International Journal of Operations Research Vol. 1, No. 1, 71−76 (2004)

A Queueing Model of General Servers in Tandem with Finite Buffer Capacities

Hsing Luh*

Department of Mathematical Science, National ChengChi University, 64, Sec. 2,Zhi-nan Rd., Wenshan, Taipei 116, Taiwan, ROC

Abstract**—**We consider a queueing model with finite capacities. External arrivals follow a Coxian distribution. Due to the limitation of the capacity, arrivals may be lost if the buffer is full. Our goal is to study the probability of blocking. In order to obtain the steady-state probability distribution of this model, we construct an embedded Markov chain at the departure points. The solution is solved analytically and its analysis is extended to semi-Markovian representation of performance measures in queueing networks.

Keywords**—**Queueing Networks, Probability Distribution, Embedded Markov Chain.

1. INTRODUCTION

In this paper we consider an open and finite capacity system which is constructed by a series of queue. There are only one server and one queue in each station. The service discipline is First-Come-First-Served.

Customers come to this system with interarrival times, denoted by a random variable A. An arrival may enter the system with a token, or he will be rejected. When a customer gets a token he will join server 1 if the server is idle, but waits for service in queue 1 when the server is busy. After finishing service in server 1, he then joins server 2 if the server is idle or stays in queue 2 if server 2 is busy. He leaves the system when finishing all services and returns the token to the token queue. Such a fundamental model can be used for studying the performance on a manufacturing line in which tokens may represent workers or machines while the customers are jobs.

If there is only one server in the network and assume A has an exponential distribution then the model is the same as the $M/G/1/N$ loss system where N is the maximum number of the customer in the system. If there is no external arrival, and N tokens are replaced by N jobs in the network, the model becomes a closed model which has been considered in Daduna (1985). In this paper, we will provide a probability that an arriving customer finds no token in the token queue, which is called a probability of blocking.

The paper is organized in the following. In Section 2, we give a detail description of this model and construct an embedded Markov chain. In Section 3, by considering the embedded Markov chain we obtain the steady-state distribution at the system. In Section 4, we will provide a formula for the probability of blocking.

2. THE MODEL

2.1. Definition and notations

To make a model whose solution analysis is attainable without loss of generality. Consider a semaphore queueing system consisting of a series of servers in which server S has the slowest service rate, namely the bottleneck, which is shown in Figure 1. Instead of a series of servers, assume S has a general service time in order to capture the property of a flow time that spent in finishing all services. The customers following a Coxian distribution with M phases will ask for the permission to enter the system. Only those customers who have tokens from queue Q_3 are legitimate to enter queue Q_2. Otherwise, they wait in Q_1 until tokens are available in Q_3. Once the customer enter queue Q_2, he shall get the service from S immediately if server S is idle. Otherwise, the customer waits in Q_2 if server S is busy. The service discipline is first-come-first-serve. The server can only serve one customer each time and its service distribution is general which is assumed independent with the arrival process. Suppose the number of tokens N is finite which may be considered as the maximal capacity of Q_2 but the size of Q_1 is infinite. Any token by itself along can not enter Q_2.

Each customer arriving at the network is assigned one token from Q_3 in order to enter Q_2 if Q_3 is not empty. This token shall not be returned to Q_3 until the customer taking

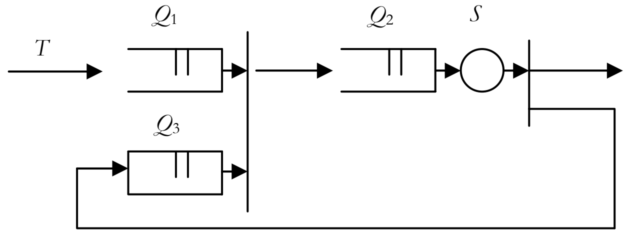

Figure 1. A semaphore queue.

it finishes the service in S and leaves the system. Assume

* Corresponding author's email: paul@math.nccu.edu.tw

there is no delay for returning a token to Q_3. This is a general description of a semaphore queue. Although it is originally designed to study the queueing problems in Telecommunication. See Fdida et al. (1990).

The interarrival times, A, are assumed characterized by a Coxian distribution with M phases, denoted by $K_M(t)$ which has mean $1/\lambda$. Its phases are numbered in backward order shown in Figure 2. This approach has been widely applied to many queueing models. For example, they were applied in Carroll et. al (1982), Perros (1983) and Bertsimas (1990). The parameters of this distribution are η_k, k = 1, 2, …, M. and a_k, k = 1, 2, …, M- 1, where η_k is the service rates with respect to phase k, and a_k is the probability that a customer is assigned to phase k from phase k+1. All the service times in each phase are independent exponentially distributed. Its Laplace-Stieltjes transform $\phi_M(s)$ is written as follows,

$$\phi_M(s)=\sum_{i=1}^{M} a_M a_{M-1}\cdots a_{M-i+1}(1-a_{M-1})\prod_{j=1}^{i}\frac{\eta_{M-j+1}}{s+\eta_{M-j+1}}$$

where $a_M \triangleq 1$ and $a_0 \triangleq 0$. The service times at S have a general distribution function $H(t)$, $t \geq 0$, with $H(0^+) < 1$ where $H(\cdot)$ has a finite expectation $1/\mu$. Furthermore, the set of all service times is assumed a stochastically independent family.

Let (τ_n: n = 0, 1, 2, …) be the sequence of departure time instants at S, where τ_0 = 0. Consider the process $\Lambda=\{(i_n, j_n, k_n): n=0, 1, 2, \ldots\}$ embedded at (τ_n: n = 0, 1, 2, …), where i_n and k_n are the number of customers presenting at Q_1 and Q_3, respectively; j_n is the number of phases left in the current interarrival times of A. It is cleared that at any instant τ_n, $i_n \cdot k_n = 0$. Note that customers arrive at the system according to a phase-type process and the interarrival times of each phase is exponential, so we have a Markov chain embedded in the process Λ. Then we may claim that the process Λ is a Markov chain with state space

$$v \equiv \{0,1,2,\ldots\}\times\{1,2,\ldots,M\}\times\{0,1,2,\ldots,N\}$$

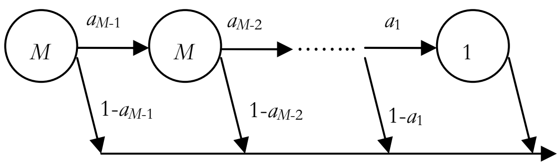

Figure 2. A M-phase Coxian distribution

Let $p(s_1; s_2)$ be the transition probability from state s_1 to state s_2 and

$$P_{(N)}=(p((i_1,j_1,k_1),(i_2,j_2,k_2)):(i_1,j_1,k_1),(i_2,j_2,k_2)\in\Lambda)$$

be the transition probability matrix of the embedded markov chain. We shall deter-mine $P_{(N)}$ and claim the Markov chain is ergodic. For this objective, we define a set of independent random variables (r.v.) with corresponding distributions in the following: Denoted by A_i the arrival time instant for the ith customer entering the system, i = 0, 1, 2, …, where A_0 = 0. Let $T_i = A_i - A_{i-1}$ be the interarrival time between the (i-1)th customer and the ith customer for i = 1, 2, …. Assume T_1, T_2, … are independent and identically distributed. Since T_i has a Coxian distribution with M phases, we may think of each T_i as being the sum of M exponential random variables with corresponding branching probabilities, that is $\sum_{i-1}^{M} X_i \prod_{j=i}^{M-1} a_j$, where $\prod_{i=j}^{k} a_i \triangleq 1$, if $j > k$. Denote by T the random variable with the distribution function $K_M(t)$. Let

Δ_k := r.v. distributed according to K_j;

X_i := r.v. distributed according to an exponential distribution with mean $1/\eta_i$, i = 1, 2, … , M;

Y := r.v. distributed according to $H(t)$

Define

$$r_{j,m} \triangleq P_r\left\{X_j + X_{j+1} + \cdots + X_{m+1} < Y\right\};\ j \geq m$$

$$\delta_{k,v,j} \triangleq P_r\left\{\Delta_k + \sum_{i=1}^{v-1}\Omega_i + \sum_{i=j+1}^{M} X_i < Y\right\}$$

$$\sigma_j \triangleq P_r\left\{X_j > Y\right\}$$

Based on the symbols defined, we derive the state transition probabilities in Λ in the following lemma.

Lemma 1 Consider all states in Λ, the probabilities of their transitions are given below:

(a) For $0 \le k \le N$, $n \geq 0$, $1 \le j \le M$, $1 \le m \le M$

$$Pr((0, j, N), (n, m, k)) = Pr((0, M, N-1), (n, m, k))$$

(b) For $0 \le k \le N$-1, $1 \le m \le j \le M$,

$$Pr((0,j,k),\ (0,m,k+1))=\begin{cases} r_{j,m}-r_{j,m-1} & \text{if } j>m \\ \sigma_j & \text{if } j=m \end{cases}$$

(c) For $1 \le j \le M$, $1 \le m \le M$, $i \geq 0$, $n > 0$

$$Pr((n, j, 0), (n+i, m, 0)) = \delta_{j,i+1,m} - \delta_{j,i+1,m-1}$$

(d) For $1 \le m \le j \le M$, $i \geq 0$,

$$\Pr((i+1,j,0),\ (i,m,0))=\begin{cases} r_{j,m}-r_{j,m-1} & \text{if } j>m \\ \sigma_j & \text{if } j=m \end{cases}$$

(e) For $1 \le j \le M$, $1 \le m \le M$, $i > k$, $0 \le k < N$

$$Pr((0, j, k), (i-k, m, 0)) = \delta_{j,i+1,m} - \delta_{j,i+1,m-1}$$

(f) For $1 \le j \le M$, $1 \le m \le M$, $0 \le i \le k$, $0 \le k < N$

$$Pr((0, j, k), (0, m, k-i)) = \delta_{j,i+1,m} - \delta_{j,i+1,m-1}$$

Proof: The proof may be found in Luh (1999).

2.2. Transition probability matrix

To be concise in matrix representation, the following $M \times M$ matrices are defined.

$B \triangleq [b_{jm}]$ for $1 \le j \le M$ and $1 \le m \le M$

where $b_{jm} = Pr((0, j, k), (0, m, k+1))$

$C_0 \triangleq [d_{jm}]$ for $1 \le j \le M$ and $1 \le m \le M$

where $d_{jm} = Pr((i+1, j, 0), (i, m, 0))$ for any fixed i, $i \ge 0$

$C_{i+1} \triangleq [c^i_{jm}]$ for $1 \le j \le M$ and $1 \le m \le M$

where $c^i_{jm} = Pr((n, j, 0), (n+i, m, 0)), i = 0, 1, 2, \cdots$

$E_i \triangleq [e^i_{jm}]$ for $1 \le j \le M$ and $1 \le m \le M$

where $e^i_{jm} = Pr((0, j, k), (i-k, m, 0)), i = 1, 2, \cdots$

$F_i \triangleq [f^i_{jm}]$ for $1 \le j \le M$ and $1 \le m \le M$

where $f^i_{jm} = Pr((0, j, k), (0, m, k-i)), i = 0, 1, 2, \cdots, N-1$

Obviously, B is equal to C_0 from our definition. However, we distinguish them be- cause they stand for the different conditions on state transitions. Define **1** as a n-dimensional column vector of 1.

$\overline{B} \triangleq \mathbf{1} \cdot [\overline{b}_{M1}\ \overline{b}_{M2} \cdots \overline{b}_{MM}]$

$\overline{E}_i \triangleq \mathbf{1} \cdot [\overline{e}^i_{M1}\ \overline{e}^i_{M2} \cdots \overline{e}^i_{MM}]$

$\overline{F}_i \triangleq \mathbf{1} \cdot [\overline{f}^i_{M1}\ \overline{f}^i_{M2} \cdots \overline{f}^i_{MM}]$

The state space of this system can be organized into three groups:
$\{(i \neq 0, \underline{j}, k = 0)\}$, $\{(i = 0, \underline{j}, k = 0)\}$, $\{(i = 0, \underline{j}, k \neq 0)\}$.
Within each group, for fixed i and j, the states can be ordered lexicographically in accordance with each phase k. For example, this ordering of $\{(i \neq 0, \underline{j}, k = 0)\}$ is described in the following table:

$\{(i=1, \underline{j}, k=0)\}$	$\{(i=2, \underline{j}, k=0)\}$	$\cdots$
$(1,1,0)$	$(2,1,0)$	
$(1,2,0)$	$(2,2,0)$	$\cdots$
$\vdots$	$\vdots$	
$(1,M,0)$	$(2,M,0)$	

The transition probability in matrix form among $\{(i \neq 0, \underline{j}, k = 0)\}$ is written as

$$\begin{array}{c} \\ \{1,j,0\} \\ \{2,j,0\} \\ \{3,j,0\} \\ \vdots \\ \vdots \end{array}
\begin{array}{c} \begin{array}{ccccc} \{(1,j,0)\} & \{(2,j,0)\} & \{(3,j,0)\} & \cdots & \end{array} \\
\begin{pmatrix} C_1 & C_2 & C_3 & C_4 & \cdots \\ C_0 & C_1 & C_2 & C_3 & \cdots \\ & C_0 & C_1 & C_2 & C_3 \\ & & C_0 & C_1 & C_2 \\ & & \ddots & \ddots & \ddots \\ & & & & \end{pmatrix} \end{array}$$

Consider the group $\{(i = 0, j, k = 0)\}$ and $\{(i = 0, \underline{j}, k \neq 0)\}$, that is no customers in queue Q_1 but tokens may or may not appear in queue Q_3. The transition matrix occurs among $\{(i = 0, \underline{j}, k = 0)\}$ and $\{(0, \underline{j}, 1), (0, \underline{j}, 2), \cdots, (0, \underline{j}, N)\}$ may be written as

$$\begin{array}{c} \\ \{(0,\underline{j},0)\} \\ \{(0,\underline{j},1)\} \\ \{(0,\underline{j},2)\} \\ \vdots \\ \{(0,\underline{j},N-1)\} \\ \{(0,\underline{j},N)\} \end{array}
\begin{array}{c} \begin{array}{ccccc} \{(0,\underline{j},0)\} & \{(0,\underline{j},1)\} & \{(0,\underline{j},2)\} & \cdots & \{(0,\underline{j},N)\} \end{array} \\
\begin{pmatrix} F_0 & B & & & & & \\ F_1 & F_0 & B & & & & \\ F_2 & F_1 & F_0 & B & & & \\ \cdots & \cdots & \cdots & \cdots & \ddots & & \\ F_{N-1} & F_{N-2} & \cdots & \cdots & F_1 & F_0 & B \\ \overline{F}_{N-1} & \overline{F}_{N-2} & \cdots & \cdots & \overline{F}_1 & \overline{F}_0 & \overline{B} \end{pmatrix} \end{array}$$

Consider the transition probability between the group $\{(i = 0, \underline{j}, k = 0)\}$, $\{(i = 0, \underline{j}, k \neq 0)\}$ and $\{(i = 0, \underline{j}, k = 0)\}$.The transition matrix occurs among, $\{(i = 0, \underline{j}, k = 0)\}$, $\{(0, \underline{j}, 1), (0, \underline{j}, 2), \cdots, (0, \underline{j}, N)\}$ and $\{(i \neq 0, \underline{j}, k =$

0)}may be written as

$$
\begin{array}{c|cccc}
 & \{(1,\underline{j},0)\} & \{(2,\underline{j},1)\} & \{(3,\underline{j},2)\} & \cdots \\
\hline
\{(0,\underline{j},0)\} & E_1 & E_2 & E_3 & \cdots \\
\{(0,\underline{j},1)\} & E_2 & E_3 & E_4 & \cdots \\
\{(0,\underline{j},2)\} & E_3 & E_4 & E_5 & \cdots \\
\vdots & \cdots & \cdots & \cdots & \cdots \\
\{(0,\underline{j},N-1)\} & E_N & E_{N+1} & E_{N+2} & \cdots \\
\{(0,\underline{j},N)\} & \overline{E}_N & \overline{E}_{N+1} & \overline{E}_{N+2} & \cdots
\end{array}
$$

Hence, the infinitesimal generator $P_{(N)}$ is written as

$$
P_{(N)} = \begin{bmatrix}
F_0 & B & & & & & & E_1 & E_2 & \cdots & \cdots \\
F_1 & F_0 & B & & & & & E_2 & E_3 & \cdots & \cdots \\
F_2 & F_1 & F_0 & B & & & & E_3 & E_4 & \cdots & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots & & & \vdots & \vdots & \cdots & \cdots \\
F_{N-1} & F_{N-2} & \cdots & & & F_0 & B & E_N & E_{N+1} & \cdots & \cdots \\
\overline{F}_{N-1} & \overline{F}_{N-2} & \cdots & & & \overline{F}_0 & \overline{B} & \overline{E}_N & \overline{E}_{N+1} & \cdots & \cdots \\
C_0 & & & & & & & C_1 & C_2 & C_3 & \cdots \\
 & & & & & & & C_0 & C_1 & C_2 & \cdots \\
 & & & & & & & & C_0 & C_1 & \cdots \\
 & & & & & & & & & C_0 & \cdots \\
 & & & & & & & & & & \ddots
\end{bmatrix}
$$

3. STATE BALANCE EQUATIONS

Let π be the vector of the steady-state probability associated with $P_{(N)}$, i.e. $\pi P_{(N)} = \pi$. To solve $\pi P_{(N)} = \pi$, we partition π conformed with the blocks of matrix $P_{(N)}$ Thus, we let

$$\pi = \left(\pi_0,\ \pi_1,\ \ldots,\ \pi_{N+1},\ \pi_{N+2},\ \ldots\right),$$

where the vector π_i, $i > 1$, are of dimension M and the vector π_0, which corresponds to the states in the boundary level 0, is of dimension M as well. The equations can be written out as follows:

$$\pi_0 F_0 + \pi_1 F_1 + \cdots + \pi_{N+1} C_0 = \pi_0 \tag{1}$$

$$\pi_{i-1} B + \sum_{i=j}^{N-1} \pi_j F_{j-1} + \pi_N \overline{F}_{N-i-1} = \pi_i \tag{2}$$
for $i = 1, 2, \ldots, N-1$

$$\pi_{N-1} B + \pi_N \overline{B} = \pi_N \tag{3}$$

$$\sum_{j=0}^{N-1} \pi_j E_{j+k-N} + \pi_N \overline{E}_{k-1} + \sum_{j=0}^{k-N} \pi_{k+1-j} C_j = \pi_k \tag{4}$$
for $k = N+1, N+2, \ldots$

$$\sum_{i=0}^{\infty} \pi_i \cdot \mathbf{1} = 1 \tag{5}$$

To solve π, it may refer to Neuts (1989). His solution procedures require a series of computational efforts which solve several nonlinear matrix equations involving the inverse of matrices of the order $(N+1)M$. In our approach, it does not need solving any system of equations because all vectors can be expressed by π_N only which is the probability of N tokens in Q_3 associated with different arrival phases. Rearranging the equations (3) (2) (1), we have

$$\pi_{N-1} = \pi_N (I - \overline{B}) B^{-1} \tag{6}$$

$$\pi_{i-1} = [\pi_i - \sum_{j=i}^{N-1} \pi_j F_{j-i} - \pi_N \overline{F}_{N-i-1}] B^{-1} \tag{7}$$

for $i = N-1, N-2, \ldots, 1$.

$$\pi_{N+1} = [\pi_0 (I - F_0) - \sum_{j=1}^{N-1} \pi_j F_j - \pi_N \overline{F}_{N-1}] C_0^{-1} \tag{8}$$

As to Eq.(4), write π_{k+1} in term of π_N for all $k \geq N+1$

$$\pi_{k+1} = \left[\pi_k + \sum_{j=0}^{N-1} \pi_j E_{j+k-N} - \pi_N \overline{E}_{k-1} - \pi_{N+1} C_{k-N} - \sum_{j=1}^{k-N-1} \pi_{k+1-j} C_j \right] C_0^{-1} \tag{9}$$

for $k = N+1$, N+2, …

In order that π is unique, it is necessary that B and C_0 are nonsingular. It implies $\sigma_j > 0$ for every j because B and C_0 are lower triangular matrices and their determinant $\prod_{j=1}^{M} \sigma_j$ must not equal to 0. Since the stability is implicitly satisfied by this assumption, this quarentees the existence and uniqueness of π, for n = 1, 2, …. Since all π_n can be written in terms of π_N, it is essential to determine the value of π_N first. Because the probability that the system is idle in $GI/G/1$ in steady-state at any time is $1-\lambda/\mu$, it gives the probability of that at a departure point $\pi_N \cdot \mathbf{1} = (1-\lambda/\mu)\nu$, where ν is the normalizing factor and can be found by a formula at closed queueing models. Therefore, if each of its elements is given, the remaining π_n for n = 0, 1, 2, … is determined in Eq.(6), (7) (8) and (9) with substitution recursively.

Then, $\pi_N \cdot \mathbf{1}$ is interpreted as the probability of no customers in the system, i.e. probability of idle. In other words, $1-\pi_N \cdot \mathbf{1}$ is the probability that one of servers is busy. The value of $\sum_{i=1}^{\infty}\pi_{N+i} \cdot \mathbf{1}$ may be interpreted as the probability of no tokens available for arrivals in the system in the long run. For a $GI/G/1$ model, $\pi_{N-n} \cdot \mathbf{1}$ is the probability of n customers in the system, as $0 \le n \le N$; when $n > N$, the probability of that is π_n. The average queue length, the average number of customers in the system and the mean waiting time are expressed as follows.

L_q : queue length in Q_1 in average,

$$L_q = \{\, 1 \cdot \pi_{N+1} + 2 \cdot \pi_{N+2} + \ldots + n \cdot \pi_{N+n} + \ldots \} \cdot \mathbf{1}$$

π_0^* : $Pr\{$no token in $Q_3\}$

$$\pi_0^* = \pi_0 + \sum_{n=1}^{\infty}\pi_{N+n} \cdot \mathbf{1}$$

L_Q : average number of customers staying in Q_2 and in service,

$$L_Q = \sum_{n=1}^{N}(N-n) \cdot \pi_n \cdot \mathbf{1} + N \cdot \pi_0^* \cdot \mathbf{1}$$

By Little's formula, the mean waiting time of the system is

$$\frac{1}{\lambda}\,[L_q + L_Q]$$

where λ is the average arrival rate.

4. A MODEL WITH FINITE CAPACITIES

The formulation previously described can be generalized to allow the case of being lost in which arrivals may be lost due to the limited buffer sizes. Suppose in the system taken into account here the buffer size of Q_1 is set to 0 while the other assumptions are the same as shown before in Figure 1. Such a system may be considered as a $GI/G/1/N$ queue in which N is the maximum capacity of the system. Thus, there are no more than N customers staying in the system at any time. In general, the probability that customers who are rejected and the effective arrival rate are of interest in this case. After the semaphore queueing model has been introduced, it becomes convenient to solve these problems.

Following the same notations, the state balance equations will take the equations (2), (3) and the normalization equation (5) which is

$$\sum_{i=0}^{N}\pi_i \cdot \mathbf{1} = 1$$

Hence, π_i is obtained by solving the system of such equations. Thus, $\pi_i \cdot \mathbf{1}$ is the probability of N-i customers in the system, for $0 < i \le N$. $\pi_0 \cdot \mathbf{1}$ is the probability of being lost, i.e. the probability when the system is full and customers are rejected. The effective arrival rate is

$$\bar{\lambda} = \sum_{i=1}^{N}\lambda\pi_i \cdot \mathbf{1} = \lambda(1-\pi_0 \cdot 1)\,.$$

The following theorems are only stated for conclusion without proofs.

Theorem 1: Assume that the service rate $\mu_j(n_j)$ is a nondecreasing function of n_j for each station j. The throughput of this system, denoted by $TH(N)$, is a concave function of the number of the tokens if the service rates $\mu_j(n_j)$ as a function of the local queue length n_j, has the same property for each j.

Theorem 2: If the service rate at every station is a nondecreasing and concave function of the queue length at that station, then the blocking probability is a nonincreasing convex function of the maximum total tokens available in the network.

The proofs are omitted here. Similar results may be found in Shanthikumar and Yao (1988).

REFERENCES

1. Berstsimas, D. (1990). An Analytic Approach to A General Class of G/G/s Qeueing Systems. *Operations Research*, 38(1): 139-155.
2. Carroll, J.L. and A. D.E. Liefvoort and L. Lipsky. (1982). Solutions of M/G/1//N- type Loops with Extensions to M/G/1 and GI/M/1 Queues. *Operations Research*, 30(3): 490-514.
3. Daduna, H. (1985). Two-Stage Cyclic Queues with Nonexponential Severs: SteadyState and Cyclic Time. *Operations Research*, 34(3): 455-459.
4. Fdida, S., H.G. Perros and A. Wilk. (1990). Semaphore Queues: Modeling Multilayered Windows Flow Control

Mechanisms. *IEEE Transaction on Computers*, COM-38: 309-317.
5. Luh, H. (1999). Derivation of the N-step interdeparture time distribution in GI/G/1 queueing systems. *European Journal of Operations Research*, 118: 194-212.
6. Neuts, F. (1989). *Structured Stochastic Matrices of M/G/1 Type and Their Applications*. Marcel Dekker, New York.
7. Perros, H.G. (1983). On the M/C_k/1 Queue. *Performance Evaluation*, 3: 83-93.
8. Shanthikumar, J.G. and Yao, D.D. (1988). Second-Order Properties of the Throughput in a Closed Queueing Network. *Mathematics of Operations Research*, 13: 524-534.