REFERENCES

[1] M. J. Atallah and S. R. Kosaraju, "Graph problems on a mesh-connected processor array," in *Proc. 14th Ann. ACM Symp. Theory Comput.*, 1982, pp. 345-353.

[2] A. Bojanczyk, R. Brent, and H. T. Kung, "Numerically stable solution of dense systems of linear equations using mesh-connected processors," *SIAM J. Sci. Stat. Comput.*, vol. 5, pp. 95-104, 1984.

[3] C. F. Ilse Ipsen, "Stable matrix computations in VLSI," Ph.D. dissertation, Dep. Comput. Sci., Pennsylvania State Univ., University Park, PA, Tech. Rep. CS-83-17, 1983.

[4] H. T. Kung and C. E. Leiserson, "Systolic arrays (for VLSI)," in *Sparse Matrix Proceedings 1978*, I. S. Duff and G. W. Stewart, Eds. Philadelphia, PA: SIAM, pp. 256-282, 1979.

[5] J. D. Ullman, *Computational Aspects of VLSI.* Rockville, MD: Computer Science, 1984.

Exact Product Form Solution for Queueing Networks with Blocking

I. F. AKYILDIZ

Abstract—**This work investigates closed queueing networks with blocking composed of two stations with multiple servers. Blocking occurs when a job wanting to enter a full station is forced to remain in its source station, thus blocking the source station until room is available at the destination station. This type of blocking is known as classical blocking. We show that, for a two-station closed queueing network with blocking, there exists an equivalent nonblocking network with the same state space structure. Utilizing this concept, we demonstrate that two-station closed queueing networks with blocking have product form solutions.**

Index Terms—**Blocking, equilibrium state probabilities, normalization constant, performance analysis, performance measures, queueing networks, state space transformations.**

I. INTRODUCTION

The basic results of product form networks are given in Baskett, Chandy, Muntz and Palacios [4]. They show that queueing networks with different classes of jobs, exponential and nonexponential service time distributions, and different queueing disciplines (FCFS, RR-PS or LCFS-PR) have product form solution. This was a remarkable result for the queueing network theory. The term "product form" means that the equilibrium state probabilities can be expressed as a product of terms for each queue in the network. The product form networks, also known as BCMP or separable networks, are based on the assumption that all stations have infinite capacities. If the stations have finite capacity, blocking can occur in the network. Since blocking causes interdependencies between stations, blocking queueing networks cannot be analyzed by existing product form algorithms.

Formally we distinguish between three types of blocking: *classical*, *rejection*, and *service* blocking.

In the first case, blocking occurs when a job completing service at station i cannot proceed to station j because station j is full. The job is forced to wait in the station i's server until it is allowed to enter the destination station j. The station i's server stops processing until station j releases a job [1], [2], [11], [13], [15], [20]. In the second case, blocking occurs when a job completing service at station i attempts to join destination station j. If station j is full at that moment, the job is refused. The rejected job goes with a certain probability (called rejection probability) back to the station i's server and receives a new service. This is repeated until some job completes a service at station j and a place becomes available [3], [9], [16], [22]. In the third 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 ith server becomes blocked, i.e., it cannot serve jobs. When a departure occurs from destination station j, the ith server becomes unblocked and the job begins receiving service [5], [8], [10], [19].

In recent years there has been a growing interest in the development of computational methods to analyze queueing networks with blocking. The interest developed primarily from the realization that these models are useful in the study of subsystem behavior in computers and communication networks, in addition to providing detailed descriptions of several computer-related applications such as flexible manufacturing systems.

Several investigators in recent years have published results on queueing networks with classical, rejection and service blocking. A bibliography concerning queueing network models with blocking is given by Perros [14]. Formal comparisons between the distinct classes of blocking have been carried out by Onvural and Perros [12].

In this paper we introduce the state space transformation concept which provides exact results for two-station closed queueing networks with classical blocking.

II. PRODUCT FORM SOLUTION

We will consider closed queueing networks with $N = 2$ stations and K single-class jobs. Each station consists of a single queue served by ($m_i \geq 1$), servers each with exponentially distributed service time with mean value $1/\mu_i$ (for $i = 1, 2$). The service discipline in each station is first-come, first-served. Each station has a fixed finite capacity M_i where M_i = (queue capacity + m_i), (for $i = 1, 2$). Cases in which the stations can have infinite capacity are also allowed. Any station whose capacity exceeds the total number of jobs in the network can be considered to have infinite capacity. It is obvious that the total number of jobs K must be smaller than the sum of the station capacities, that is,

$$K < \sum_{i=1}^{2} M_i.$$

A job which is serviced by the ith station proceeds to the jth station with probability p_{ij}, (for $i, j = 1, 2$), if the jth station is not full. That is, if the number of jobs in the jth station k_j is less or equal to M_j for $i, j = 1, 2$. Otherwise, the job is blocked in the ith station until a job in the jth station has completed its servicing and a place becomes available. Note that the case $i = j$ is allowed. A station can have a transition back to itself, which will be shown by an example in Section IV. The queueing network with the above assumptions is known as the network with classical blocking which we will investigate in this work.

As is generally known, the following binomial coefficient formula is valid for closed queueing networks without station capacity limits. It indicates the number of possible ways that K jobs can be distributed into N stations.

$$Z = \binom{N+K-1}{N-1} \tag{1}$$

where Z is the total number of states in a closed queueing network.

For queueing networks with two stations (1) is simplified to

$$Z = K + 1. \tag{2}$$

Structure for the state space of the two-station network is illustrated in Fig. 1.

The state (k, n) denotes that k jobs are in the first station and n jobs are in the second station. The transition rates from one state to another are equal to the service rates μ_i multiplied by the number of servers m_i in each station. Fig. 1 shows that there must be at least ($K - m_i$) waiting places in each queue (m_i jobs can be in service) to ensure that all states are feasible. Since each station in the network

Manuscript received May 13, 1985; revised September 11, 1985.

The author is with the Department of Computer Science, Louisiana State University, Baton Rouge, LA 70803.

IEEE Log Number 8611265.

has a limited capacity, it is clear that all states in Fig. 1 cannot be feasible. The feasible states for blocking networks are obtained by realizing that the number k_i of jobs in the ith station may not exceed the station capacity M_i, ($k_i \leq M_i$).

We define s_i as the number of servers that can be blocked at the ith station in a given state.

$$s_i = \begin{cases} m_i & \text{if } K > M_j + m_i \\ K - M_j & \text{if } K \leq M_j + m_i \end{cases} \quad \text{for } i = 1,\ 2. \tag{3}$$

Note that the first case of (3) states that all the servers of the ith station can be blocked while the second case states that some servers of the ith station can be empty. The s_i 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 in the network and that the state is a blocking state, i.e., the job resides in the source station.

In this fashion the complete state space for the blocking network is obtained (Fig. 2).

Obviously, Fig. 2 is a subgraph of Fig. 1. Except for the blocking states denoted in Fig. 2 by a "*," all the other states violating the station capacities are nonfeasible and are cancelled. From the reduced state space, Fig. 2, we obtain the number of states Z' of the blocking queueing network, which is the sum of the number of feasible states and blocking states. Formally,

$$Z' = \min\{K,\ M_1 + m_2\} + \min\{K,\ M_2 + m_1\} - K + 1. \tag{4}$$

We can now state the following theorem.

Theorem: For a two-station closed queueing network with *classical blocking* there exists an equivalent two-station closed queueing network *without blocking* (i.e., without station capacity limits) with the same structure. The equilibrium state probabilities $p(k_1,\ k_2)$ for the blocking network can be obtained from the equilibrium state probabilities of the equivalent nonblocking network.

$$p(k_1,\ k_2) = \frac{1}{G(\hat{K})} \prod_{i=1}^{2} \frac{x_i^{k_i - K + M_j + s_j}}{\beta_i(k_i - K + M_j + s_j)} \quad \text{for } i,\ j = 1,\ 2;\ i \neq j \tag{5}$$

where the number of jobs k_i is defined in the following range.

$$\{K - M_j - s_i\} \leq k_i \leq \{M_i + m_j\}.$$

For the k_i values outside of this range, the equilibrium state probabilities $p(k_1,\ k_2)$ are zero.

The other quantities in (5) are defined as follows:

$K = \sum_{i=1}^{2} k_i$ represents the total number of jobs in the blocking network

$\hat{K} = \sum_{i=1}^{2} \hat{k}_i$ represents the total number of jobs in the equivalent nonblocking network

$G(\hat{K})$ is the normalization constant so that all probabilities sum up to unity

$x_i = (e_i/\mu_i)$ is the relative utilization of the ith station (for i = 1, 2) where e_i is the expected number of visits a job makes to the ith station (for i = 1, 2):

$$e_i = \sum_{j=1}^{2} e_j p_{ji} \quad \text{for } i = 1,\ 2. \tag{6}$$

As there are (N − 1) linear independent equations in the system it is assumed that e_1 = 1. The function $\beta_i(k_i)$ is defined as

$$\beta_i(k_i) = \begin{cases} k_i! & \text{if } k \leq m_i \\ m_i!\, m_i^{k_i} - m_i & \text{if } k > m_i. \end{cases}$$

Proof: As we stated in the theorem, the equivalent nonblocking network has the same structure as the given blocking network. In other words the equivalent nonblocking network has the same number of stations $\hat{N} = N = 2$; the same exponentially distributed mean service time with rate $\hat{\mu}_i = \mu_i$; the same number of servers $\hat{m}_i = m_i$; the same transition probabilities $\hat{p}_{ij} = p_{ij}$; and, accordingly, the same mean number of visits a job makes to the ith station $\hat{e}_i = e_i$, (6). The only difference is that the station capacities are unlimited, hence no blocking occurs and the total number of jobs in both systems is not equal, i.e., $\hat{K} \neq K$.

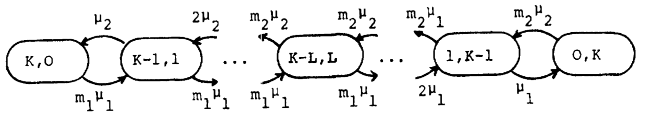

Fig. 1.

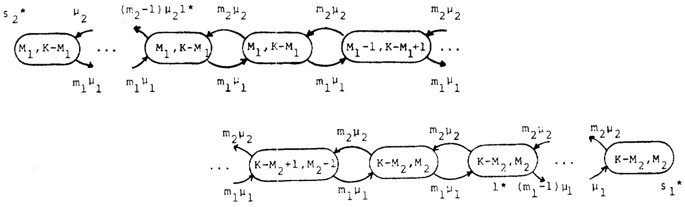

Fig. 2.

Since the nonblocking network satisfies local-balance, it has the product form solution

$$p(\hat{k}_1,\ \hat{k}_2) = \frac{1}{G(\hat{K})} \prod_{i=1}^{2} \frac{\hat{x}_i^{\hat{k}_i}}{\hat{\beta}_i(\hat{k}_i)}. \tag{7}$$

By considering the above equalities of both networks we get

$$p(\hat{k}_1,\ \hat{k}_2) = \frac{1}{G(\hat{K})} \prod_{i=1}^{2} \frac{x_i^{\hat{k}_i}}{\beta_i(\hat{k}_i)}. \tag{8}$$

The goal is to find the total number of jobs $\hat{K}$ in this equivalent nonblocking network.

As shown in Fig. 2, the number of states Z is reduced to Z', (4), by considering the station capacities. From both (2) and (4) we arrive at the total number of jobs $\hat{K}$ in the equivalent two-station closed queueing network without blocking

$$\hat{K} = \min\{M_1 + m_2,\ K\} + \min\{M_2 + m_1,\ K\} - K. \tag{9}$$

The state space for the nonblocking network with $\hat{K}$ jobs is given in Fig. 3.

It is immediately seen that the state space for the equivalent nonblocking network with $\hat{K}$ jobs (Fig. 3) has exactly the same structure as the state space of the blocking network shown in Fig. 2. The number of states and the transitions between the states are identical. The Markov processes describing the evolution of the networks over time have the same structure up to an isomorphism of the state spaces. This isomorphism is compatible with the transition rates. This implies that the equilibrium state probability of the blocking network is equivalent to the equilibrium state probability of the nonblocking network

$$p(k_1,\ k_2) = p(\hat{k}_1,\ \hat{k}_2).$$

It is easily seen that the number of jobs $\hat{k}_i$ in the ith station of the equivalent nonblocking network in Fig. 3 corresponds to the (k_i − K + M_j + s_j) value in Fig. 2. In (8) we replace $\hat{k}_i$ by (k_i − K + M_j + s_j) and obtain (5). This completes the proof.

III. Performance Measures

We have shown that the equilibrium state probabilities for two station blocking networks have a product form solution, (see (5)).

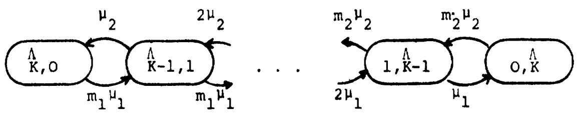

Fig. 3.

The normalization constant $G(\hat{K})$, which occurs in the solution can be determined by any product form network algorithm, e.g., convolution algorithm [6], [7] or LBANC technique [18] where the total number of jobs $\hat{K}$ in the equivalent nonblocking network is computed by (9).

Performance measures, such as throughput λ and utilization ρ, that are dependent on the equilibrium state probabilities and the service rates of each station can be computed directly from the transformed state space of Fig. 3. The throughput $\lambda_{NB}(\hat{K})$ of the equivalent nonblocking network is exactly equal to the throughput $\lambda_B(K)$ of the blocking network, i.e.,

$$\lambda_B(K)=\lambda_{NB}(\hat{K}) \tag{10}$$

where $\lambda_{NB}(\hat{K})$ can either be determined by

$$\lambda_{NB}(\hat{K})=\frac{G(\hat{K}-1)}{G(\hat{K})} \tag{11}$$

or by mean value analysis [17].

The *throughput* of each station can be obtained by using $\lambda_B(K)$ in the blocking network

$$\lambda_i(K)=e_i\lambda_B(K) \quad \text{for } i=1,\ 2. \tag{12}$$

The *mean number of jobs* in the ith station is given by the following well-known formula:

$$\overline{k_i}(K)=\sum_{n=1}^{K} np_i(n) \tag{13}$$

where $p_i(n)$ is the marginal probability that there are n jobs in the ith station. In two-station closed networks the following values are legal.

$$p_1(n)=p(n,\ K-n)=p_2(K-n) \quad \text{for } n=1,\ \cdots,\ K. \tag{14}$$

By using the equivalence of nonblocking and blocking networks (13) is modified to

$$\bar{k}_i(K)=\sum_{n=K-M_j}^{M_i} np_i(n)+\sum_{n=1}^{m_j} M_ip_i(M_i+n)$$
$$+\sum_{n=1}^{m_i}(K-M_j)p_i(K-M_j-n) \quad \text{for } i,\ j=1,\ 2;\ i\neq j. \tag{15}$$

An informal interpretation of this formula is that the right-hand side of the first term includes the feasible states while the second and third terms show the blocking states (see Fig. 2).

Other performance measures: the mean residence time (time spent by a job in queue, in service, and in the blocking phase) ($\overline{t_i}$); the mean queue length ($\overline{Q_i}$); the mean waiting time ($\overline{w_i}$); and the utilization (ρ_i) can be computed by the well-known product form network formulae [18]. Another performance measure that is important for blocking networks is the *blocking probability* of the ith station

$$P_{B_i}=\sum_{n=1}^{m_i} p_i(K-M_j-n) \tag{16}$$

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

IV. Example

In Section II we assumed that a station may have a transition back to itself. However, we did not explicitly cover this case as it would have complicated the presentation. In this example we examine this case and demonstrate that the state space is still one-dimensional. This implies that our theorem holds.

Consider a closed queueing network with $N = 2$ stations and $K = 10$ jobs. The stations have exponentially distributed service times with mean values $(1/\mu_1) = 2$ and $(1/\mu_2) = 0.9$, respectively, and FCFS service disciplines. The first station has a single server ($m_1 = 1$) and constant capacity $M_1 = 7$. The second station has $m_2 = 2$ servers and the constant capacity $M_2 = 5$ jobs. The transition probabilities are given as $p_{11} = 0.3$; $p_{12} = 0.7$; $p_{21} = 1$.

Using (2) we obtain the number of states for this network without considering the capacities: $Z = K + 1 = 11$. The state space for this network without capacity limits is shown in Fig. 4.

By considering the station capacities the substate space in Fig. 5 is obtained, containing the feasible states and the blocking states (denoted by $*$; obtained by (3)) for the blocking network.

Note that the following situation which could occur in a classical blocking network demands an explanation. Assume that the first station is full, its server is busy serving and the second station is blocked. Upon service completion at the first station, the departing job chooses to go back to the end of the first queue (with probability $p_{11} = 0.3$), the station it just left. The question that arises is which of the two jobs will enter the queue, the job which is blocked in the second station or the job just serviced by the first station which wants to go back to the queue? As can easily be seen in Fig. 4 the job that just left the first station is allowed to enter it again.

From the theorem, it follows that an equivalent closed network without blocking can be developed, which has $\hat{K} = 5$ total jobs. The state space, Fig. 6, for $\hat{K} = 5$ for the equivalent closed network without blocking has an identical structure to Fig. 5.

The throughput value $\lambda_{NB}(5)$ for the equivalent nonblocking network is obtained by mean value analysis and is exactly equal to the throughput value $\lambda_B(10)$ of the blocking network, (10)

$$\lambda_B(10)=\lambda_{NB}(5)=0.496$$

The *mean number* of jobs in the ith station of the blocking network is determined by (15)

$$\overline{k_1}=6.875 \quad \overline{k_2}=3.124.$$

The *blocking probabilities* are computed from (16)

$$P_{B_1}=0.007 \quad P_{B_2}=0.687.$$

Other performance measures can be computed by according formulae.

V. Conclusion

We have treated the classical blocking closed two-station queueing networks with multiple servers. The concept is based on the transformation of the state space of blocking network into a state space for a nonblocking network. The equilibrium state probability distribution for the blocking network is then computed from the nonblocking network which has a product form solution.

Since our solution is not restricted solely to stations of finite capacity or to serially switched stations, it extends to networks in which the capacity of one or two stations is infinite and one or two stations possess a transition probability back to itself. Since the state space of two-station closed queueing networks is one-dimensional in all these cases, our theorem holds.

Queueing networks with $N > 2$ stations have $(N - 1)$-dimensional state spaces and the results are no more exact because the state spaces of both networks do not agree exactly, i.e., they cannot be transformed bijectively into each other as is the case of two-station networks.

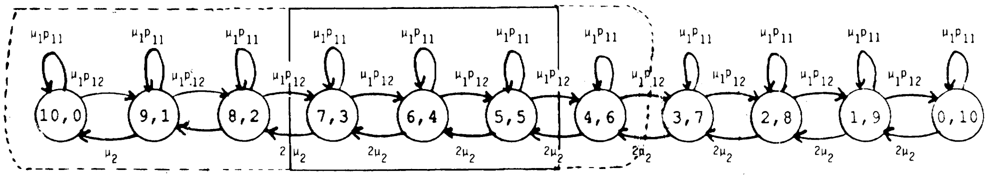

Fig. 4.

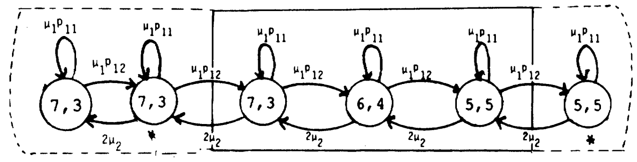

Fig. 5.

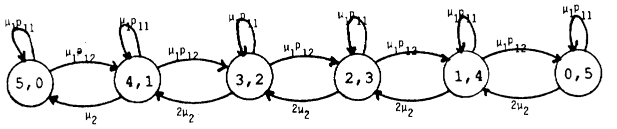

Fig. 6.

We demonstrate the difference between classical, rejection, and service blocking in the following five examples. Note that Onvural and Perros [12] have shown that the service blocking and rejection blocking types are identical, if the network is tandem. Therefore, the results given in the following are valid for both blocking types. The results for classical blocking are computed by our method and the results for rejection and service blocking are obtained by using the formulas suggested in [8]. The inut parameters for the two-station, single-server queueing network models to investigate are given as follows:

number	K	M_1	M_2	$1/\mu_1$	$1/\mu_2$
1	20	18	10	2	3.33
2	25	18	13	0.8	0.5
3	15	10	10	1.111	1.5
4	8	4	7	10	2.5
5	50	31	24	4	2.

The results for the mean number of jobs $\bar{k}_i$, the mean residence time $\bar{t}_i$ and the throughput λ are shown below:

	$\bar{k}_1$	$\bar{k}_2$	$\bar{t}_1$	$\bar{t}_2$	λ
Classical	10.85	9.14	36.28	30.54	0.291
Rejection service	11.41	8.59	38.28	28.83	0.298
Classical	17.09	7.90	13.79	6.38	1.230
Rejection service	16.60	8.39	13.59	6.87	1.221
Classical	6.31	8.69	9.81	13.51	0.640
Rejection service	6.67	8.37	10.75	13.42	0.620
Classical	3.92	4.08	39.21	40.84	0.090
Rejection service	3.68	4.32	37.26	43.69	0.098
Classical	30.53	19.46	122.61	78.17	0.25
Rejection service	30.01	19.90	122.32	80.90	0.246

note that all results are exact in both cases.

Acknowledgment

I would like to thank Prof. U. Herzog, Prof. R. Hofman, and Dr. G. Bolch for their continuous support of my research at the University of Erlangen-Nuremberg. Thanks also to Prof. H. Perros and to the referees for their constructive criticisms and patient advise.

References

[1] I. F. Akyildiz, "Leistungsanalyse von Multiprozessorsystemen mit Prozesskommunikation," Univ. Erlangen-Nuernberg, vol. 17, Erlangen, F. R. G., Dec. Tech. Rep., 1984.

[2] ——, "Analysis of closed queueing network models with blocking," Louisiana State Univ., Baton Rouge, Tech. Rep. TR-85-046, Sept. 1985.

[3] S. Balsamo and G. Iazeolla, "Some equivalence properties for queueing networks with and without blocking," in *Proc. Performance 83 Conf.*, 1983, pp. 351-360.

[4] F. Baskett, K. M. Chandy, R. R. Muntz, and G. Palacios, "Open, closed, and mixed network of queues with different classes of customers," *J. Ass. Comput. Mach.*, vol. 22, pp. 248-260, Apr. 1975.

[5] O. I. Boxma and A. G. Konheim, "Approximate analysis exponential queueing systems with blocking," *Acta Inform.*, vol. 15, pp. 19-66, Jan. 1981.

[6] J. P. Buzen, "Queueing network models of multiprogramming," Ph.D. dissertation Div. Eng. and Appl. Sci., Harvard Univ., Cambridge, MA, Aug. 1971.

[7] K. M. Chandy, U. Herzog, and L. Woo, "Parametric analysis of queueing network models," *IBM J. Res. Dev.*, vol. 19, pp. 43-49, Jan. 1975.

[8] W. J. Gordon and G. F. Newell, "Cyclic queueing systems with restricted queues," *Operat. Res.*, vol. 15, pp. 266-277, Apr. 1967.

[9] A. Hordijk and N. van Dijk, "Networks of queues with blocking," in *Proc. 8th Int. Symp. Comput. Perform. Modeling, Meas., and Eval.*, Amsterdam, The Netherlands, Nov. 4-6, 1981.

[10] A. G. Konheim and M. Reiser, "A queueing model with finite waiting room and blocking," *J. Ass. Comput. Mach.*, vol. 23, pp. 328-341, Apr. 1976.

[11] G. Latouche and M. Neuts, "Efficient algorithmic solutions to exponential tandem queues with blocking," *SIAM, Alg. Dis. Meth.*, vol. 1, Mar. 1980.

[12] R. O. Onvural and H. G. Perros, "On equivalences of blocking mechanisms in queueing networks with blocking," North Carolina State Univ., Raleigh, NC, Tech. Rep. 85-09, 1985.

[13] H. G. Perros, "A symmetrical exponential open queue network with blocking and feedback," *IEEE Trans. Software Eng.*, vol. SE-7, pp. 395-402, 1981.

[14] ——, "Queueing networks with blocking: A bibliography," *ACM Sigmet. Perform. Eval. Rev.*, Aug. 1984.

[15] H. G. Perros and T. Altiok, "Approximate analysis of open networks of queues with blocking: Tandem configurations," *IEEE Trans. Software Eng.*, vol. SE-12, pp. 450-462, Mar. 1986.

[16] B. Pittel, "Closed exponential networks of queues with saturation: The jackson type stationary distribution and its asymptotic analysis," *Math. Operat. Res.*, vol. 4, pp. 367-378, 1979.

[17] M. Reiser and S. S. Lavenberg, "Mean value analysis of closed multichain queueing networks," *J. Ass. Comput. Mach.*, vol. 27, pp. 313-322, Apr. 1980.

[18] C. H. Sauer and K. M. Chandy, *Computer Systems Performance Modeling*. Englewood Cliffs, NJ: Prentice-Hall, 1981.

[19] R. Suri and G. W. Diehl, "A new building block for performance evaluation of queueing networks with finite buffers," in *ACM Sigmet. Conf. Proc.*, Cambridge, MA, Aug. 1984, pp. 134-142.

[20] ——, "A variable buffer-size model and its use in analyzing closed queueing networks with blocking," *Management Sci.*, vol. 32, pp. 206-225, Feb. 1986.

[21] Y. Takahashi, H. Miyahara, and T. Hasegawa, "An approximation method for open restricted queueing networks," *Operat. Res.*, vol. 28, pp. 594-602, May-June 1980.

[22] D. D. Yao and J. A. Buzacott, "Modeling a class of state-dependent routing in flexible manufacturing systems," *Annals Operat. Res.*, vol. 3, pp. 153-167, 1985.