Revised

SOME EXACT RESULTS ON CLOSED QUEUEING NETWORKS WITH BLOCKING

Raif O. Onvural

H. G. Perros

Computer Science Department

and

Center for Communication and Signal Processing

North Carolina State University

Raleigh, NC, 27695-8206

CCSP-TR-87/16

October 1987

ABSTRACT

We obtain equivalencies between closed queueing networks with blocking with respect to buffer capacities and the number of customers in the network. These results can be used in approximations, in the buffer allocation problem as well as to explain the behavior of closed queueing networks with finite buffers. We also present exact efficient solutions for special cases.

Key words: blocking, closed queueing networks

1. Introduction

Closed queueing networks have proved to be useful in modeling computer systems, distributed systems, production systems and flexible manufacturing systems. In recent years, there has been a growing interest in the development of computational methods for the analysis of queueing networks with finite buffers. This is primarily due to a growing need to model actual systems which have finite capacity resources. An important feature of systems with finite buffers is that a server may become blocked when the capacity limitation of the destination queue is reached. Various blocking mechanisms have been considered in the literature so far. These blocking mechanisms arose out of various studies of real life systems. A discussion on these different blocking mechanisms can be found in Onvural and Perros [15].

Closed queueing networks with infinite buffer capacities, under certain restrictions, have been shown to have product-form steady state queue length distributions (see Gordon and Newell [8], Baskett, Chandy, Muntz, Palacios [4]). Efficient algorithms to calculate the performance measures of these networks have been developed based on this property. In general, closed queueing networks with finite buffers (hereafter referred to as CQN-B) could not be shown to have product form solutions. Hence, these algorithms are not applicable when limitations are imposed on buffer capacities. However, certain CQN-B have been reported in the literature as having product form solutions when: a) the routing matrix is reversible; b) the branching probability depends on the state of the originating node and the state of the destination node; c) the probability of blocking does not depend on the number of customers in the destination node but simply is constant; d) the service rate at each node is constant but there is zero probability that a queue is empty; (see Kelly [10], Le Ny[11], Akyildiz [1], and Hordijk and Van Dijk [9]). Also, a CQN-B has always a product form solution if it consists of two nodes (see, Akyildiz[1], Gordon and Newell[7], Perros [16]). A survey of closed queueing networks with blocking can be found in Onvural [13].

Since CQN-B, in general, could not be shown to have closed form solutions, most of the techniques that are used to analyze such networks are in the form of approximations, simulation and numerical techniques. In this paper, we obtain equivalencies between CQN-B with respect to buffer capacities and the number of customers in the network under two different blocking mechanisms. Two CQN-B are equivalent if they have the same rate matrix. These results are applicable to CQN-B under some other blocking mechanisms through equivalencies between different blocking mechanisms (see Onvural and Perros [15]). The equivalencies presented in this paper do not constitute a solution technique for CQN-B. Rather, they provide an insight into the behavior of their performance. We also show that in certain cases the CQN-B studied in this paper have a product form solution.

We will define the queueing network under study in the next section. Equivalencies between CQN-B with respect to buffer capacities and the number of customers in the network are given in Section 2.2. In most cases, we establish an equivalency between two CQN-B by first showing that both networks have the same number of states. We then define equivalencies between the states of the two networks so that the transition rates into and out of corresponding equivalent states are the same, thus showing that they have the same rate matrix. Some of the results obtained in Section 2.2 are generalized to another blocking mechanism in section 3. Special networks are investigated in Section 4. Finally, the conclusions are given in Section 5.

2. CQN-B Under Type 1 Blocking Mechanism

We will consider closed queueing networks consisting of N nodes and K customers. Each node consists of a single queue served by a server with an exponentially distributed service time with rate μ_i , i=1,..,N. Let B_i denote the capacity of node i including the service space in front of the server. A customer upon completion of its service at node i attempts to enter destination node j with probability p_{ij}, i=1,..,N; j=1,..,N. If at that moment, node j is full, the customer will be forced to wait in front of server i until a space would become available at node j. Server i remains blocked for this period of time, and it can not serve any other customer waiting in its queue. If more than one server is blocked by the same node, then these servers will get unblocked in a first-blocked-first-unblocked fashion. According to the classification given in Onvural and Perros [15] this blocking mechanism will be referred to as type 1 blocking mechanism.

Due to the blocking mechanism described above, and due to the fact that these N nodes are arbitrarily interconnected, it is possible that deadlocks may occur. For instance, assume that node i is blocked by node j. Now it is possible that a customer in node j may, upon completion of its service, choose to go to node i. If node i is at that time full, then a deadlock will occur. In this paper, it is assumed that deadlocks are detected immediately and resolved by instantaneously exchanging blocking units. This may violate the first-blocked-first-unblocked rule described above. For further details, see Onvural and Perros [14].

Now, let $n = \min_{i=1,..,N} \{B_i\}$. Clearly the number of customers in the network, K, is such that $1 \leq K \leq \sum_{i=1}^{N} B_i$. For $1 \leq K \leq n$, blocking does not occur and hence the network has a product form solution. This product form solution can be obtained by treating the queueing network as if the queue at each node has an infinite capacity. In particular, let $x_i = e_i/\mu_i$ be the relative utilization of node i where e_i is the mean number of visits per unit time a customer makes to the ith node and is a solution of the following system of linear equations:

$$e_i = \sum_{j=1}^{N} e_j p_{ji}, \qquad i=1,...,N \tag{1}$$

Let us assume for a moment that each node has an infinite capacity. Now, define $(i_1,..,i_j,..,i_N)$ to be the state of a closed queueing network with infinite buffer capacities where i_j is the number of customers at node j with $0 \leq i_j \leq K$ and $\sum_{j=1}^{N} i_j = K$. Let $\pi(i_1,..,i_N)$ be the steady state queue length distribution of the network. Then, $\pi(.)$ is the solution of the global balance equations :

$$\pi(i_1,..,i_N)\sum_{j=1}^{N}\sum_{k=1}^{N} p_{jk}\delta_j\mu_j = \sum_{j=1}^{N}\sum_{k=1}^{N} p_{jk}\delta_k\mu_j\pi(i_1,..,i_j+1,..,i_k-1,..,i_N) \qquad (2)$$

$$\sum_{(i_1,..,i_N)\in Z} \pi(i_1,..,i_N)=1$$

where

$$\delta_j = \begin{cases} 1 & \text{if } i_j > 0 \\ 0 & otherwise \end{cases}$$

and Z is the set of all feasible states.

The queue length distribution of the network defined in Section 2 for $1 \leq K \leq n$ is:

$$\pi(i_1,..,i_N)=G_K^{-1}\prod_{j=1}^{N} x_j^{i_j} \qquad (3)$$

where G_K^{-1} is a normalizing constant, chosen so that the distribution sums to unity (Gordon and Newell[8]).

When the number of customers is greater than the minimum buffer capacity, then blocking will occur. In this case, product form solutions are, in general, not available.

2.1. Aggregation Principle

In this section, we will introduce a concept called "aggregation principle", which is used in some of the results presented below.

Consider a node j of a CQN-B under type 1 blocking with buffer capacity B_j. If B_j=K-1, where K is the number of customers in the network, then there can be at most one node blocked by node j at any time. In this case, the blocked node has exactly one customer (which has completed its service), node j is full and all other nodes are empty. For states in which node l is blocked by node j, we will use the superscript *, i.e., i_l^*, to denote that node l is blocked while superscript l, i.e. $i_j=B_j^l$, denotes that node j is full and it is blocking node l. Hence, the state $(0,..,i_l^*=1,..,i_j=(K-1)^l,..,0)$ denotes that node l is blocked by node j. Then:

$$p_{lj}\mu_l \ P(0,..,i_l=1,..,i_j=(K-1),..,0)=\mu_j \ P(0,..,i_l^*=1,..,i_j=(K-1)^l,..,0) \quad (4)$$

where, $P(.)$ is the steady state queue length distribution of the CQN-B under consideration. Note that the right hand side of Eq. (4) is the rate out of the state in which node j is blocking node l and the left hand side is the rate into such state. Furthermore, define a state $(0,...,i_j=B_j^*+1,..,0)$ such that:

$$P(0,...,i_j=B_j^*+1,..,0)=\sum_{l=1}^{N} P(0,..,i_l^*=1,..,i_j=(K-1)^l,..,0) \quad (5)$$

That is, $P(0,...,i_j=B_j^*+1,..,0)$ is the probability that node j is blocking a node. Summing both sides of Eq. (4) over l and using equation (5) we have:

$$\sum_{l=1}^{N} p_{lj}\mu_l \; P(0,..,i_l=1,..,i_j=(K-1),..,0)=\mu_j \; P(0,...,i_j=B_j^*+1,...,0) \tag{6}$$

Now, let us increase the buffer capacity of node j by one (i.e. let B_j=K) while keeping all the other parameters the same. Then, there can not be any node blocked by node j. Writing down the equilibrium equation for the state (0,...,i_j=K,..,0) we have:

$$\sum_{l=1}^{N} p_{lj}\mu_l \; P(0,..,i_l=1,..,i_j=(K-1),..,0)=\mu_j \; P(0,..,i_j=K,..,0) \tag{7}$$

We note that the aggregated state in the CQN-B with B_j=K-1, (i.e. (0,...,$i_j=B_j^*+1$,..,0), defined by equation (5)) has the same behavior as the state (0,..,i_j=K,..,0) in the CQN with B_j=K, i.e., equation (7), while keeping all the other parameters the same. This property will be called the **aggregation principle.**

2.2. Some Properties of CQN-B Under Type 1 Blocking

We now proceed to investigate some of the properties of CQN-B under type 1 blocking. This is done mainly through a number of theorems presented in this section.

In the following theorem we show that when a node becomes blocked, the service space in front of the blocked server behaves like an additional buffer space for the blocking node. This is shown only for the case when there are no customers in the blocked node waiting for service during the blocking period.

THEOREM 1:Consider a CQN-B under type 1 blocking as described in section 2 with buffer capacities B_i. Let S={i: B_i=K} be the set of nodes with buffer capacities equal to the number of customers in the network. We shall refer to this network as CQN-B-1. Now, consider another CQN-B identical to CQN-B-1 except that the buffer capacity of node j, j$\in$ S, is reduced by one. We shall refer to this network as CQN-B-2. Then the two networks have the same rate matrix if all the states in which node j is blocking another node are aggregated into one state (i.e. applying the aggregation principle) in CQN-B-2.

Proof: Without loss of generality, let node j in CQN-B-1 be such that B_j=K. Let CQN-B-2 be identical to CQN-B-1, except that B_j=K-1. The only difference between these two networks is that in CQN-B-2, node j may cause blocking of a node while node j in CQN-B-1 will not cause any blocking. We note that there can be at most one node blocked by node j at any time in CQN-B-2. After the aggregation principle is applied to CQN-B-2, i.e. equation (5), the two networks have the same rate matrix.

The equivalency of the steady state queue length distributions can be summarized as follows: Let $P^{B_j=K}(.)$,$P^{B_j=K-1}(.)$ be the steady state probability distributions of CQN-B-1 and CQN-B-2 respectively. Then for the states in which node j is not blocking a node, we have

$$P^{B_j=K}(.)=P^{B_j=K-1}(.)$$

For the states in which node j is blocking node l in CQN-B-2 we have,

$$\sum_{l=1}^{N} P^{B_j = K-1}(0,..,i_l^{*}=1,..,B_j^{l},0,..,0) = P^{B_j = K}(0,..,i_j = B_j,..,0)$$

Note that, in the above proof we assume that there is only one node with B_j=K. If more than one node has the buffer capacity K in CQN-B-1, then the above discussion should be repeated for each of these nodes. □

Let us consider a CQN-B under type 1 blocking with K= $\min\limits_{i=1,..,N} (B_i)+1$ customers in it. In such a network blocking may arise when the node with the smallest buffer becomes full. The blocked node will contain exactly one customer which has completed its service and has attempted to enter the full node. Furthermore, all nodes other than the blocked node and the blocking node are empty. In the following theorem, we prove that a CQN-B with K= $\min\limits_{i=1,..,N} \{B_i\}+1$ customers has a product form solution. We note that this result is immediate from Theorem 1. However, in this special case we are able to present the form of product form solution of the blocking network without applying the aggregation principle.

THEOREM 2: Consider a CQN-B under type 1 blocking as described in section 2 with buffer capacities B_i. If the number of customers in the network, K, is equal to n+1, where n= $\min\limits_{i=1,..,N} \{B_i\}$, then the network has a product form solution.

Proof: Let $(i_1, i_2, .., i_N)$ be the state of a CQN-B where i_j is the number of customers at node j with $\sum_{j=1}^{N} i_j = n+1$, and $i_j \le B_j$, j=1,..,N. Also, without loss of generality, let j be the node with the smallest buffer and let $(0,..,i_l^*=1,..,i_j=B_j^l,..,0)$ denote a state of the network in which node l is blocked by node j. The steady state queue length distribution of the CQN-B, $P(i_1,..,i_N)$, is the solution of the following system:

$$P(i_1,..,i_N)\sum_{j=1}^{N}\sum_{k=1}^{N} p_{jk}\delta_j\mu_j = \sum_{j=1}^{N}\sum_{k=1}^{N} p_{jk}\delta_k\mu_j P(i_1,..,i_j+1,..,i_k-1,..,i_N), \quad (i_1,\ldots,i_N) \in Z^1$$

$$P(0,..,i_l=1,..,i_j=B_j,..,0)p_{lj}\mu_l = P(0,..,i_l^*=1,..,i_j=B_j^l,..,0)\mu_j, \quad (0,..,i_l^*=1,..,i_j=B_j^l,..,0) \in Z^2$$

$$\sum_{(i_1,..,i_N)\in Z} P(i_1,..,i_N)=1$$

where Z^1 and Z^2 is the set of all feasible states in which no node is blocked and in which one node is blocked respectively. $Z=Z^1 \cup Z^2$ is the set of all feasible states. Let,

$$P(i_1,..,i_N) = \begin{cases} \pi(i_1,..,i_N) & \text{if } i_j \le B_j,\ j=1,\ldots,N \\ \dfrac{p_{lj}e_l}{e_j(1-p_{jj})}\pi(0,..,i_j=K,..,0) & \text{if } i_l^*=1 \text{ and } i_j=B_j^l \end{cases} \tag{9}$$

where e_j is given by (1) and $\pi(.)$ is the solution given by (3) obtained by assuming that CQN-B has infinite buffers and K=n+1. Clearly, $\sum_{(i_1,..,i_N)\in Z} P(i_1,..,i_N)=1$.

By substituting these expressions for $P(.)$ into equation (8), it can be easily verified that the balance equations are satisfied. □

Under certain restrictions, the rate matrix of a CQN-B under type 1 blocking is identical for a range of values of K. In particular, let there be a node m in a CQN-B with $B_m \geq (\sum_{i=1}^{N} B_i) - B_m$, i.e., the buffer capacity of node m is greater than or equal to the remaining total capacity of the network. Then, the following theorem proves that this is a sufficient condition for the network to have the same rate matrix for a range of values of K.

THEOREM 3: Consider a CQN-B under type 1 blocking as described in section 2 with buffer capacities B_i. Let $B_m = \max_{i=1,..,N} (B_i)$. If $B_m \geq (\sum_{i=1}^{N} B_i) - B_m$ then the network has the same rate matrix for all K ∈ S, where S={ L : $(\sum_{i=1}^{N} B_i)$-B_m+1 ≤ L ≤ B_m }. Furthermore, the network with K=B_m+1 customers in it will have the same rate matrix as for all K∈ S if all the states in which node m is blocking another node are aggregated into one state. (i.e. if the aggregation principle is applied)

Proof: Consider all nodes other than node m. We first note that the distribution of K customers over the remaining N-1 nodes is independent of K. This is because K takes values greater than $(\sum_{i=1}^{N} B_i)$-B_m, the total capacity of the remaining nodes. (Obviously, this will not be the case if K < $(\sum_{i=1}^{N} B_i)$-B_m). In order to clarify this point, let us consider the states of the system for K > $(\sum_{i=1}^{N} B_i)$-B_m. There is a state in which all N-1 nodes are full (i.e. $(i_1=B_1,..,i_m=K-\sum_{\substack{j=1\\ j\neq m}}^{N} B_j,..,i_N=B_N)$) , and another state in which all N-1 nodes are empty (i.e. $(i_1=0,..,i_m=K,..,i_N=0)$). The remaining states reflect all the possible combinations in between with $i_m=K-\sum_{\substack{j=1\\ j\neq m}}^{N} i_j$. Now, let K^*=K+k, where $K^*\leq B_m$. Then, we can easily verify that for K^* we still obtain the same states, only i_m will contain k more customers. Similarly, it can be easily verified that the number of states where some node(s) is blocked is independent of K for K$\geq(\sum_{i=1}^{N} B_i)$-B_m. Hence, the state space of the N-1 nodes and the transitions between them are independent of K. Furthermore, for $(\sum_{i=1}^{N} B_i)$-B_m+1 ≤ K ≤ B_m node m can not block any node and can not be empty. Hence, transitions into and out of node m are independent of K, K∈S. Therefore, for all K∈S, we have the same rate matrix. That is, for K and K^* such that $(\sum_{i=1}^{N} B_i)$-B_m +1≤K≤$K^*\leq B_m$ we have $P^K(i_1,..,i_N)=P^{K^*}(i_1,..,i_m+K^*$-$K,..,i_N)$, where $P^K(.)$ and $P^{K^*}(.)$ are the steady state queue length distributions with K and K^*

customers in the network respectively.

To complete the proof, we need to show that the network with K= B_m+1 customers has the same rate matrix as the rate matrix of the network with K=B_m customers after all the states in which node m is blocking a node are aggregated into one state (i.e. equation (5)). However, this is immediate from Theorem 1. □

Now, consider a CQN-B with K customers such that the node with the maximum buffer capacity can not be empty. Then, the following theorem shows that if the buffer capacity of this node and the number of customers in a CQN-B are increased by the same amount, then the stochastic behavior of the network (i.e. its rate matrix) does not change.

THEOREM 4: Consider a CQN-B-1 under type 1 blocking as described in section 2 with buffer capacities B_i and let $B_m = \max_{i=1,\ldots,N} \{B_i\}$. Also, consider a CQN-B-2 with the same parameters as CQN-B-1 except that the buffer capacity of node m is increased to B_m^*. Now, let the number of customers in CQN-B-1, K, be such that node m can not be empty, i.e. K$\geq (\sum_{i=1}^{N} B_i)-B_m+1$. Then, CQN-B-1 with K customers has the same rate matrix as CQN-B-2 with K^*=K+B_m^*-B_m customers.

Proof: Let i_m and i_m^* be the number of customers at node m in CQN-B-1 and CQN-B-2 respectively. We first note that, K-$\sum_{i=1}^{N} B_i \leq i_m \leq min(B_m\ , K)$ in CQN-

B-1, and $K^{*}-\sum_{i=1}^{N} B_i \le i_m^{*} \le min(B_m^{*}, K^{*})$ in CQN-B-2. Hence, i_m and i_m^{*} can take the same number of (different) values in both networks. Furthermore, given the number of customers, i_m, in CQN-B-1 and $i_m^{*}= i_m+K^{*}$-K in CQN-B-2, node m in CQN-B-1 and CQN-B-2 can not be empty and the number of customers left to be distributed over remaining N-1 nodes are the same in both networks. Similarly, it can be argued that the number of states in which one or more nodes are blocked is the same in both networks. (We note that this is not the case if node m is allowed to be empty.) Hence, both networks have the same number of states. Furthermore, a state $(i_1, \ldots, i_m, \ldots, i_N)$ of CQN-B-1 is equivalent to the state $(i_1, \ldots, i_m+K^{*}-K,\ldots,i_N)$ of CQN-B-2 in the sense that both states have the same transition rates into and out of corresponding equivalent states. Hence,

$$P^1(i_1, \ldots, i_m, \ldots, i_N)=P^2(i_1, \ldots, i_m+K^{*}-K,\ldots,i_N)$$

where, $P^1(.)$ and $P^2(.)$ are the steady state queue length distributions of CQN-B-1 and CQN-B-2 respectively. □

COROLLARY 1: Consider a CQN-B-1 under type 1 blocking as described in section 2 with buffer capacities B_i. Also, consider a CQN-B-2 identical to CQN-B-1 but with buffer capacities C_i, i=1,..,N. Let K_1 and K_2 be the number of customers in CQN-B-1 and CQN-B-2 respectively such that no node can be empty. Then, the two networks have the same rate matrix if they have the same number of free spaces, i.e. $(\sum_{i=1}^{N} B_i)-K_1= (\sum_{i=1}^{N} C_i)-K_2$.

Proof: We first note that if no node is allowed to be empty then the node with the maximum buffer capacity can not be empty. Then, the proof follows from the successive applications of theorem 4 to the CQN-B with the larger number of customers in it. For presentation purposes, let K_1 and K_2 be the number of customers in CQN-B-1 and CQN-B-2 respectively such that no node can be empty (i.e. $K_1 \geq (\sum_{i=1}^{N} B_i) - \min_{i=1,..,N} B_i + 1$ and $K_2 \geq (\sum_{i=1}^{N} C_i) - \min_{i=1,..,N} C_i + 1)$. Furthermore, let l_1 and l_2 be the total number of free spaces in CQN-B-1 and CQN-B-2 respectively (i.e. $l_1 = (\sum_{i=1}^{N} B_i) - K_1$ and $l_2 = (\sum_{i=1}^{N} C_i) - K_2)$. If $l_1 = l_2 = l$ then the number of states in which no node is blocked is the same in both networks since it is equal to the number of ways in which l free spaces can be distributed over N nodes. Furthermore, if the number of customers in the two networks is such that no node can be empty, i.e. $l < \min_{i=1,..,N} \{B_i, C_i\}$, then the number of states in which one or more nodes are blocked is the same. Hence, the two networks have the same number of states. Now, let $d_i = B_i - C_i$, i=1,..,N. Then, a state $(i_1,..,i_N)$ of CQN-B-1 is equivalent to a state $(i_1 - d_1, ..., i_N - d_N)$ of CQN-B-2 in the sense that both states have the same transition rates into and out of corresponding equivalent states. Hence,

$$P^1(i_1,..,i_N) = P^2(i_1 - d_1, \ldots, i_N - d_N)$$

where $P^1(.)$ and $P^2(.)$ are the steady state queue length distributions of CQN-B-1 and CQN-B-2 respectively. □

We now proceed to demonstrate how some of the above theorems can be used. Let us first define λ_i(K) and L_i(K) to be the throughput and the mean queue length of node i respectively with K customers in the network. Furthermore, let λ(K) denote the throughput of the network. Then, $\lambda(K)e_i=\lambda(K)_i$, $i=1,...,N$, where e_i is the mean number of visits a customer makes to ith node per unit time and is given by equation (1). For presentation purpose, let us consider a three node CQN-B under type 1 blocking with buffer capacities 12,5,3. We shall refer to this network as CQN-B-1.

From theorem 3, CQN-B-1 has the same rate matrix for K=9,..,13. Hence, from the equivalency of the states of CQN-B-1 with K=10,11,12 customers, we have L_1(K)=1+L_1(K-1), K=10,11,12, while L_2(K) and L_3(K) is the same for K=9,..,12. Clearly, the throughput of the network is the same for all K=9,..,13. Now, let $\lambda_{\max}= \max_{K=1,..,M} \lambda(K)$ be the maximum throughput of the network w.r.t. K, where M is the total capacity of the network, i.e. M=B_1+8. If $B_1\geq 9$ then the value of $\lambda_{\max}$ will be the same for $9\leq K\leq B_1+1$. Now, consider a CQN-B identical to CQN-B-1 with buffer capacities 2,2,2. We shall refer to this network as CQN-B-2. From corollary 1, CQN-B-1 with 19 customers has the same rate matrix as CQN-B-2 with 5 customers. Then, $L_1^1(19)=10+L_1^2(5)$, $L_2^1(19)=3+L_2^2(5)$, and $L_3^1(19)=1+L_3^2(5)$, where $L_i^1(19)$ and $L_i^2(5)$ are the mean queue lengths of node i in CQN-B-1 and CQN-B-2 respectively.

Now, let us consider two three-node-CQN-B with identical parameters and buffer capacities 4,3,2 with 6 customers and 5,3,2 with 7 customers. Then, from theorem 4 the two networks have the same rate matrix, and, hence, $L_1^2(7)=1+L_1^1(6)$, $L_2^2(7)=L_2^1(6)$, $L_3^2(7)=L_3^1(6)$, and, $\lambda_i^1(7)=\lambda_i^2(6)$ for i=1,2,3.

Further results on CQN-B under type 1 blocking are given in section 4, where we investigate some special cases.

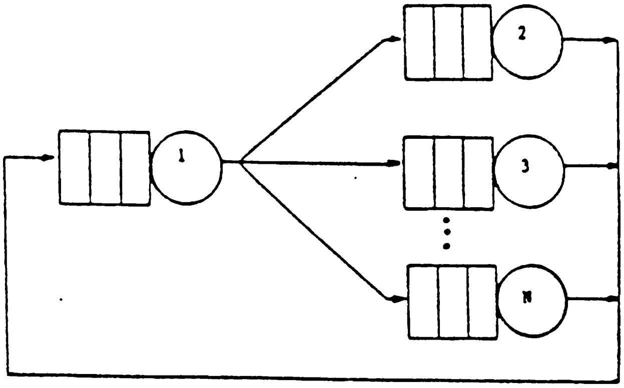


Figure 1: The Central Server Model

3. CQN-B Under Type 2.2 Blocking Mechanism

In this section, we will discuss the applicability of the results presented above for type 1 blocking to CQN-B under another blocking mechanism first introduced by Gordon and Newell [7]. According to the classification given in Onvural and Perros [15], this blocking mechanism will be referred to as type 2.2 blocking. In Type 2.2 blocking mechanism, a customer in queue i declares its destination queue j just before it starts its service. If queue j is full then the ith server becomes blocked. When a departure occurs from destination queue j, the ith server becomes unblocked and the customer begins receiving service. While the server is blocked, the position in front of the server is occupied by the customer.

This blocking mechanism for closed queueing network with finite buffers, arbitrary topology, and any number of customers in it, is not always well defined. This is because, in this general case deadlocks may occur. These deadlocks can not be resolved without violating the rules of type 2.2 blocking. For example, let us consider the central server model shown in figure 1.

Let B_i be the buffer capacity of node i, i=1,2,3. Let p_{1i}, i=2,3 be the branching probabilities with $p_{11}=0$. All service times are assumed to be exponentially distributed. For simplicity, let $B_i=2$, i=1,2,3, and K=4. Now let us consider the state (1,2,1) where the customer in the first node is blocked (as it wants to join queue 2 upon completion of its service), and queue 2 is full and its server is busy serving. If the customer at queue 3 completes its service first, the system will shift to the state (2,2,0), where the destination node of queue 2 is now full. This will cause the service at queue 2 to be suspended and the server to become blocked. Neither server 1 nor server 2 can start service because its destination node is full. This deadlock can not be resolved without violating the rules of type 2.2 blocking mechanism. For instance, it could be resolved by forcing the blocked customer at node 1 to change destination node or by temporarily switching to blocking type 1. In view of the above discussion, we will only study CQN-B under type 2.2 blocking such that deadlocks can not occur. Akyildiz and Kundu [3] showed that a CQN-B is deadlock-free if and only if for each cycle C in the network the following condition holds:

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

That is, the number of customers in the network should be less than the total capacity of each cycle in the network. Hereafter, we will refer to CQN-B which satisfy the above condition as **deadlock-free** CQN-B.

Gordon and Newell [7] introduced the concept of holes in cyclic networks under type 2.2 blocking. Let $M = \sum_{i=1}^{N} B_i$ be the capacity of the network. They proved that the network has a product form solution with M-l customers, where $1 \leq l < \min_{i=1,..,N} B_i$. This result also presents an equivalency between two cyclic networks with the same number of free spaces similar to theorem 3 given for type 1 blocking.

The following theorem is an extension of theorem 3 to type 2.2 blocking.

THEOREM 5: Consider a deadlock-free CQN-B under type 2.2 blocking with buffer capacities B_i and let $M = \sum_{i=1}^{N} B_i$ be the total capacity of the network. Now, let $B_m = \max_{i=1,..,N} (B_i)$. Then, if $B_m \geq M - B_m$, the network has the same rate matrix for all K $\in$ S, where S={ L : M-B_m $\leq$ L $\leq$ B_m }.

We note that the set, S, given above in theorem 5 is different than the set given in theorem 3. In theorem 3, the equivalency of the network with K=B_m

and $K=B_m+1$ customers is obtained using theorem 1. This theorem is not applicable to networks under type 2.2 blocking. Furthermore, it can be easily shown that a network under type 2.2 blocking does not have the same rate matrix with $K=B_m$ and $K=B_m+1$ customers. Also, for networks under type 2.2 blocking, it is possible that node m may be empty for the values in the set S. This is not the case for networks under type 1 blocking. This is due to the difference in the definitions of two blocking mechanisms. In type 1, blocking occurs after service completion while, in type 2.2, blocking occurs before service starts. In type 1 blocking, although the number of states in which no node is blocked is the same with $K=M-B_m$ and $K=M-B_m+1$ customers, the state $(i_1=B_1,..,i_m=0,..,i_N=B_N)$ of the network with $K=M-B_m$ customers is not equivalent to the state $(i_1=0,..,i_m=1,..,i_N=0)$, since the total rate out of these two states are not the same. Furthermore, it can be easily shown that two networks have different number of states. In type 2.2 blocking, the total rate out of the above two states are the same since there can not be any departure from node m in either case. Given these differences, theorem 5 can be proved following similar arguments as in theorem 3.

Finally, we note that theorem 4 given above for type 1 blocking mechanism is also applicable to deadlock-free CQN-B under type 2.2 blocking mechanism.

Further results on CQN-B under type 2.2 blocking are given in the following section.

4. Special Cases

In this section, we study special cases including symmetric queues. With some restrictions on the parameters of CQN-B, we will present three cases which have a product form solution. For symmetric queues, we present an algorithm to calculate the exact steady state queue length distribution on a reduced state space.

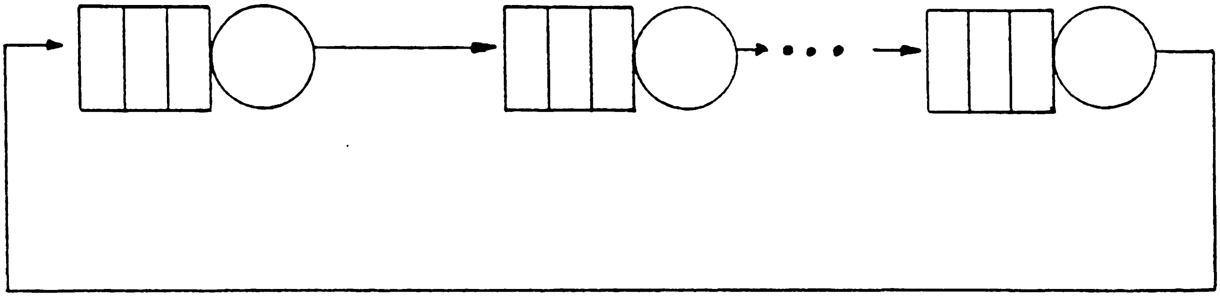
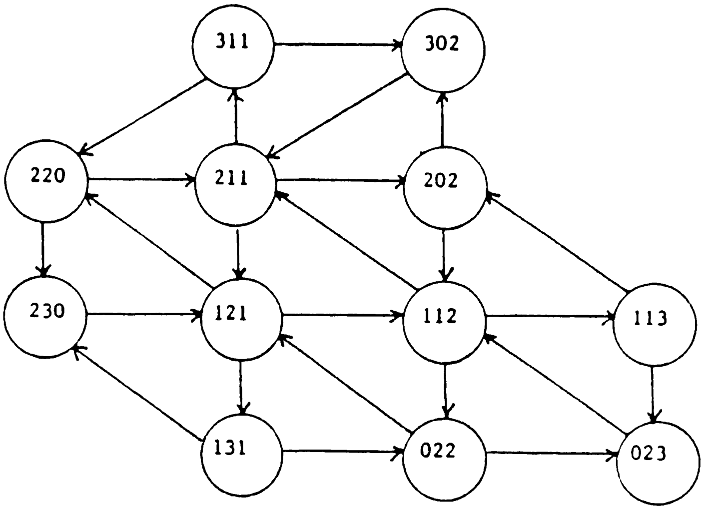


Figure 2: A Cyclic Network

4.1. Cyclic Networks-Symmetric Queues

A cyclic network is a closed queueing network consisting of queues in tandem as shown in Figure 2. Now, let us assume that $B_i = B < \infty$, $\mu_i = \mu$, $i=1,..,N$, and type 1 blocking. All service times are assumed to be distributed exponentially. Then, $1 \le K \le NB$. The algorithm given in this section utilizes an aggregate state space obtained from the original state space of the network after it is reduced by a factor of N. For presentation purposes, consider a cyclic network with $B=2$, $K=4$ and $N=3$. The state space has the following structure with all transition rates equal to μ.



Solving the system numerically (see Perros, Nilsson and Liu [11]), we have:

$$P(2,2,0)=P(0,2,2)=P(2,0,2)=0.071429$$

$$P(2,3,0)=P(0,2,3)=P(3,0,2)=0.11905$$

$$P(2,1,1)=P(1,2,1)=P(1,1,2)=0.095238$$

$$P(3,1,1)=P(1,3,1)=P(1,1,3)=0.047619$$

This result is not surprising seeing that the nodes are indistinguishable. In view of this, let us define the following classes, where a state is a member of a class if that state has the same steady state probability as all the other states in the same class.

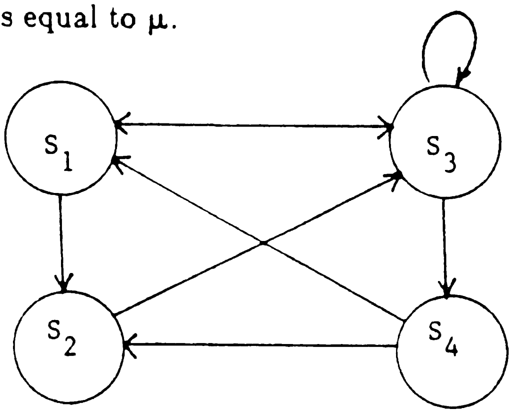
$$S_1=\{(2,2,0),(0,2,2),(2,0,2)\}$$

$$S_2=\{(2,3,0),(0,2,3),(3,0,2)\}$$

$$S_3=\{(2,1,1),(1,2,1),(1,1,2)\}$$

$$S_4=\{(3,1,1),(1,3,1),(1,1,3)\}$$

Then, we have the following state space structure for these equivalence classes with all transition rates equal to μ.



Solving this system numerically, we have:

$P(S_1)=0.214287$; $P(S_2)=0.35715$; $P(S_3)=0.28571$; $P(S_4)=0.142857$

We note that, $P(S_i)= \sum_{(i_1,i_2,i_3)\in S_i} P(i_1,i_2,i_3)$, i=1,..,4. Hence, to solve the original network, we can form the equivalence classes, S_i, create the rate matrix for these classes and solve the system. Then we can obtain the queue length distribution of the original network. The following algorithm summarizes this procedure.

ALGORITHM

1.Generate the equivalence classes, S_i, and set up the rate matrix.

2.Solve the system to obtain $P(S_i)$.

3.Calculate the normalizing constant, G_K, for the original network as follows:

$$G_K = \sum_{i=1}^{S} R_i P(S_i)$$

where S is the number of equivalence classes and R_i is the number of states in equivalence class i.

4.$P(i_1,..,i_N)=G_K^{-1}P(S_i)$ where $(i_1,...,i_N) \in S_i$

The equivalence class of a state $(i_1,...,i_N)$, S_i, is the cyclic permutation of customers in that state and it can be determined easily.

Finally, we note that this algorithm is also applicable to cyclic networks under type 2.2 blocking without any modification.

4.2. Central Server Model-Symmetric Queues

A central server model is shown in figure 1. Each node has a single server, buffer capacity B_i, and service times are exponentially distributed with rate μ_i. p_{1i} is the probability that a customer upon completion of its service at node 1 attempts to go to node i, i=2,...,N.

Let us assume that B_i=B < ∞ and $\mu_i=\mu$, i=2,..,N and consider the network under type 1 blocking. We note that, in general, $\mu_1 \neq \mu$ and $B_1 \neq$B. Furthermore, let p_{1i}=1/(N-1), i=2,..,N. Then nodes 2 through N are indistinguishable and hence the algorithm given for cyclic networks above can be applied to nodes 2 through N.

4.3: Merge Configuration-Symmetric Queues

Let us now consider the central server model under type 1 blocking with $B_1 < \infty$ and $B_i=\infty$, and p_{1i}=p=1/(N-1), i=2,..,N. With these parameters, the network has the local balance property after the service rates are modified as follows:

$$\mu_i = \begin{cases} \mu_i & i=2,..,N \\ \mu_1 & i=1 \text{ and } \textit{no node is blocked} \\ \dfrac{\mu_1}{p} & i=1 \text{ and } \textit{some node(s) are blocked} \end{cases}$$

Then, we have:

$$p\mu_1 P(i_1,...,i_j,...,i_N)=\mu_j\ P(i_1-1,..,i_j+1,..,i_N) \quad \text{if } \textit{no node is blocked}$$

$$\mu_1 P(i_1+^*,...,i_N)=\mu_j P(i_1,...,i_N) \qquad \text{if } \textit{some node(s) are blocked.}$$

where i_1+^* denotes that node 1 is blocking a node.

Hence, one can use these two sets of equations to obtain the joint queue length distributions of such networks instead of working with the rate matrix .

5.4. Type 2^* Blocking Mechanism

In this section, we will consider a modification of type 2.2 blocking mechanism defined in section 4. In type 2.2 blocking mechanism, a customer chooses its destination node before it starts receiving service and if the destination node is full then the server can not start serving the customer. Let us modify this blocking mechanism as follows. A customer in queue i checks its all possible destination nodes before starting its service. If any one of its possible destinations is full then the ith server becomes blocked. We will refer to this blocking mechanism as type

2^* blocking. We note that the only difference between type 2^* and type 2.2 blocking mechanisms is that in type 2.2 blocking, a server gets blocked if the destination node of its customer is full while in type 2^* blocking a server is blocked if any one of the destination node is full. The following result follows from Kelly [10].

THEOREM 6: Consider a reversible CQN as described in section 2 with infinite buffer capacities. Then a deadlock-free CQN-B derived from this network by imposing capacity restrictions on the buffer capacities is also reversible under type 2^* blocking, and, hence, it has a product form solution.

Proof: Kelly [10] showed that if a network is reversible with state space S then any truncated process with state space A, $A \subseteq S$ is also reversible. The form of the solution of the truncated process is the same as the original process normalized over the state space A. A reversible CQN-B under type 2^* blocking is a truncated process of a reversible network with infinite buffer capacities. Furthermore, since the network with infinite buffer capacities has a product-form solution, CQN-B under type 2^* blocking has a product form solution. The form of the product-form solution is given in equation (3) normalized over the state space of the network. □

5. Conclusions

We presented some equivalencies between CQN-B under two types of blocking mechanisms. These results are also applicable to CQN-B under some other blocking mechanisms through equivalencies between different blocking mechanisms (see, Onvural [12] and Balsamo, Persone, and Iazeolla [5]).

Theorem 2 presents a product-form solution for networks under type 1 blocking when the number of customers in the network is equal to the minimum buffer capacity plus one. It is also shown that the merge configuration under type 1 blocking has the local balance property with modified service rates and, hence, it possesses a product-form solution. We also defined a modification of type 2.2 blocking, i.e. 2^* blocking, which has a product form solution. For some special cases, type 2.2 blocking reduces to type 2^* blocking and, hence, it has a product form solution. For example, consider the central server model with $B_1<\infty$ and $B_i=\infty$, i=2,...,N. Let, p_{1i}, i=2,3,..,N be the routing probabilities, and let μ_i, i=1,..,N, be the service rates where the service time at each node is distributed exponentially. Then, it can be shown that this network under type 2.2 blocking behaves like a network under type 2^* blocking, and hence it has a product-form solution.

For symmetrical CQN-B we presented a numerical solution for calculating the exact steady state queue length distributions on a reduced state space.

Finally, theorems 1 and 3 to 5 do not present a solution technique for CQN-B. Instead, they provide an insight into the behavior of their performance. Some of these theorems have been used to construct an approximation algorithm for estimating the throughput of cyclic queueing networks under type 1 or type 2.2 blocking. For further details see Onvural and Perros [14]. These theorems may also be used in the buffer allocation problem as they provide some constraints on the buffer capacities. Finally, in some of the approximation algorithms developed for CQN-B, the analysis is based on a mapping from the states of the network under study to the states of a similar network but with infinite buffer capacities (see Akyildiz [1,2], Diehl [6]). If some of these theorems are applicable, then we can find an equivalent network with smaller buffer capacities and/or fewer customers in the network which may be easier to analyze, or, for which the approximation may yield better values.

REFERENCES

[1] **Akyildiz, I.F.**, "On the Exact and Approximate Throughput Analysis of Closed Queueing Networks with Blocking", to appear in IEEE Trans. Soft. Eng.

[2] **Akyildiz, I.F.**, "Mean Value Analysis for Blocking Queueing Networks", to appear in IEEE Trans. Soft. Eng.

[3] **Akyildiz, I.F. and Kundu, S.**, "Buffer Allocation in Deadlock Free Blocking", CS Dept, (1986), Lousiana State University

[4] **Baskett, F., Chandy, K.M., Muntz, R.R. and Palacios, J.**, "Open, Closed and Mixed Networks of Queues with Different Classes of Customers", J. of ACM, 22, 2, pp 249-260, (1985)

[5] **Balsamo, S. and Persone V. De Nitto, and Iazeolla, G.**, "Some Equivalencies of Blocking Mechanisms in Queueing Networks with Finite Capacity", Manuscript, Dipartimento di Informatica, Universita di Pisa, Pisa Italy, (1986)

[6] **Diehl, G.W.**, "A Buffer Equivalency Decomposition Approach to Finite Buffer Queueing Networks", Ph.D. Thesis, Eng. Sci., Harvard University, (1984)

[7] **Gordon, W.J. and Newell, G.F.**, "Cyclic Queueing Systems with Restricted Length Queues", Oper. Res., 15, 266-278, (1967)

[8] **Gordon, W.J. and Newell, G.F.**, "Closed Queueing Systems with Exponential Servers", Oper. Res., 15, 254-265, (1967)

[9] **Hordijk, A. and Van Dijk, N.**, "Networks of Queues with Blocking", Performance'81, Klystra (Ed.), 51-65, North Holland, (1981)

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

[11] **Le Ny, L.M.**, "Etude Analytique de Reseaux de Files d'Attante Multiclasses a Routage Variables", R.A.I.R.O., 14, 331-347, (1980)

[12] **Onvural, R.O.**, "Closed Queueing Networks with Finite Buffers", Ph.D. Thesis, CSE/OR, NCSU, (1987)

[13] **Onvural, R.O.**, "A Survey of Closed Queueing Networks with Blocking", CS Dept., NCSU, (1987)

[14] **Onvural, R.O. and Perros, H.G.**, "Throughput Analysis of Closed Queueing Networks with Blocking", CS Dept., NCSU

[15] **Onvural, R.O. and Perros, H.G.**, "On Equivalencies of Blocking Mechanisms in Queueing Networks with Blocking", OR Letters, 5, No:6, 293-298, (1986)

[16] **Perros, H.G.**, "A Survey of Queueing Networks with Blocking (Part I)", Dept. of CS, 86-04, NC State Univ., (1986)