0.15	5.4900	0.54	4.5100	0.66
0.5	1	1.9469	5.3887	4.6113	1.9469	-	5.3887	-	4.6113	-
0.5	4	1.7769	5.2071	4.7929	1.7992	1.24	5.2333	0.50	4.7667	0.55
0.5	9	1.6632	5.1821	4.8179	1.6906	1.67	5.2010	0.36	4.7990	0.39
0.5	100	1.5186	5.1833	4.8167	1.5273	0.57	5.1863	0.06	4.8137	0.06
1	0.5	1.9402	5.3382	4.6618	1.9402	-	5.3382	-	4.6618	-
1	1	1.9038	5.2601	4.7399	1.9038	-	5.2601	-	4.7399	-
1	4	1.7510	5.1463	4.8537	1.7510	-	5.1463	-	4.8537	-
1	9	1.6497	5.1290	4.8710	1.6497	-	5.1290	-	4.8710	-
1	100	1.5152	5.1322	4.8678	1.5152	-	5.1322	-	4.8678	-
100	0.5	1.6772	4.7461	5.2539	1.6791	0.11	4.7496	0.07	5.2505	0.06
100	1	1.6498	4.7942	5.2058	1.6498	-	4.7942	-	5.2058	-
100	4	1.5719	4.8910	5.1090	1.5682	0.24	4.8883	0.06	5.1117	0.05
100	9	1.5318	4.9325	5.0675	1.5267	0.33	4.9298	0.05	5.0701	0.05
100	100	1.4831	4.9771	5.0229	1.4789	0.28	4.9757	0.03	5.0243	0.03

where δ denotes the relative deviation of approximate results from exact. Note that the results in the Table are exact for cases where one station has exponential service time distribution (i.e., $c_i = 1$)

Example 2. $N = 3$ Stations; $K = 9$ Jobs;

Station	B_i	μ_i	p_{i1}	p_{i2}	p_{i3}
1	5	2	0	0.5	0.5
2	5	2.5	0.5	0	0.5
3	5	3	0.5	0.5	0

i) $c_1^2 = 0.5$; $c_2^2 = 0.5$; $c_3^2 = 0.5$;

	Simulation	Approximate	Deviation
λ	1.8643	1.7092	8.32
$\bar{k}_1$	4.0505	4.2217	4.23
$\bar{k}_2$	2.9075	2.7217	6.39
$\bar{k}_3$	2.0419	2.0566	0.72

ii) $c_1^2 = 100;\ c_2^2 = 100;\ c_3^2 = 100;$

	Simulation	Approximate	Deviation
λ	1.13O7	1.0807	4.42
$\bar{k}_1$	3.6094	3.4390	4.72
$\bar{k}_2$	2.9467	2.9377	0.31
$\bar{k}_3$	2.4439	2.6233	7.34

iii) $c_1^2 = 1;\ c_2^2 = 1;\ c_3^2 = 1;$

	Simulation	Approximate	Deviation
λ	1.7544	1.5885	9.46
$\bar{k}_1$	3.8454	4.0681	5.79
$\bar{k}_2$	2.9265	2.7908	4.64
$\bar{k}_3$	2.2280	2.1410	3.90

Example 3. $N = 4$ **Stations;** $K = 10$ **Jobs;**

Station	B_i	μ_i	p_{i1}	p_{i2}	p_{i3}	p_{i4}
1	5	3	0	0.6	0.4	0
2	7	2	0	0	0.7	0.3
3	4	4	0	0	0	1
4	5	4	0.8	0.2	0	0

i) $c_1^2 = 0.5;\ c_2^2 = 0.5;\ c_3^2 = 0.5;\ c_4^2 = 0.5$

	Simulation	Approximate	Deviation
λ	2.2703	2.1618	4.78
$\bar{k}_1$	2.1711	2.2159	2.06
$\bar{k}_2$	4.9950	4.7341	5.22
$\bar{k}_3$	0.9961	1.1102	11.45
$\bar{k}_4$	1.8378	1.9398	5.55

ii) $c_1^2 = 0.5;\ c_2^2 = 9;\ c_3^2 = 9;\ c_4^2 = 0.5$

	Simulation	Approximate	Deviation
λ	1.7243	1.6801	2.56
$\bar{k}_1$	2.5932	2.3427	9.66
$\bar{k}_2$	3.9388	4.4082	11.92
$\bar{k}_3$	1.2871	1.3095	1.74
$\bar{k}_4$	2.1801	1.9395	11.04

iii) $c_1^2 = 9;\ c_2^2 = 0.5;\ c_3^2 = 0.5;\ c_4^2 = 9$

	Simulation	Approximate	Deviation
λ	1.7239	1.6498	4.30
$\bar{k}_1$	2.1848	2.3519	7.65
$\bar{k}_2$	4.0418	4.0038	0.94
$\bar{k}_3$	1.1878	1.2004	1.06
$\bar{k}_4$	2.5856	2.4439	5.48

iv) $c_1^2 = 9;\ c_2^2 = 9;\ c_3^2 = 9;\ c_4^2 = 9$

	Simulation	Approximate	Deviation
λ	1.4417	1.3422	6.90
$\bar{k}_1$	2.3942	2.3897	0.19
$\bar{k}_2$	3.7142	3.9009	5.03
$\bar{k}_3$	1.3621	1.3606	0.11
$\bar{k}_4$	2.5295	2.3487	7.15

Example 4. $N = 5$ Stations;

Station	B_i	μ_i	c_i^2
1	5	2	0.9
2	30	1	9
3	10	1.3	0.5
4	20	1.5	4
5	7	0.9	16

i) $K = 50$ Jobs

	Simulation	Approximate	Deviation
λ	0.6784	0.6992	3.07
$\bar{k}_1$	1.8622	1.4738	20.86
$\bar{k}_2$	21.6968	18.1720	16.25
$\bar{k}_3$	6.3194	6.7092	6.20
$\bar{k}_4$	14.4997	16.6882	15.09
$\bar{k}_5$	5.6148	6.9579	23.92

ii) $K = 60$ Jobs

	Simulation	Approximate	Deviation
λ	0.6314	0.6676	5.73
$\bar{k}_1$	2.2822	1.6878	25
$\bar{k}_2$	27.1543	24.5067	9.25
$\bar{k}_3$	6.8353	7.7180	12.91
$\bar{k}_4$	17.9929	19.0892	6.1
$\bar{k}_5$	5.8264	6.9982	20

iii) $K = 70$ Jobs

	Simulation	Approximate	Deviation
λ	0.5067	0.5423	7.03
$\bar{k}_1$	4.3069	3.5158	18.32
$\bar{k}_2$	29.8812	29.9568	0.25
$\bar{k}_3$	9.4194	9.5399	1.28
$\bar{k}_4$	19.6738	19.9876	1.6
$\bar{k}_5$	6.6983	7.0000	4.5

References

1. I. F. Akyildiz, "Exact Product Form Solution for Queueing Networks with Blocking", *IEEE Transactions on Computers*, Vol. 1, January 1987, pp. 121-127.

2. I. F. Akyildiz, "On the Exact and Approximate Throughput Analysis of Closed Queueing Networks with Blocking", *to appear in IEEE Transactions on Software Engineering*.

3. I. F. Akyildiz, "Product Form Approximations for Closed Queueing Networks with Multiple Servers and Blocking", *to appear in IEEE Transactions on Computers*.

4. I. F. Akyildiz, "Mean Value Analysis for Closed Queueing Networks with Blocking", *to appear in IEEE Transactions on Software Engineering*.

5. I. F. Akyildiz and H. von Brand, "Duality in Open and Closed Queueing Networks with Rejection Blocking", *Technical Report of LSU*, TR-87-011, March 1987.

6. T. Altiok, "Approximate Analysis of Exponential Tandem Queues with Blocking", *European Journal of Operations Research*, Vol. 11, pp. 390-398, 1982.

7. T. Altiok and H. G. Perros, "Approximation Analysis of Arbitrary Configurations of Open Queueing Networks with Blocking", *AIIE Transactions*, March 1986.

8. S. Balsamo and G. Iazeolla, "Some Equivalence Properties for Queueing Networks with and without Blocking", *Proceedings of Performance 83 Conference*, editors A. K. Agrawala and S. Tripathi, North Holland Publ. Co., 1983, pp. 351-360.

9. O. I. Boxma and A. G. Konheim, "Approximate Analysis of Exponential Queueing Systems with Blocking," *Acta Informatica*, vol. 15, January 1981, pp. 19-66.

10. A. Brandwajn and Y. L. Jow, "An Approximation Method for Tandem Queues with Blocking", *Technical Report of the University of California at Santa Cruz*, 1985.

11. L. W. Dowdy and K. D. Gordon, "Algorithms for Nonintegral Degrees of Multiprogramming in Closed Queueing Networks", *Performance Evaluation*, Vol. 4, No. 1, February 1984, pp. 19-31.

12. W. J. Gordon and G. F. Newell, "Closed Queueing Systems with Exponential Servers", *Operations Research*, 15, Nr. 2, April 1967, pp. 266-277.

13. W. J. Gordon and G. F. Newell, "Cyclic Queueing Systems with Restricted Queues", *Operations Research*, 15, Nr. 2, April 1967, pp. 266-277.

14. A. Hordijk and N. van Dijk, "Networks of Queues with Blocking," *Proceedings*, 8th International Symposium on Computer Performance Modelling, Measurement, and Evaluation, Amsterdam, November 4-6, 1981.

15. A. G. Konheim and M. Reiser, "A Queueing Model with Finite Waiting Room and Blocking," *Journal of the ACM*, vol. 23, Number 2, April 1976, pp. 328-341.

16. R. Marie, "Calculating Equilibrium Probabilities for λ(n)/C/1-N Queues", ACM Sigmetrics Conference, Performance Evaluation Review, Vol. 9, No. 2, pp. 117-125, Summer 1980.

17. R. Onvural and H. G. Perros, "On Equivalences of Blocking Mechanisms in Queueing Networks with Blocking", *Operations Research Letters*, December 1986.

18. R. O. Onvural and H. G. Perros, "Some Exact Results for Closed Queueing Networks with Blocking", *Technical Report of North Carolina State University, 1987.*

19. H. G. Perros, "Queueing Networks with Blocking: A Bibliography", *ACM Sigmetrics Performance Evaluation Review*, August 1984.

20. H. G. Perros, "A Survey of Queueing Networks with Blocking - Part I", Technical Report of North Carolina State University, September 1985.

21. H. G. Perros and T. Altiok, "Approximate Analysis of Open Networks of Queues with Blocking: Tandem Configurations", *IEEE Transactions on Software Engineering*, Vol. SE-12, No. 3, March 1986, pp. 450-462.

22. H. G. Perros, A. A. Nilsson and Y. C. Liu, "Approximate Analysis of Product Form Type Queueing Networks with Blocking and Deadlock", *to appear in Performance Evaluation.*

23. B. Pittel, "Closed Exponential Networks of Queues with Saturation: The Jackson Type Stationary Distribution and Its Asymptotic Analysis," *Mathematics of Operations Research*, vol. 4, 1979, pp. 367-378.

24. M. Reiser ans S. S. Lavenberg, "Mean Value Analysis of Closed Multichain Queueing Networks", *Journal of the ACM*, Vol. 27, No. 2, April 1980, pp. 313-322.

25. C. H. Sauer and K. M. Chandy, "Computer Systems Performance Modeling" Prentice Hall, N.J., 1981.

26. R. Suri and G. W. Diehl, "A New Building Block for Performance Evaluation of Queueing Networks with Finite Buffers", *ACM Sigmetrics Conference Proceedings*, Cambridge, Mass., Aug. 1984, pp.134-142.

27. R. Suri and G. W. Diehl, "A Variable Buffer-Size Model and its Use in Analyzing Closed Queueing Networks with Blocking", *Management Science*, Vol. 32, No. 2, February 1986, pp. 206-225.

28. Y. Takahashi, H. Miyahara and T. Hasegawa, "An Approximation Method for Open Restricted Queueing Networks", *Operations Research*, Vol. 28, Nr. 3, May-June 1980, pp. 594-602

Acknowledgement: I am grateful to the referees for their careful review and constructive criticisms.