1987

Closed-Network Duals of Multiques with Application to Token-Passing Systems

Vernon J. Rego
Purdue University, rego@cs.purdue.edu

Wojciech Szpankowski
Purdue University, spa@cs.purdue.edu

Report Number:
87-660

CLOSED-NETWORK DUALS OF MULTIQUES WITH APPLICATION TO TOKEN-PASSING SYSTEMS

Vernon Rego
Wojciech Szpankowski

CSD-TR-660
February 1987

Abstract

Asymmetric multiqueue systems (e.g., Token-passing systems) are known to raise analytically difficult questions in the sense of modelling tractabilty. A prime example is the asymmetric multiqueue system with a single cyclic-server and one-at-a-time customer service. In the interests of sound approximation techniques, we examine the feasibility of approximation schemes using known results in the realm of closed queueing networks. Focusing on unit capacity buffer systems and cycle-times of the server on such systems, a notion of duality is introduced. In essence, this duality says that the single server on the multiqueue system can be viewed as a single customer on a dual closed network. Correspondingly, the customers in the multiqueueing system take on the role of servers in the dual closed queueing network. In this work, we examine conditions under which this duality is strong (or weak). In order to do this, we solve two open problems: the cycle-time distribution on an asymmetric system, and the probability of any station being empty (nonempty) at its scan instant. Numerical results indicate that closed network duality is strong under testable conditions and can thus be used as a guide to useful approximations.

1. INTRODUCTION

Queueing systems that are essentially multiqueues which share a single server, i.e., a scarce resource such as a channel, have received a considerable amount of attention in the recent literature, especially in modelling token-passing systems (see for example [FeAm85, Ferg86, PaDo86, KiTa86, BuxW83, BuTr83, RuDe83, etc.]). It appears that the determination of sound measures of performance for such systems, under realistic assumptions such as asymmetric traffic, finite buffers, nonexhaustive-service, general input, general token-passing time, and general transmission-time probability distributions are fairly difficult to obtain, as can be witnessed from the literature. For example, it is known that obtaining the distributions of the number of packets queued at each station (either embedded or arbitrary-time) is a formidable open problem, as is the problem of obtaining the waiting-time distribution. Further, note that the problem is difficult no matter what the buffer size at each station (i.e., all infinite, all finite with the same size, all finite with different sizes, etc.) or the service-discipline at each queue (i.e., exhaustive, nonexhaustive, etc.).

It must be mentioned that some results are available for token-passing systems. Notable results include the *mean* waiting times and queue-lengths for *symmetric* exhaustively served queues [KoMe74], a generalization of this to *asymmetric* queues [Swar80], *mean* waiting times of packets in symmetric systems with *nonexhaustive service* [Fuhr85], *mean* waiting times in queues with gated service (i.e., where only those customers found waiting for the server at a given queue, when the server arrives at the queue, are served) [FeAm85], and the *variance* of the *cycle-time* of an asymmetric system with an infinite number of stations and exhaustive service [Ferg86]. There is a version of the token-ring problem, one on which the the original token-passing LAN protocol was defined [IEEE84a, IEEE84b], which is particularly formidable. This is the problem of *nonexhaustive* service on an *asymmetric* system, where at most k_j packets can be transmitted by station j, each time station j acquires the *free* token, $1 \le k_j < \infty$. To the best knowledge of the authors, the only version of this problem that has been solved is that by Boxma [Boxm84], where the number of stations is $N = 2$, and $k_j = 1$.

The present paper has been inspired mainly by two things. First, there are the many difficulties associated with exact analyses of nonexhaustive service token-passing systems. Next, consider the tremendous increase in the complexity of exact analyses of such systems when the number of servers (or tokens on a token ring) is more than one. In a pioneering paper, Morris and Wang [MoWa84] present some interesting approximations for multiqueue systems with multiple cyclic servers. A key conclusion of this study was that for fairly wide ranges of system parameters, the many servers tended to behave somewhat *independently* of one another. The net effect observed was an "averaging out" of cycle-time variability, which in turn effectively reduced waiting times of jobs at the various queues.

In a similar spirit, the present paper examines the feasibility of a new kind of approximation for multiqueues. Since this is preliminary work, we focus our attention on the case of a single cyclic server (token). Basically, we claim that it is possible to view certain multiqueue systems as *duals* of certain closed networks, with respect to *specific* performance measures. In particular, we restrict our attention to the *distribution of cycle-time* (where the cycle-time is the random time between two consecutive visits of the server at any station on the multiqueueing system).

Our motivation for studying multiqueues with a cyclic-server in the framework of a closed network dual is to exploit results that are well known for closed networks (e.g., [Klei76], [Chow80], [ScDa83], [Dadu86a], [Dadu86b]). For example, consider the problem of determining the position of the token on a token ring at an arbitrary instant in time. Alternately, consider the difficulty involved if one was to try to compute the joint distribution of the positions of N different tokens on a multiple-token ring (i.e., cyclic-servers on a multiqueue). In the following sections we explain how one can use the closed network duality and the results thus available to obtain useful performance measures that are either impossible to obtain or simply too difficult to obtain by analyzing multiqueues in a direct manner. In the course of establishing duality, we solve two open problems. The first is obtaining the distribution of cycle-time on an N station, asymmetric, unit-capacity buffer token ring system. The second is obtaining the probability that a given station is found empty when the free token arrives at the station.

In the following sections, we proceed as follows. In section **2**, the multiqueue and cyclic-server model is presented formally, along with the closed network dual. In section **3**, we demonstrate the existence of a duality by computing the cycle-time distribution *exactly* on dual systems and comparing them. The first is the closed network system, and the second is an asymmetric multiqueue dual system with a cyclic-server and station buffers each of unit capacity. We use Poisson arrivals but otherwise general distributions. In this sense the solution is very general. A disadvantage of our solution is that it is computationally feasible only for small systems, i.e., the number of stations less than ten. The complexity of computation grows exponentially with N (the number of stations), since a $2^N \times 2^N$ probability transition matrix is involved. However, we stress that this is an exact solution, and sufficient to demonstrate the duality that we are interested in. A somewhat similar results can be seen in [TeTH86] but the system analyzed in this reference is a *symmetric* one , while our analysis is for an asymmetric system. Additionally, this reference utilizes an assumption of inter-departure time independence to obtain the Laplace-Stieltjes transform of customer inter-departure time distribution, not a distribution in any computational form. Thus the present work is considerably different in that we have a more general system, obtain explicit cycle-time (not inter-departure time) distributions that are both exact and approximate. These are obtained in computationally convenient forms. Also, we focus on a duality relationship between queueing systems, rather than performance measurements. For completeness, we also mention that symmetric systems have been well studied in the form of the machine-interference problem (see for example, [Mack57], [MaMW57], also see [Taka86]). In section 4 we demonstrate some numerical results for cycle-time distributions on the dual systems. Finally, we present an immediate application of the duality and conclude the paper.

2. DUAL QUEUEING MODELS

An illustration of the multiqueue and cyclic server model is shown in Figure 1a for $N = 3$ queues. Customer interarrival times at queue j are assumed to be Poisson with rate λ_j, $j = 1, 2, \ldots, N$. Additionally, customer arrivals within and between queues are mutually independent. A single server walks around the ring of queues, from queue j to queue $(j \text{ mod } N) + 1$, serving at most one customer per queue if he finds the queue nonempty. This is called the *one-at-a-time* (nonexhaustive) service discipline.

When a customer is served at some queue j, the server spends a random time X_j at station j, where $Pr[X_j \le t] = B_j(t)$. That is, X_j is the service time of a queue-j customer. If queue j is empty when the server gets there, it takes the server a small random time Y_j to detect this, and switch past this queue, with $Pr[Y_j \le t] = S_j(t)$. Thus, each time the server arrives at station j,

$j = 1, 2, ..., N$, either a random time X_j or a random time Y_j is spent by the server at station j. On leaving station j, the server must walk to station k, $k = (j \ mod \ N) + 1$. This walking time is taken to be a random time W_k, where $Pr[W_k \le t] = V_k(t)$. It is assumed that the distributions $B_j(\cdot)$, $S_j(\cdot)$, and $V_j(\cdot)$ are general, $j = 1, 2, \cdots, N$, with finite first and second moments. The quantities just described are summarized below.

λ_j	customer arrival rate at station j
X_j , $X_j \sim B_j(\cdot)$	customer *service* time at station j
Y_j , $Y_j \sim S_j(\cdot)$	time for server to *switch* past empty station j
W_k , $W_k \sim V_k(\cdot)$	time for server to *walk* from station j to station $(j \ mod \ N)+1$.

Let p_j denote the (stationary) probability that the cyclic-server finds queue j nonempty on any visit to this queue, $1 \le j \le N$. Consider the following scenario (see Figure 1b). We treat the cyclic-server multiqueue system as a closed network with a *single* customer by interchanging the notion of server and customer. That is, the cyclic-server on the multiqueue system is viewed as the only customer on a certain closed network. The N queueing stations of the multiqueueing system could be considered as stations at which two kinds of servers provided service on the closed network. At each station j on the closed network, one server provides service of length X_j, and the other server provides service of length Y_j, $1 \le j \le N$. On arriving at station j, $1 \le j \le N$, the cyclic-customer chooses the server with service-time X_j with probability p_j, and the server with service-time Y_j, with probability $(1-p_j)$. On completing of a service at station j the cyclic-customer is forced to visit a single server which provides service of length $W_{(j \ mod \ N)+1}$, $1 \le j \le N$.

In the preceding description, observe that the closed network is really a dual of the cyclic-server multiqueue, with the notion of servers and customers interchanged. The cyclic-server on the multiqueue system becomes the single cyclic-customer on the closed network. The amount of time that the cyclic-server spends at each queue j on the multiqueue system, at steady-state, is either X_j if a customer (from the multiqueueing system) is present (i.e., with probability p_j), or Y_j if the queue is empty (i.e., with probability $(1-p_j)$). On the closed network, the latter times are represented via routing probabilities and two different types of servers.

The existence of a duality between closed networks and multiqueues has interesting possibilities, as was indicated earlier. At this stage, two important questions arise quite naturally. These are:

(1) How strong is the duality?

(2) If duality (either weak or strong) can be established, with respect to a given performance measure, how can this duality be exploited?

Since our work is preliminary, we attend to the first question with greater vigour. By focusing our attention on cycle-times, we show that the duality is strong under certain system parameters and weak under others. This means that if one is solving a multiqueue and cyclic server problem, with the one-at-a-time service discipline, the duality will ensure that closed network approximations can be very good under certain system parameters. Alternately, certain multiqueue approximations will also hold good for our closed networks. In the past a similar study, in the spirit of an approximation, was done by Kuehn [Kueh79]. Besides introducing duality, our results differ from Kuehn's in that we compute exact results for verification (while Kuehn uses simulation), and we use unit-capacity buffers (while Kuehn uses infinite capacity buffers). The extent to which this duality can be exploited is a very interesting question. It is extremely important that the practitioner be aware of when the duality works and the quantities that can accurately be obtained via this duality.

3. ESTABLISHING DUALITY

One way to establish that cycle-times on the dual systems can be close for certain system parameters is to provide a formal proof. While this approach can be effective in answering the question, it does not provide quantitative answers. A nontrivial issue in constructing such a proof is that not much is known either qualitatively or quantitatively about the *exact* nature of cycle-times on the multiqueue system with the one-at-a-time service discipline, asymmetric traffic, and either finite or infinite buffers.

We choose an alternate method of establishing duality. We compute the distribution of the server's cycle-time on the multiqueue system and also the distribution of the single customer's cycle-time on the closed network. In doing this, two important cases must be distinguished. The first is the case of an asymmetric multiqueue system in which each station has a buffer of infinite capacity and its corresponding closed network dual. Establishing duality in this case proves to be an intractable problem because the cycle-time distribution for this multiqueuing system is as yet undetermined (see [FeAm85, Boxm84]). Interestingly enough, for this case the probability p_j' that station j is nonempty when the server arrives at this station, $1 \le j \le N$, is easily computed (see [Kueh79]). In this sense, the closed network dual of the multiqueueing system is tractable, even though the multiqueueing system is not. However, we cannot establish the conditions under which the duality holds since we cannot compute the server's cycle-time distribution on this multiqueueing system exactly.

The second case (i.e., the one that we examine in detail) is the case of an asymmetric multiqueue system in which each station has a buffer of unit capacity and its corresponding closed network dual. Again, interestingly enough, the probability p_j that station j is nonempty on the server's arrival at station j, $1 \le j \le N$ is *not* known. Hence, the closed network dual of the multiqueue system is not readily available. We show how the probability p_j can be obtained. Additionally, the server's cycle-time distribution on this multiqueue system is also unknown. We also demonstrate how this distribution can be obtained for an N-station system in order to establish duality.

Thus, in order to establish the duality, it is necessary to solve two unsolved problems. These are (for asymmetric systems),

(1) obtaining the cycle-time distribution in a multiqueue system with cyclic server, one-at-a-time service discipline, and unit capacity buffers, and

(2) obtaining the probability that an arbitrary station is empty on the servers' arrival at that station, in the above system.

3.1 Cycle-time on the closed network-dual

Define the cycle-time C_d of the *single* customer on the closed network (i.e., the dual of the cyclic-server's cycle-time on the multiqueueing system) as the random time between two consecutive visits of the customer to an arbitrary station. The random variable C_d can be computed as a sum of independent random variables,

$$C_d \quad = \quad \sum_{j=1}^{N} W_j \; + \; R_j \tag{3.1}$$

where the newly introduced random variable R_j has a mixture distribution, i.e., $Pr[R_j \le t] = p_j B_j(t) + (1-p_j) S_j(t)$, $1 \le j \le N$. In other words, the amount of time spent by the customer at a two-server station j is a mixture of the random times X_j and Y_j. Let $F_d(\cdot)$ denote the distribution of C_d. The distribution $F_d(\cdot)$ is easily obtained once p_j is known, $1 \le j \le N$. For example, if the distributions $B_j(\cdot)$, $S_j(\cdot)$, and $V_j(\cdot)$ are exponential, for each j, it is clear from (3.1) that F_d is a discrete mixture distribution, given by the convolution of N different exponential distributions and N different two-phase generalized Erlangian (i.e., hyperexponential) distributions. Consequently, the computation of the customer's cycle-time distribution is an O(N) operation, assuming that the N probabilities $\{p_j \,;\, 1 \le j \le N\}$ are known. In the case of infinite capacity buffers, the amount of work required to obtain these probabilities is O(N) (see [Kueh79]).

3.2 Cycle-time on the multiqueue system

Consider an N station mutliqueue system with unit capacity buffers (and one-at-a-time service discipline). In order to illustrate the ideas clearly, we first give the construction for $N = 3$ queueing stations, and the general construction immediately after. Let C denote the random time between two consecutive visits of the cyclic-server at station N (which we always take to be a reference station). Suppose that we place an observer at the exit point of the server from station N. On each cycle, the server constructs an N-bit binary vector $\mathbf{Z} = \langle Z_1, Z_2, \ldots, Z_N \rangle$, where

$Z_j = 1$ if a customer is served at station j during that cycle, and $Z_j = 0$ otherwise, $1 \le j \le N$. Each time the server exits from station N, the observer obtains the most recent vector $\mathbf{Z}$ from the server. Given a vector $\mathbf{Z}$, the observer can reconstruct the service or switching events at the various stations for the last cycle. Additionally, the observer can compute the length of the most recent cycle of the server made with respect to station N as $\sum_{j=1}^{N} \{Z_j X_j + (1-Z_j) Y_j + W_j\}$.

The *state* of the observer at any time t is defined to be the most recently acquired vector $\mathbf{Z}$ from the server. Let t_n denote the time of the n^{th} exit of the server from station N, at which time the vector $\mathbf{Z}_n$ is given to the observer. At time t_{n+1}, $t_{n+1} > t_n$, the observer receives the vector $\mathbf{Z}_{n+1}$ from the server, and is said to make a transition from state $\mathbf{Z}_n$ to state $\mathbf{Z}_{n+1}$. Since each buffer is of unit capacity, it is easy to see that whether a station is empty or not a given visit by the server only depends on the amount of time since the server's last visit to the station. Thus, events at each station are Markovian. However, these events are also strongly dependent on events at other stations. We now state a theorem whose formal proof can be found in [ReNi86], but for which we constructively show how probability transitions can be computed below.

THEOREM

Let $\mathbf{Z}_n$ be the state that the observer moves into on the n^{th} exit of the server from station N, $n = 0, 1, 2, \ldots$. Then the sequence $\{\mathbf{Z}_n\}$ is a time-homogeneous Markov chain.

■

In order to obtain the cycle-time distribution of the server, we must be able to compute the probability transition matrix of this Markov chain. We illustrate the computation for an $N = 3$ station system, and then we generalize it to an arbitrary integer N. Let $\mathbf{Z}_n = \langle z_1, z_2, z_3 \rangle$ and $\mathbf{Z}_{n+1} = \langle z_1', z_2', z_3' \rangle$ be two consecutive states visited by the observer (i.e., two consecutive service-vectors defined by the server with respect to station N). We wish to compute the probability $Pr[\langle z_1', z_2', z_3' \rangle \mid \langle z_1, z_2, z_3 \rangle]$, for every pair of triples $\langle z_1, z_2, z_3 \rangle$,

$<z_1', z_2', z_3'> \in \Theta_3$, where Θ_3 is the set of all three-bit binary vectors.

In order to computer $Pr[\mathbf{z}' \mid \mathbf{z}]$ for each $\mathbf{z}, \mathbf{z}' \in \Theta_3$, it is necessary to obtain only one general formula, as we shall demonstrate in the following discussion. Let $T_j = z_j X_j + (1 - z_j) Y_j$ be the amount of time spent by the server at station j while the vector $\mathbf{z} = <z_1, z_2, z_3>$ is being defined, $1 \le j \le 3$. If $\mathbf{z}' = <z_1', z_2', z_3'>$ is to be the *next* state, let $T_j' = z_j' X_j + (1 - z_j') Y_j$ be the time spent by the server at station j during this next state $\mathbf{z}'$. Let C_i be the time between the two consecutive visits of the server to station i during the observer transition from z to $\mathbf{z}'$. It can be seen that (see Figure 2),

$$\begin{aligned} C_1 &= T_1 + W_2 + T_2 + W_3 + T_3 + W_1' \\ C_2 &= T_2 + W_3 + T_3 + W_1' + T_1' + W_2' \\ C_3 &= T_3 + W_1' + T_1' + W_2' + T_2' + W_3' \end{aligned} \tag{3.2}$$

where the unprimed terms correspond to random variables accounted for in the *current* state (i.e., current server cycle), and the primed terms are their identically distributed counterparts, respectively, accounted for in the *next* state (i.e., next server cycle) of the observer. In Figure 2 we show the server intervisit-times C_1 and C_2 for the first and the second stations, respectively, for server cycles measured with respect to the exit point of station 3. Observe that C_1 and C_2 actually overlap (shown in Fig. 2 by a dotted line). In fact, for any pair of states $\mathbf{z}, \mathbf{z}' \in \Theta_3$, the times C_1, C_2 and C_3 overlap, and hence are dependent random variables. It is important to note that this particular form of overlapping is due to our studying the embedded chain with respect to station 3. In computing $Pr[\mathbf{z}' \mid \mathbf{z}]$ the presence of dependence between C_1, C_2 and C_3 indicates that their joint distribution function is required. However, an explicit computation of this joint distribution can be avoided due to the special manner in which cycle-times overlap (see Eq. (3.2)).

Suppose that we wish to compute the transition probability $Pr[\mathbf{z}' \mid \mathbf{z}]$, for some $\mathbf{z}, \mathbf{z}' \in \Theta_3$. For any given $\mathbf{z} \in \Theta_3$, first assume that $\mathbf{z}' = <0, 0, 0>$, and define the station times $T_j = z_j X_j + (1 - z_j) Y_j$, and $T_j' = z_j' X_j + (1 - z_j') Y_j$ for $1 \le j \le 3$. Let $F_{T_j}(\cdot)$ and $F_{T_j'}(\cdot)$ denote the distribution functions for T_j and T_j', respectively, $1 \le j \le 3$. Note that $F_{T_j}(\cdot)$ is really the (given) distribution $B_j(\cdot)$ if $z_j = 1$, and the distribution $S_j(\cdot)$ if $z_j = 0$. The situation is the same with $F_{T_j'}(\cdot)$, except that z_j' is involved instead of z_j. Given that the observer is in some state $\mathbf{z}$, we now compute the probability $Pr[0, 0, 0 \mid \mathbf{z}]$ that the next state of the observer will be $\mathbf{z}' = <0, 0, 0>$ (i.e., a cycle in which all stations are found empty). Using the fact that arrivals at station j are Poisson with rate λ_j, $1 \le j \le 3$, we write

$$
\begin{aligned}
Pr[0, 0, 0 \mid \mathbf{z}] = & \int_0^\infty e^{-\lambda_1 t}\, dF_{T_1}(t) \cdot \int_0^\infty e^{-\lambda_1 t}\, dV_2(t) \\
& \times \int_0^\infty e^{-(\lambda_1 + \lambda_2)t}\, dF_{T_2}(t) \cdot \int_0^\infty e^{-(\lambda_1 + \lambda_2)t}\, dV_3(t) \\
& \times \int_0^\infty e^{-(\lambda_1 + \lambda_2 + \lambda_3)t}\, dF_{T_3}(t) \cdot \int_0^\infty e^{-(\lambda_1 + \lambda_2 + \lambda_3)t}\, dV_1(t) \\
& \times \int_0^\infty e^{-(\lambda_2 + \lambda_3)t}\, dF_{T_1'}(t) \cdot \int_0^\infty e^{-(\lambda_2 + \lambda_3)t}\, dV_2(t) \\
& \times \int_0^\infty e^{-\lambda_3 t}\, dF_{T_2'}(t) \cdot \int_0^\infty e^{-\lambda_3 t}\, dV_3(t) \qquad (3.3)
\end{aligned}
$$

The reason that (3.3) takes on this particular "product of integrals" form is explained as follows. We are computing the probability of *no arrivals* at each of the three stations on the systems, or correspondingly, that a vector $\mathbf{z}' = <0, 0, 0>$ will be constructed by the server on a certain cycle *given* that on the previous cycle the server constructed some vector $\mathbf{z}$ in Θ_3. From (3.2) we see that C_1 is the time between the two consecutive visits of the server at station 1. Notice that T_1 and W_2 are times that are unique to station 1 for any observer transition. Hence, the first pair of integrals in (3.3) describe the probability that no customer arrives at station 1 during the time $(T_1 + W_2)$. However, when the server is at station 2, he is already in the process of making a cycle with respect to station 1, and has just started making a cycle with respect to station 2. Hence, the segment of time $(T_2 + W_3)$ overlaps for stations 1 and 3. That is why the second pair of integrals in (3.3) is computed jointly (for stations 1 and 2) as the probability of no arrivals at *both* stations during the segment of time $(T_2 + W_3)$. The rest of the integral product is explained in identical fashion. Note that W_j and W_j' are really different realizations of the same random variable, and hence the same distribution function is used for these in (3.3).

We next take a case-by-case approach in showing how an arbitrary transition probability $Pr[\mathbf{z}' \mid \mathbf{z}]$, $\mathbf{z}, \mathbf{z}' \in \Theta_3$, can be computed. Given any two vectors $\mathbf{z}, \mathbf{z}'$ in Θ_3, first obtain the distribution functions F_{T_j} and $F_{T_j'}$, for $1 \le j \le N$. Next, keep $\mathbf{z}$ fixed and compute $Pr[\mathbf{z}' \mid \mathbf{z}]$ for all possible vectors $\mathbf{z}'$, as explained below. The first step is to obtain the integral product on (3.3), i.e., for $\mathbf{z}' = <0, 0, 0>$. The other cases of $\mathbf{z}'$ are obtained in stepwise fashion.

Case 1*:* $\mathbf{z}'$ contains a single nonzero bit. Then

$$\begin{aligned}
Pr[0, 0, 1 \mid \mathbf{z}] &= Pr[0, 0, \cdot \mid \mathbf{z}] - Pr[0, 0, 0 \mid \mathbf{z}] \\
Pr[0, 1, 0 \mid \mathbf{z}] &= Pr[0, \cdot, 0 \mid \mathbf{z}] - Pr[0, 0, 0 \mid \mathbf{z}] \\
Pr[1, 0, 0 \mid \mathbf{z}] &= Pr[\cdot, 0, 0 \mid \mathbf{z}] - Pr[0, 0, 0 \mid \mathbf{z}]
\end{aligned}$$

Note that the above computations require that the conditional joint probability that any two out of three stations are found by the server to be nonempty. These are easily computed *using* (3.3) . For example, $Pr[0, 0, \cdot \mid \mathbf{z}]$ is obtained from (3.3) by setting $\lambda_3 = 0$.

Case 2*:* $\mathbf{z}'$ contains two nonzero bits. Then

$$\begin{aligned}
Pr[0, 1, 1 \mid \mathbf{z}] &= Pr[0, \cdot, \cdot \mid \mathbf{z}] - (Pr[0, 0, 0 \mid \mathbf{z}] + Pr[0, 0, 1 \mid \mathbf{z}] + Pr[0, 1, 0 \mid \mathbf{z}]) \\
Pr[1, 0, 1 \mid \mathbf{z}] &= Pr[\cdot, 0, \cdot \mid \mathbf{z}] - (Pr[0, 0, 0 \mid \mathbf{z}] + Pr[0, 0, 1 \mid \mathbf{z}] + Pr[1, 0, 0 \mid \mathbf{z}]) \\
Pr[1, 1, 0 \mid \mathbf{z}] &= Pr[\cdot, \cdot, 0 \mid \mathbf{z}] - (Pr[0, 0, 0 \mid \mathbf{z}] + Pr[0, 1, 0 \mid \mathbf{z}] + Pr[1, 0, 0 \mid \mathbf{z}])
\end{aligned}$$

The above computations require the marginal probability that any one station is found by the server to be nonempty upon his arrival at that station. This is easily obtained from (3.3). For example, to obtain $Pr[0, \cdot, \cdot \mid \mathbf{z}]$, we simple compute (3.3) by setting $\lambda_2 = 0$ and $\lambda_3 = 0$.

Case 3*:* $\mathbf{z}'$ contains only nonzero bits.

In this case, we are merely computing the complement of the the sum of all the probabilities computed in cases one and two. That is,

$$\begin{aligned}
Pr[1, 1, 1 \mid \mathbf{z}] = 1 - \{&Pr[0, 0, 0 \mid \mathbf{z}] + Pr[0, 0, 1 \mid \mathbf{z}] + Pr[0, 1, 0 \mid \mathbf{z}] + Pr[0, 1, 1 \mid \mathbf{z}] \\
&+ Pr[1, 0, 0 \mid \mathbf{z}] + Pr[1, 0, 1 \mid \mathbf{z}] + Pr[1, 1, 0 \mid \mathbf{z}]\}
\end{aligned}$$

Extension of the computation scheme to general N

For the case of general N, we can develop a similar scheme to compute transition probabilities for transitions from any given $\mathbf{z} \in \Theta_N$ to all possible $\mathbf{z}' \in \Theta_N$. In fact, the extension of (3.3) to the N station case is fairly straightforward. However, the number of cases to investigate is now N, i.e., case k would be the case in which the N-bit vector $\mathbf{z}'$ contained exactly k nonzero bits. Within case k, the number of distinct transition probabilities to be computed would be equal to the number of ways of choosing k bits from N bits without repetition, that is, $\begin{bmatrix} N \\ k \end{bmatrix}$.

Let $\mathbf{z} \in \Theta_N$ be fixed. In order to compute $Pr[\mathbf{z}' \mid \mathbf{z}]$ for any $\mathbf{z}' \in \Theta_N$, we must first develop a general expression for $Pr[0, 0, 0, ..., 0 \mid \mathbf{z}]$ (i.e., the N-station extension of (3.3)). We first introduce some compact notation. Define vectors $\Lambda^{(j)} = <\lambda_1, ..., \lambda_j>$, and $\Lambda_{(j)} = <\lambda_j, ..., \lambda_N>$, for each j, $1 \le j \le N$. Next, for generic (nonnegative) random variables X and Y, with distributions $F_X(\cdot), F_Y(\cdot)$, respectively, we define the joint integral products

$$D(\Lambda^{(j)}, X, Y) = \int_0^\infty e^{-(\sum_{k=1}^{j} \lambda_k)t} \, dF_X(t) \, . \int_0^\infty e^{-(\sum_{k=1}^{j} \lambda_k)t} \, dF_Y(t) \tag{3.4}$$

and

$$D(\Lambda_{(j)}, X, Y) = \int_0^\infty e^{-(\sum_{k=j}^{N} \lambda_k)t} \, dF_X(t) \, . \int_0^\infty e^{-(\sum_{k=j}^{N} \lambda_k)t} \, dF_Y(t). \tag{3.5}$$

For a fixed $\mathbf{z} \in \Theta_N$, we obtain the distributions F_{T_j}, $1 \le j \le N$. Before attempting to compute $Pr[\mathbf{z}' \mid \mathbf{z}]$ for an arbitrary $\mathbf{z}' \in \Theta_N$, we must first obtain the integral product

$$Pr[0, 0, \cdots, 0 \mid \mathbf{z}] = \{ \prod_{k=1}^{N} D(\Lambda^{(k)}, T_k, W_{(k \bmod N)+1}) \} \cdot \{ \prod_{i=1}^{N} D(\Lambda_{(i)}, T_i', W'_{(i \bmod N)+1}) \} \tag{3.6}$$

From the integral product in (3.6) it is possible to generate $Pr[\mathbf{z}' \mid \mathbf{z}]$ for all $\mathbf{z}' \in \Theta_N$, and a fixed $z \in \Theta_N$. The idea is essentially a generalization of what was done for the $N = 3$ case. For example,

$$Pr[0, 0, 0, ..., 1 \mid \mathbf{z}] = Pr[0, 0, ..., \cdot \mid \mathbf{z}] - Pr[0, 0, 0, ..., 0 \mid \mathbf{z}] \tag{3.7}$$

where $Pr[0, ..., \cdot \mid \mathbf{z}]$ is obtained from (3.6) by setting $\lambda_N = 0$. The the other transition probabilities are obtained similarly. Thus, to obtain the probability transition matrix for the case of general N, one only needs to obtain (3.6) in terms of an arbitrary $\mathbf{z} \in \Theta_N$.

Explicit form of the cycle-time distribution

Let the random time between two consecutive visits of the server at station N be denoted by the random variable C, and its distribution by $F_C(\cdot)$. Let $\mathbf{P}_E$ be the probability transition matrix for the 2^N state Markov chain $\{\mathbf{Z}_n\}$, as given by (3.6) and the associated sequence of computations. Let $\{\phi_{\mathbf{z}}\,;\, \mathbf{z} \in \Theta_N\}$ denote the invariant vector of $\mathbf{P}_E$. The exact cycle-time distribution $F_C(\cdot)$ on the asymmetric, unit-capacity buffer, cyclic-server multiqueue is given by

$$F_C(t) = \sum_{\mathbf{z} \in \Theta_N} \phi_{\mathbf{z}} [V_1 * \cdots * V_N * \{z_1 B_1 + (1 - z_1) S_1\} * \cdots * \{z_N B_N + (1 - z_N) S_N\}](t) \tag{3.7}$$

where the * is used to denote the convolution operation.

As an illustrative example, suppose that the distributions $B_j(\cdot)$, $S_j(\cdot)$ and $V_j(\cdot)$ are all exponential, with means $1/\mu_j$, $1/\beta_j$, and $1/\alpha_j$, respectively, for $1 \le j \le N$. Further, let $a_i = \alpha_i$ and $a_{N+i} = z_i \mu_i + (1 - z_i) \beta_i$ for $1 \le i \le N$. In this case, (3.7) can be explicitly written as

$$F_C(t) = \sum_{\mathbf{z} \in \Theta_N} \phi_{\mathbf{z}} \left[\sum_{j=1}^{2N} \xi_j(\mathbf{z})\, a_j\, e^{-a_j t} \right] \tag{3.8}$$

where $\xi_j(\mathbf{z}) = \prod_{\substack{k=1 \\ k \neq j}}^{N} \frac{a_k}{a_k - a_j}$, $1 \le j \le 2N$. The distribution in the squared brackets is the generalized Erlangian distribution.

3.3 Closed network and other cycle-time approximations

In this section we derive an expression for the cycle-time distribution of the single customer on the closed network dual of an asymmetric, unit capacity-buffer, cyclic-server multiqueue system. Even though the cycle-time random variable C_d of the single customer on the closed network dual was explicitly given in (3.1), the distribution F_d was not because the routing probabilities p_j, $1 \le j \le N$ are yet to be determined. Further, the machinery of section 3.2 is required for these probabilities.

The marginal probability that station j is found nonempty when the cyclic-server arrives at station j in the asymmetric, unit-capacity buffer multiqueue can be obtained as

$$p_j \;=\; \sum_{\{\mathbf{z} \in \Theta_N : z_j = 1\}} \phi_{\mathbf{z}} \tag{3.9}$$

where $\{\phi_{\mathbf{z}} ; \mathbf{z} \in \Theta_N\}$ is the invariant vector of the probability transition matrix $\mathbf{P}_E$ (see section 3.2). The fact that C_d is merely a sum of independent random variables means that we are computing the cycle-time of the single customer on the closed network as the sum of all his sojourn times at the $2N$ stations he must visit to complete a cycle. If this type of cycle-time is used to *approximate* the cycle-time of the serve on the cyclic-server multiqueue system, then in effect, we are assuming that the random variable C (i.e., the exact cycle-time random variable of the server on the multiqueue) can be approximated by *assuming that events at the* N multiqueueing stations are mutually independent. This assumption of *station-independence* is precisely that which yields the closed network dual representation for the cyclic-server multiqueue. An application of the station independence assumption in different forms can be found in [Lieb62], [HaOh72], and [Kueh79].

Closed Network or Station Independence Cycle-time Approximation

Once p_j is known, $1 \le j \le N$, the distribution F_d is easily obtained. Let $\{\kappa_{\mathbf{z}} ; \mathbf{z} \in \Theta_N\}$ be defined by

$$\kappa_{\mathbf{z}} = \prod_{j=1}^{N} \left[z_j p_j + (1-p_j)(1-z_j) \right] \tag{3.10}$$

for each $\mathbf{z} \in \Theta_N$. The probability $\kappa_{\mathbf{z}}$ is the steady state probability that the observer sees the vector $\mathbf{z}$ under the station independence assumption. In effect, (3.10) is using the assumption of station independence to describe the limiting service-vector seen by the observer on the multiqueue system. Finally, the distribution F_d is explicitly given by

$$F_d(t) = \sum_{z \in \Theta_N} \kappa_z [V_1 * \cdots * V_N * \{z_1 B_1 + (1-z_1) S_1\} * \cdots * \{z_N B_N + (1-z_N) S_N\}](t). \quad (3.11)$$

If the distributions $B_j(\cdot)$, $S_j(\cdot)$ and $V_j(\cdot)$ are all exponential, with means $1/\mu_j$, $1/\beta_j$, and $1/\alpha_j$, respectively, for $1 \le j \le N$, and we define $a_i = \alpha_i$ and $a_{N+i} = z_i \mu_i + (1-z_i)\beta_i$ for $1 \le i \le N$, then (3.11) can be written as

$$F_C(t) = \sum_{z \in \Theta_N} \kappa_z \left[\sum_{j=1}^{2N} \xi_j(z) a_j e^{-a_j t} \right] \quad (3.12)$$

The duality can now be established through the comparison of (3.7) and (3.11).

Customer or Packet Independence Cycle-Time Approximation

Besides the closed network (station independence) approximation (i.e., equation (3.11)) to the exact distribution of the cyclic-server on the multi-queue (i.e., equation (3.5)), we demonstrate how another interesting approximation arises. Consider an $N = 3$ station cyclic-server multiqueue, and the system of cycle-times given in equation (3.2). If we assume that the random variables C_1, C_2, and C_3 are mutually independent, we are effectively assuming that each random variable T_j and W_j (or T_j' and W_j') appearing in two different C_i's, $1 \le i \le 3$, are independent. In other words, each station sees the server's realization of a cycle-time to be *independent* of any other station's realization of a cycle-time. This assumption is equivalent to the assumption that customers' service times (or server's walk times) times at a given station can be arbitrarily drawn from the service-time distribution (or walk time distribution) without regard to the particular cycle in progress. If we model the customers on our multiqueue as packets on a token passing system, this assumption would be very similar to the assumption of packet independence made by Kleinrock in the Arpanet models [Klei76].

As an example of the computational simplifications introduced by the customer (or packet) independence assumption, consider an $N = 3$ station cyclic-server multiqueue. Using the customer (or packet) independence idea, equation (3.3) becomes

$$
\begin{aligned}
Pr[0,0,0 \mid \mathbf{z}] = \quad & \int_0^\infty e^{-\lambda_1 t}\, dF_{T_1}(t) \cdot \int_0^\infty e^{-\lambda_1 t}\, dV_2(t) \\
& \left[\prod_{j=1}^{2} \int_0^\infty e^{-\lambda_j t}\, dF_{T_2}(t)\right] \left[\prod_{j=1}^{2} \int_0^\infty e^{-\lambda_j t}\, dV_3(t)\right] \\
& \left[\prod_{j=1}^{3} \int_0^\infty e^{-\lambda_j t}\, dF_{T_3}(t)\right] \left[\prod_{j=1}^{3} \int_0^\infty e^{-\lambda_j t}\, dV_1(t)\right] \qquad (3.13) \\
& \left[\prod_{j=2}^{3} \int_0^\infty e^{-\lambda_j t}\, dF_{T_1}'(t)\right] \left[\prod_{j=2}^{3} \int_0^\infty e^{-\lambda_j t}\, dV_2(t)\right] \\
& \int_0^\infty e^{-\lambda_3 t}\, dF_{T_2}(t) \cdot \int_0^\infty e^{-\lambda_3 t}\, dV_3(t)
\end{aligned}
$$

from which we can construct a probability transition matrix for $N = 3$. The extension to general N follows just as was explained in section 3.2.

For an N station asymmetric cyclic-server multiqueue with unit-capacity buffers, let $\mathbf{P}_P$ denote the probability transition matrix obtained by using customer (or packet) independence. Let $\{\pi_{\mathbf{z}}\,;\, \mathbf{z} \in \Theta_N\}$ denote its invariant vector. The approximate cycle-time distribution $F_p(\cdot)$ obtained via customer independence is given by

$$F_p(t) = \sum_{\mathbf{z} \in \Theta_N} \pi_{\mathbf{z}} [V_1 * \cdots * V_N * \{z_1 B_1 + (1 - z_1) S_1\} * \cdots * \{z_N B_N + (1 - z_N) S_N\}](t) \quad (3.14)$$

If the distributions $B_j(\cdot)$, $S_j(\cdot)$ and $V_j(\cdot)$ are all exponential, with means $1/\mu_j$, $1/\beta_j$, and $1/\alpha_j$, respectively, for $1 \le j \le N$, and we define $a_i = \alpha_i$ and $a_{N+i} = z_i \mu_i + (1 - z_i) \beta_i$ for $1 \le i \le N$, then (3.11) can be written as

$$F_p(t) = \sum_{\mathbf{z} \in \Theta_N} \pi_{\mathbf{z}} \left[\sum_{j=1}^{2N} \xi_j(\mathbf{z})\, a_j\, e^{-a_j t} \right] \qquad (3.15)$$

where $\xi_j(\mathbf{z})$ is defined after Eq. (3.8).

4. COMPUTATIONAL RESULTS AND APPLICATIONS

For a system of N queueing stations, the complexity of computing the cycle-time distribution using P_E is $O(2^N)$. If packet independenced is used the complexity still remains as $O(2^N)$, but the constant involved is reduced due to the simplification in computing (5.3) over (4.2). If unit capacity buffers are used, then station independence *also* requires $O(2^N)$ complexity since the distribution $\{\kappa_z ; z \in \Theta_N\}$ depends on $\{\phi_z ; z \in \Theta_N\}$. However, if infinite capacity buffers are used, then the complexity of computation via station independence becomes $O(N)$, since the mean cycle-time is directly obtainable from (3.2), and the probability p_j that the server finds queue j not empty upon arrival there is simply $p_j = \lambda_j E(C_j)$ and $E(C_j)$ is easy to compute as shown in [Kueh79].

For the unit buffer capacity situation, the largest amount of computation is required for P_E. Observe that when $N = 10$, P_E requires a matrix of size 1024 x 1024, thus implying more than a million computations to obtain the matrix. It appears that our computations are reasonable for smaller systems, i.e., $N \leq 9$.

4.1 Numerical results

In this section we obtain (computationally) conditions under which station independence yields cycle-time distributions that are reasonably good approximations to exact cycle-time distributions. In other words, we establish conditions under which duality between the cyclic-server multiqueue model and the closed network model is strong. For completeness, we also include the results of computation under the assumption of customer (or packet) independence.

In a series of graphical comparisons (see Figures 3a through 4.2d) we compare the limiting distributions of service-vectors and cycle-times obtained via our three methods. Using $N = 3$, each graph compares

(1) the exact cycle-time distribution of the server on the multiqueue (labelled E)

(2) the cycle-time distribution of the single customer on the closed-network dual, that is, the cycle-time distribution given by station independence (labelled D), and

(3) the cycle-time distribution obtained via the customer (or packet) independence assumption (labelled P).

We group the figures into four different categories. Those in the "a" category (Figures 3a, 4.1a, 4.2a) refer to a system with moderate traffic (i.e., load) . The "b" category (Figures 3b, 4.1b, and 4.2b) refer to a system with higher-than-moderate traffic. The "c" category (Figures 3c, 4.1c, and 4.2c) refers to a system with still higher traffic. Finally, the "d" category refers to a system with lower-than-moderate traffic. The moderate-traffic case (loosely defined as traffic that causes cycles in which no stations transmit packets to occur roughly as frequently as cycles in which all stations transmit packets) is of special importance as will be seen shortly.

With the exception of λ_1, λ_2 and λ_3, all other input parameters are kept fixed for all computations. The λ_j are varied to obtain different traffic levels, $j = 1, 2, 3$. The distributions $B_j(\cdot)$, $S_j(\cdot)$, and $V_j(\cdot)$ are all taken to be exponential distributions, $1 \le j \le 3$. The packet transmission times (customer service-times) are taken to have a mean of 100, 300, and 200 time-units for stations 1, 2, and 3, respectively. The mean switching times at stations 1, 2 and 3 are taken to be 1/10, 1/30, and 1/20 time-units, respectively. Finally, the mean walk-time from station 1 to station 2 is 1 time-unit, the mean walk-time from station 2 to station 3 is 3 time-units, and from station 3 to station 1 this time is 2 time-units.

In Figures 3a through 3d can be seen the steady state probability that the observer at station 3 is in any given state of the set Θ_3. That is, a "0" on the z- axis corresponds to the state "000", a "1" corresponds to the state "001", etc. The continuous line connects points representing the exact vector distribution, the circles represent the vector distribution given by the closed-network dual (i.e., station independence), and the squares give the distribution obtained via customer (or packet) independence. In Figure 3a (i.e., moderate traffic), observe that the customer independence assumption performs better than the station independence assumption. More importantly, for moderate traffic, the closed-network dual yields a vector distribution that is unimodal, while the other two methods yield multimodal vector distributions. When traffic is increased (se Figs. 3b and 3c), the three distributions converge, with customer independence consistently outperforming station independence. In these two cases, all distributions are multimodal. When traffic is very low (see Fig. 3d), both independence assumptions can be seen to perform very well. Once again, all three distributions are seen to be multimodal. The discrete distributions indicate that the duality between a multiqueue system and a closed network tends to be weak for moderate traffic and strong for low and high traffic.

In Figures 4.1a and 4.2a can be seen the cycle-time distribution obtained via the three different methods, for moderate traffic. The distribution is actually broken up into two pieces in each traffic situation, in order to give a clear picture demonstrating the presence of bimodality. Figure 4.1a shows the cycle-time distribution for cycles of length less than 40 time-units. Correspondingly, Figure 4.1b shows the distribution for cycles of length greater than 40 time-units. The bimodal nature of the exact cycle-time distribution at moderate load may be surprising, at first glance. The multimodal vector distribution (show in Fig 3a) plays a part in this. In effect, this says that for a system in which buffers are able to hold only one packet (customer) at a time, the cycle-time of the server can take one certain large values (in the range 200 - 400 time-units) with a higher probability than other large values. We believe that the effect is largely due to the effect of unit-capacity buffers. The provision of more buffer space for arriving packets is likely to have a smoothing effect on the cycle-time distribution, possibly removing such bimodal behaviour.

In Figures 4.1b and 4.2 b, the cycle-time distributions are displayed for higher-than-moderate load. In Figure 4.1c and 4.2c, the cycle-time distribution is displayed for even higher traffic, and in Figure 4.1d and 4.2d we see the results for very low traffic. Clearly, at high and low traffic situations, the duality between the multiqueue system and the closed network system appears to be strong, especially for large cycle-times. It is interesting to note that for very low loads (Figure 4.1d and 4.2d), and correspondingly, for very high loads, both approximating assumptions yield a distribution that is extremely close to the exact distribution. We thus conclude that duality is strongest at extreme loads. Further, our method of computing the cycle-time distribution yields numerical values that tell how well the duality relationship performs in terms of closeness to the exact result. Observe that for moderate traffic, th duality result gives only a crude approximation to the cycle-time distribution. In *all* traffic situations, the closed-network cycle-time distribution lacks the bimodality that is characteristic of the exact (and even the approximate, via customer-independence) cycle-time distribution.

4.2. Applications

In the previous section, it was shown that the closed-network dual of the multiqueue yields better approximations as system loads become more extreme (i.e., higher, or lower). This implies that the entire theory of closed-networks may be applied to analyze token-passing systems to obtain useful pieces of information at these loads. As an example, we demonstrate how one can compute the position of a token on a token ring (or token bus). The method presented here is not restricted to a single token. Multiple tokens on a token passing system can be handled without any additional work.

Let us assume that there are K tokens on a token-passing system. By duality, we have K customers in the closed network system. Referring to Figure 2, we note that the closed network consists of 3N queues; N queues are associated with walking times, N queues with switching times, N queues correspond to "real" service times. For $i = 1, 2, \ldots, N$, let k_{3i-2} denote the number of customers (i.e., tokens from the multiqueue are customers in the closed network) in the walking-phase at the $(3i-2)^{nd}$ queue, let k_{3i-1} denote the number of customers in the switching-phase at the $(3i-1)^{th}$ queue, and let k_{3i} denote the number of customers in the service-phase at the $3i^{th}$ queue. The closed network is described by a $3N$-dimensional vector $\mathbf{k} = (k_1, k_2, k_3, \ldots, k_{3N})$ with the probability of being in state $\mathbf{k}$ denoted by $p(\mathbf{k}) = p(k_1, k_2, \ldots, k_{3N})$. We refer to a walking queue followed by two parallel queues, namely a switching queue and a service queue, as a *node*. Such a node also corresponds to a node in the token-passing dual network. The j^{th} node of our closed network is shown in Figure 5. The branching probability p_j in the closed-network is equivalent to the probability that the j-th queue in the token-passing dual is found *not* empty on the server's arrival. The walking time, switching time and service time at the j^{th} node are assumed to be exponentially distributed random variables with means α_j, β_j and μ_j respectively.

Using standard techniques for closed networks [Klei76], the steady state probability $p(k_1, \ldots, k_{3N})$ is

$$p(k_1, \ldots, k_{3N}) = \frac{1}{G(K)} \prod_{j=1}^{3N} x_j^{k_j}, \tag{4.1}$$

where $k_1 + k_2 + \cdots + k_{3N} = K$, and $G(K)$ is a normalizing constant. The x's are computed as follows. For $i = 1, \ldots, N$ we find

$$x_{3i-2} = \frac{1}{\alpha_i}, \quad x_{3i-1} = \frac{\bar{p}_i}{\beta_i}, \quad x_{3i} = \frac{p_i}{\mu_i}. \tag{4.2}$$

Hence, (4.1) and (4.2) imply

$$p(k_1, \ldots, k_{3N} = \frac{1}{G(K)} \prod_{i=1}^{N} \frac{\bar{p}_i^{k_{3i-1}} \, p_i^{k_{3i}}}{\alpha_i^{k_{3i-2}} \, \beta_i^{k_{3i-1}} \, \mu_i^{k_{3i}}} \tag{4.3}$$

where $\bar{p}_i = 1 - p_i$. For the unit capacity buffer system the probability p_i is given by (3.9), while for the infinite capacity buffer model $p_i = \lambda_i \, EC_i$, where EC_i is the average cycle time as it is computed in [Kueh 79].

In the case of a single token the formula (4.3) becomes simpler since now $k_1 + k_2 + \cdots + k_{3N} = 1$ and all k's except one are equal to zero. Let $p(0, 0, \ldots, 1^j, \ldots, 0)$ denote the probability that the token is at queue j, $j = 1, 2, , \ldots, 3N$. Then for $i = 1, \ldots, N$ we obtain

$$p(0, 0, \ldots, 1^j, \ldots, 0) = \begin{cases} \dfrac{1}{G(1)\alpha_j} & \text{for} \quad j = 3i - 2 \\ \dfrac{\bar{p}_j}{G(1)\beta_j} & \text{for} \quad j = 3i - 1 \\ \dfrac{p_j}{G(1)\mu_j} & \text{for} \quad j = 3i \end{cases} \tag{4.4}$$

In particular, using (4.3) and (4.4) we can compute the position of the token in the token-passing system, that is, the probability that a token is currently at a given node. Note that a node in the token-passing system corresponds to three queues in the closed network dual system, namely "walking" queue, "switching" queue and "service" queue. Let $p(l_1, \ldots, l_N)$ be the probability that l_i tokens are at node i. Then, by (4.3) we find that

$$p(l_1, l_2, \ldots, l_N) = \sum_{\{\mathbf{l} \in A\}} p(k_1, k_2, \ldots, k_{3N}),$$

where the set A is defined as $A = \{\mathbf{l}: l_i = k_{3i} + k_{3i-1} + k_{3i-2};\ i = 1, 2, \ldots, N\}$.

The duality property just established has some more potential advantages especially in the case of multiple tokens. For example, even under station independence, it is not clear how to compute the cycle time distribution for multiple-token systems. Using our dual system (closed network approach) the cycle time distribution might be relatively easy to compute. We need to replace the two parallel queues by a single queue with hyperexponential service time. Then, the closed network becomes a cyclic queueing model. Much work has been done in the past on the cycle time distributions in closed cyclic queueing networks. For example, the recent work of Daduna, i.e., [Dadu 86a], [Dadu 86b] and Boxma [Boxm85] are the most promising.

5. COCLUSIONS

We have shown that known results for closed queueing networks can be used to obtain sound approximations in asymmetric multiqueueing problems. In this effort, we solved two open problems to test the notion of duality between closed queueing networks and multiqueueing systems. From our computational experiments, it appears that the duality can be strong for a wide range of interesting parameter values. Additionally, when using closed network results to approximate multiqueueing systems, the strength of the approximations can often be tested by testing the strength of the duality relationship. Ongoing research examines more general systems (i.e., finite and infinite buffers, more general service patterns, etc.). Particularly appealing features of the closed-network approximation include testable accuracy, simplicity, and small computational requirements in comparison to exact results.

References

[Boxm84] O.J. Boxma, "Two Symmetric Queues with Alternating Service and Switching Times," Performance '84, Ed. Gelenbe (North-Holland, Amsterdam), pp. 409–431,1984.

[Boxm85] O.J. Boxma, "Response times in cyclic queues - The influence of the slowest server", Tech. Rep., No. 380, University of Utrecht, 1985

[BuTr83] W. Bux and H. Truong, "Token-ring Performance: Mean Delay Approximation," *Proc. 10th Teletraffic Cong.*, pp. 3.3.3.1-1–3.3.3-7, June 1983.

[BuxW81] Werner Bux, "Local Area Subnetworks: A Performance Comparison," *IEEE Trans. Commun.*, Vol. COM-29, pp. 1465–1473, October 1981.

[Chow80] W. Chow, "The Cycle Time Distribution of Exponential Cyclic Queues," *J.A.C.M.* Vol. 27, No. 2, pp. 281–286, April 1980.

[CoBo83] J. Cohen and O. Boxma, "Boundary Value Problems in Queueing Systems Analysis," North-Holland, Amsterdam, 1983.

[Dadu86a] H. Daduna, "Cycle-Times in two-staged closed queueing networks: Applications to multiprogrammed computer systems with virtual memory," *Oper. Res*, 34, pp. 281-288, 1986.

[Dadu86b] H.Daduna, "Two-stage cyclic queues with non-exponential servers: Steady-state and Cycle-time," *Oper. Res*, 34, pp. 455-459, 1986.

[FeAm85] M.J. Ferguson and M. Aminetzah, "Exact Results for Nonsymmetric Token Ring Systems," *IEEE Trans. Commun.*, Vol. COM-33, pp. 223–231, March 1985.

[Ferg86] M.J. Ferguson, "Computation of the Variance of the Waiting Time for Token Rings," *IEEE Journal on Selected Areas in Communications*, Vol. SAC-4, No. 6, September 1986.

[Fuhr85] S.W. Fuhrmann, "Symmetric Queues Served in Cyclic Order," *Operations Research Letters*, October 1985.

[HaOh72] O. Hashida and K. Ohara, "Line Accommodation Capacity of a Communication Control Unit," Review of the Electr. Comm. Labs., Nippon Telegraph and Telephone Public Corp., Vol. 20, pp. 231–239, 1972.

[IEEE84a] IEEE Project 802, Local Area Network Standards, *Draft IEEE 802.5, Token Ring Access Method and Physical Layer Specification*, 1984.

[IEEE84b] IEEE Project 802, Local Area Network Standards, *Draft IEEE 802.4, Token Passing Bus Access Method and Physical Layer Speficications*, 1984.

[Kaye72] A.R. Kaye, "Analysis of a Distributed Control Loop for Data Transmissions," P.I.B. International Symposium XXII on Computer Communications and Teletraffic, April 4–6, 1972.

[KiTa86] G. Kimura and Y. Takahashi, "Diffusion Approximation for a Token Ring System with Nonexhaustive Service," *IEEE Journal on Selected Areas in Communications*, Vol. SAC-4, No. 6, September 1986.

[Klei76] L. Kleinrock, *Queueing Systems*, Vol. 2, Wiley, New York, 1976.

[Kueh79] P. J. Kuehn, "Multiqueue systems with Nonexhaustive Cyclic Service," *B.S.T.J.*, 58, No. 3,pp. 671-698, March 1979.

[KoMe74] A. Konheim and B. Meister," Waiting Lines and Times in a System with Polling," *J.A.C.M.*, 21, No. 3, pp. 470–490, July 1974.

[Mack57] C. Mack, "The Efficiency of N Machines Uni-directionally Patrolled by One Operative When Working Time is Constant and Repair Times are Variable," *J. Royal Stat. Soc.*, Series B, Vol. 19, pp. 173–178, 1957.

[MaMW57] C. Mack, T. Murphy, and N. Webb, "The Efficiency of N Machines Uni-directionally Patrolled by One Operative When Walking Time and Repair Times Are Constants," *J. Royal Stat. Soc.*, Series B., Vol. 19, pp. 166–172, 1957.

[MoWa84] R.J.T. Morris and Y. Wang, "Some Results for Multi-Queue Systems with Multiple Cyclic Servers," Performance '84, Ed. H. Rudin and W. Bux, North-Holland, Amsterdam, pp. 245–257, 1984.

[PaDo86] J. Pang and R. Donaldson, "Approximate Delay Analysis and Results for Asymmetric Token-Passing and Polling Networks," *IEEE Journal on Selected Areas in Communications*, Vol. SAC-4, No. 6, September 1986.

[ReNi86] V. Rego and L. M. Ni, "Analytic modelling of Cyclic-Service Systems and their application to Token-passing networks," submitted for publication, 1986.

[RuDe83] I. Rubin and L. DeMoraes, "Message Delay Analysis for Polling and Token Multiple-access Schemes for Local Communication Networks," *IEEE Journal on Selected Areas in Communications*, Vol. SAC-1, No. 5, pp. 935–947, November 1983.

[ScDa83] R. Schassberger and H. Daduna, "The Time for a Round Trip in a Cycle of Exponential Queues," *J.A.C.M.*, Vol. 30, No. 1, pp. 146–150, January 1983.

[Swar80] G. Boyd Swartz, "Polling in a Loop System," *J.A.C.M.*, Vol. 27, No. 1, pp. 42–59, January 1980.

[Taka86] H. Takagi, Analysis of polling systems, MIT Press, 1986.

[TeTH86] T. Takine, Y. Takahashi, and T. Hasegawa, "Performance Analysis of a Polling System with Single Buffers and Its Application to Interconnected Networks," *IEEE Journal on Selected Areas in Communications*, Vol. SAC-4, No. 6, September 1986.

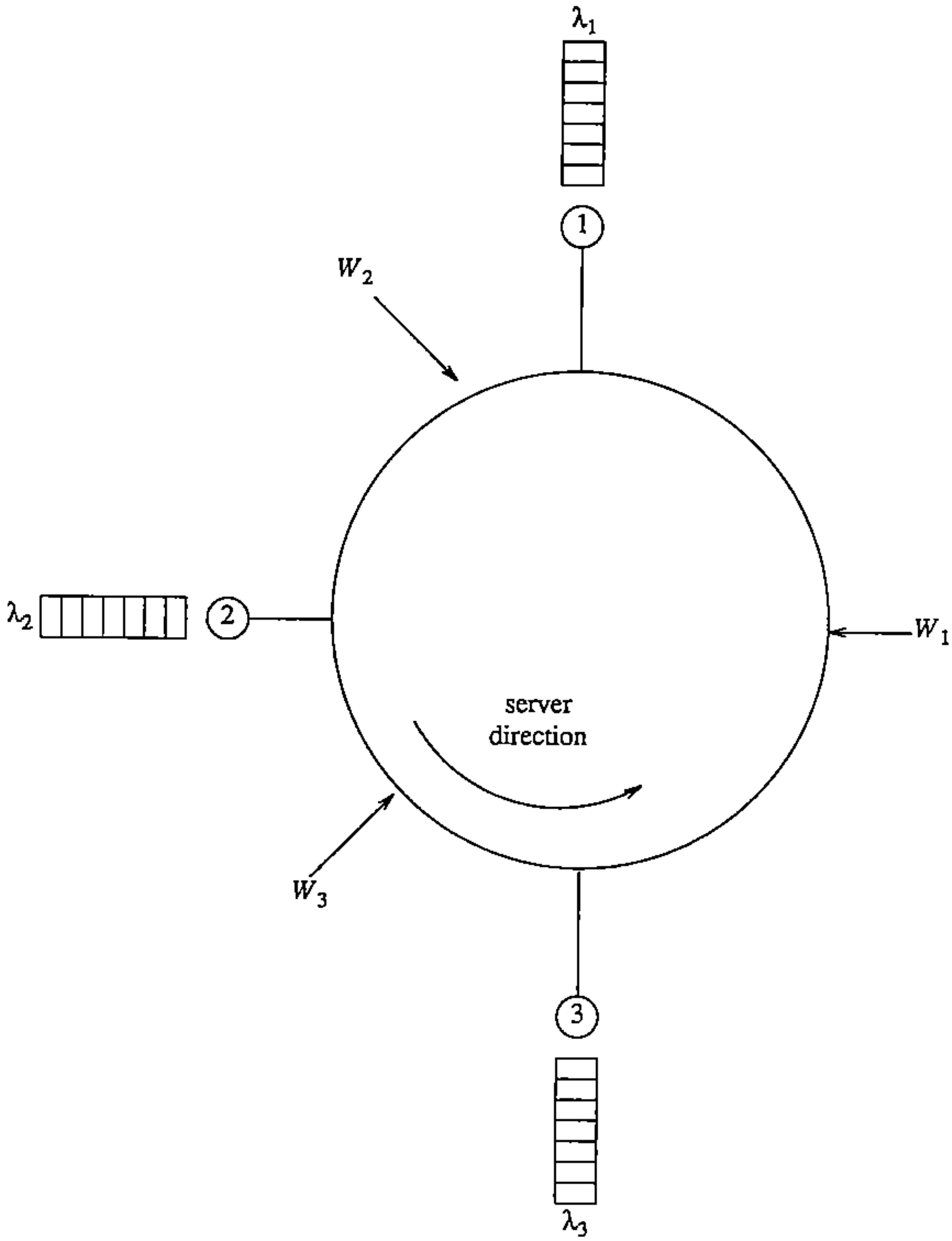

Figure 1a. Multiqueue and cyclic server (or token ring model), $N = 3$.

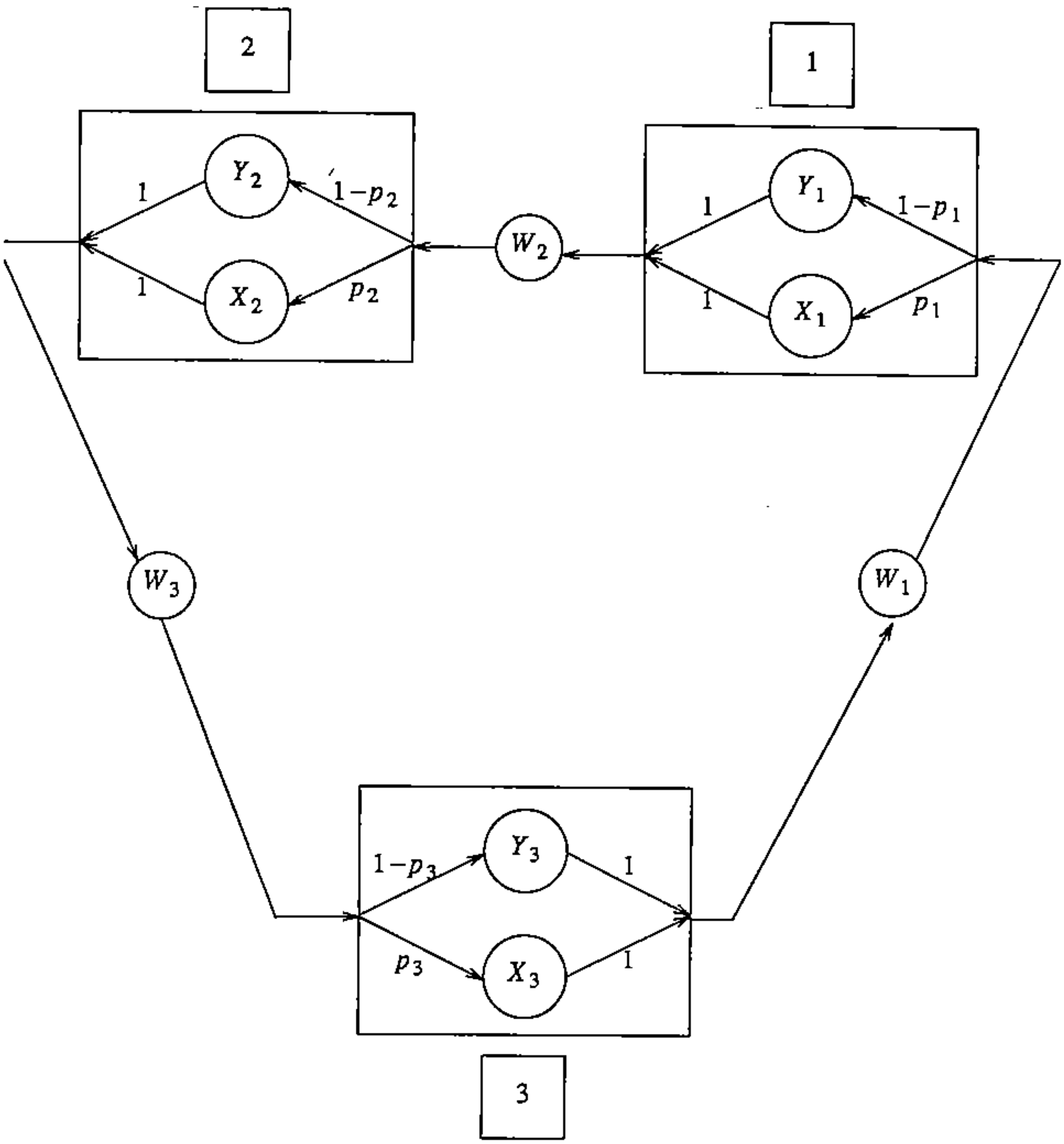

Figure 1b. Closed network representation (dual) of multiqueue and cyclic server.

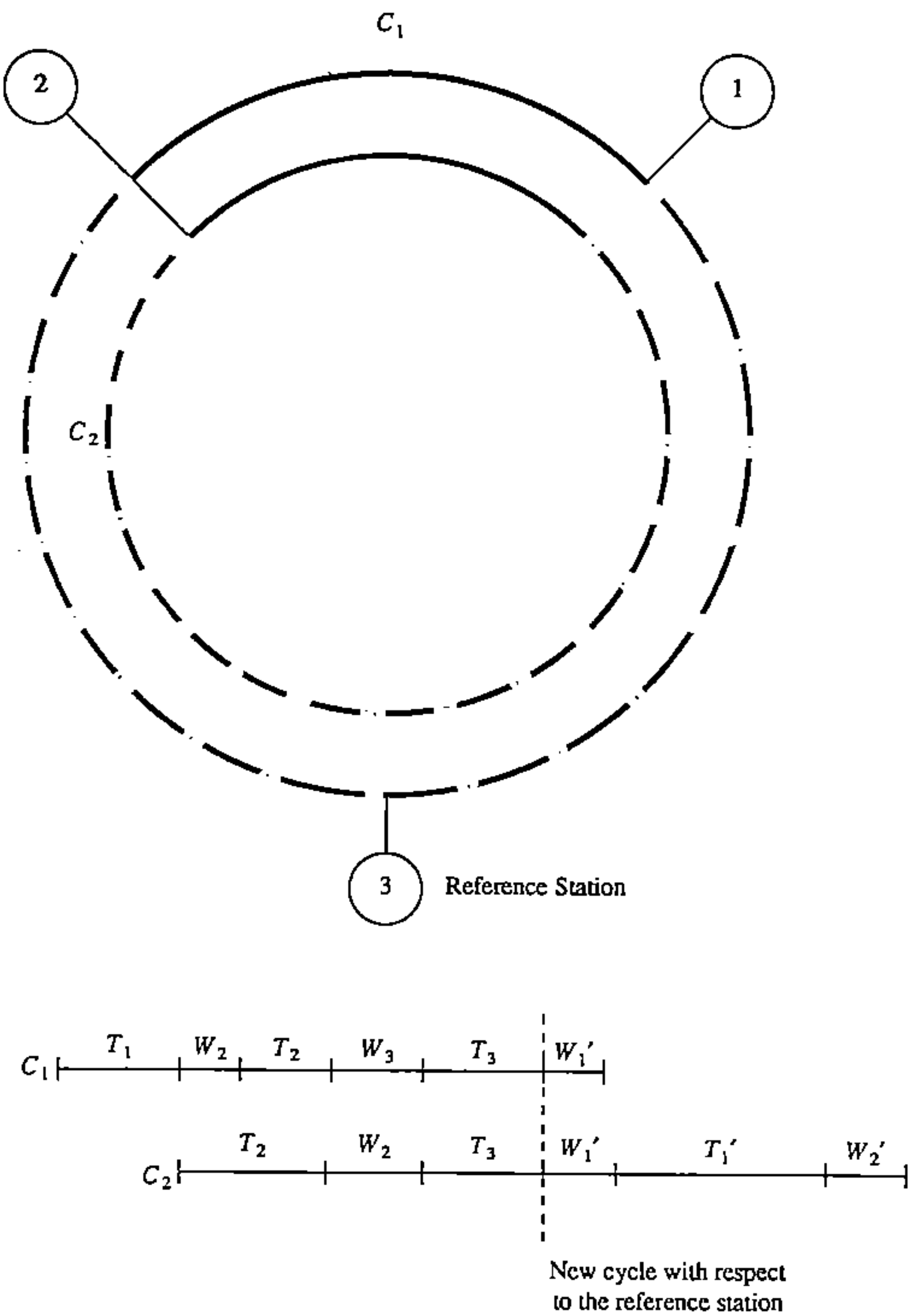

Figure 2. Illustration of cycle-times C_1 and C_2.

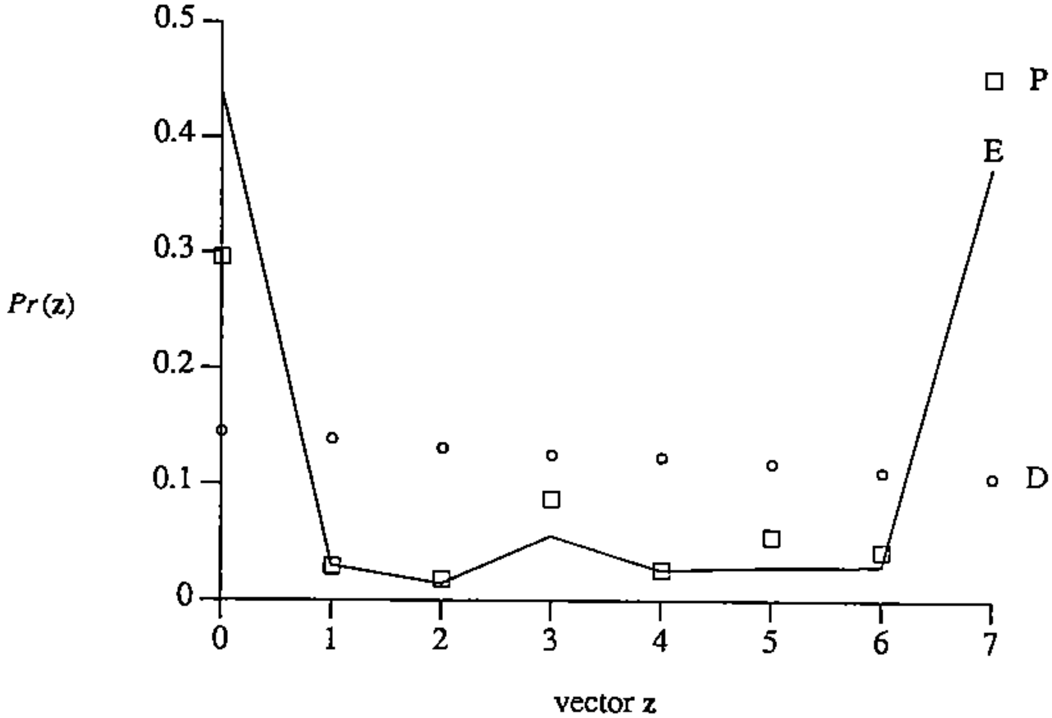

Figure 3a (Limiting vector distribution)

Moderate Traffic ($\lambda_1=0.005$, $\lambda_2=0.006$, $\lambda_3=0.007$)

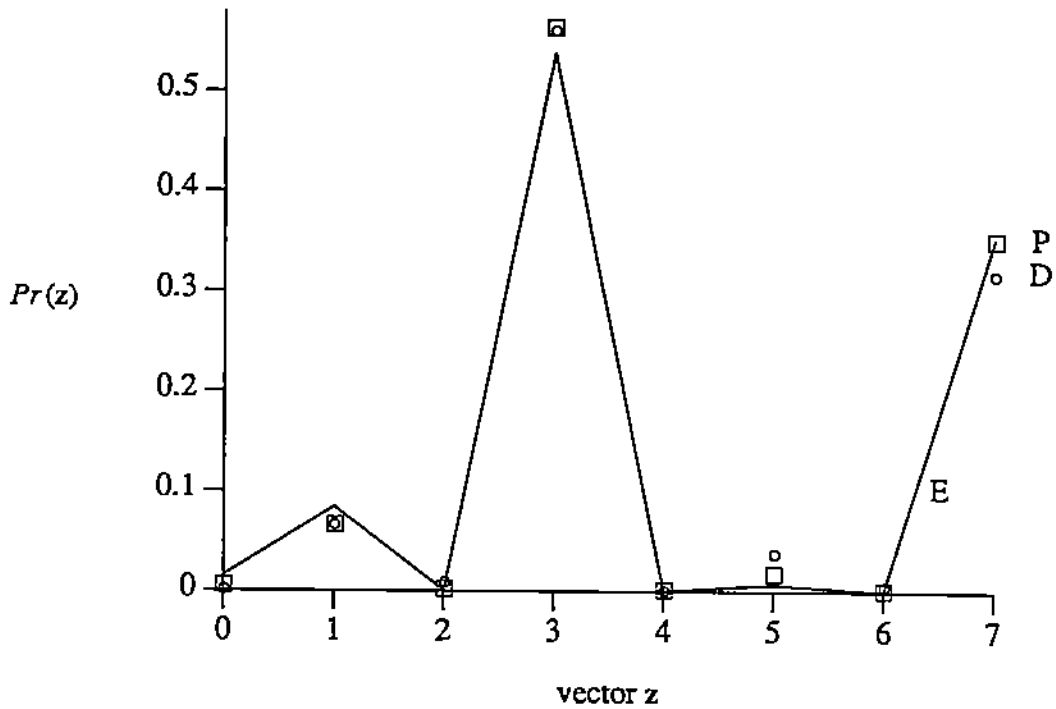

Figure 3b (Limiting vector distribution)

Higher Traffic ($\lambda_1=0.001$, $\lambda_2=0.01$, $\lambda_3=0.1$)

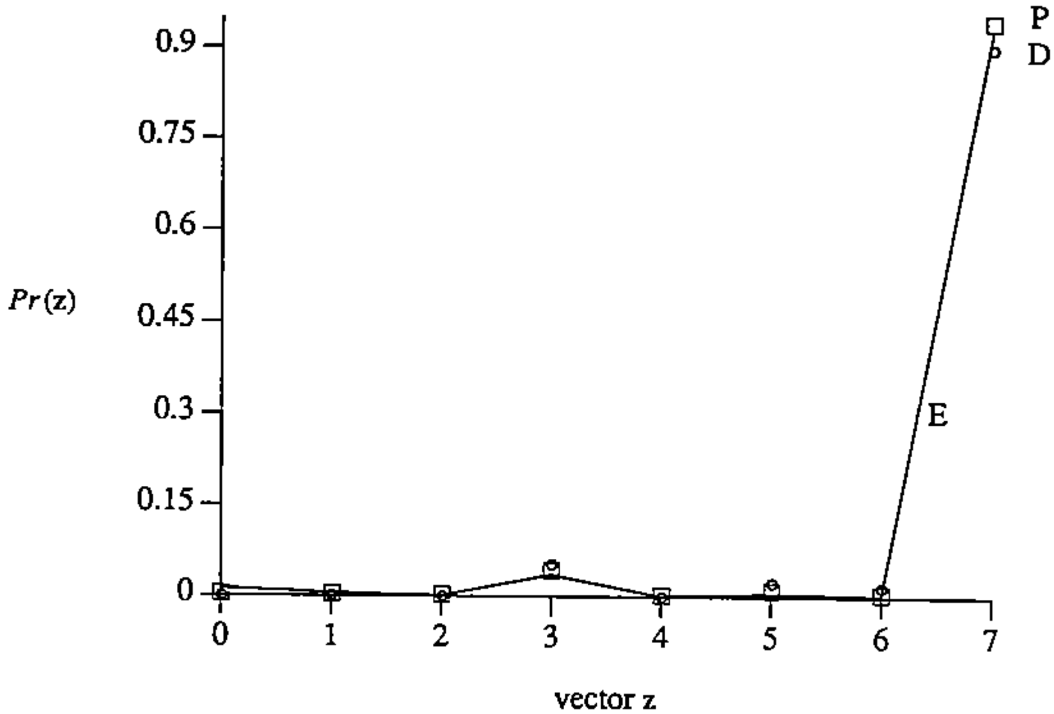

Figure 3c (Limiting vector distribution)

Very High Traffic ($\lambda_1=0.01$, $\lambda_2=0.02$, $\lambda_3=0.03$)

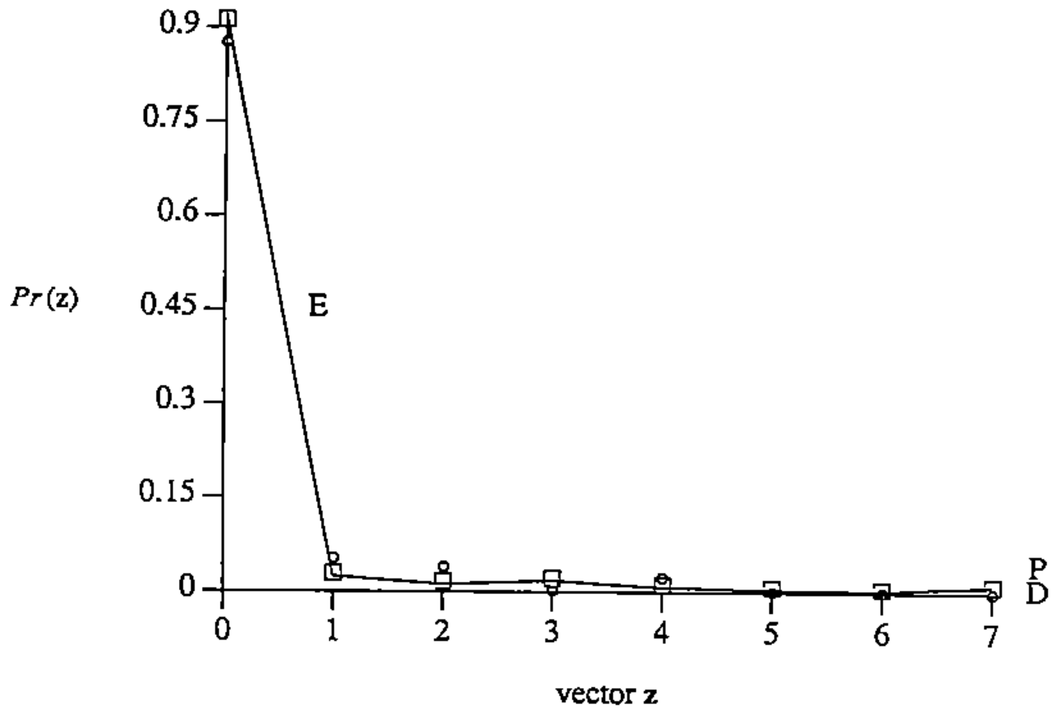

Figure 3d (Limiting vector distribution)

Very Low Traffic (λ_1=0.001 , λ_2=0.002, λ_3=0.003)

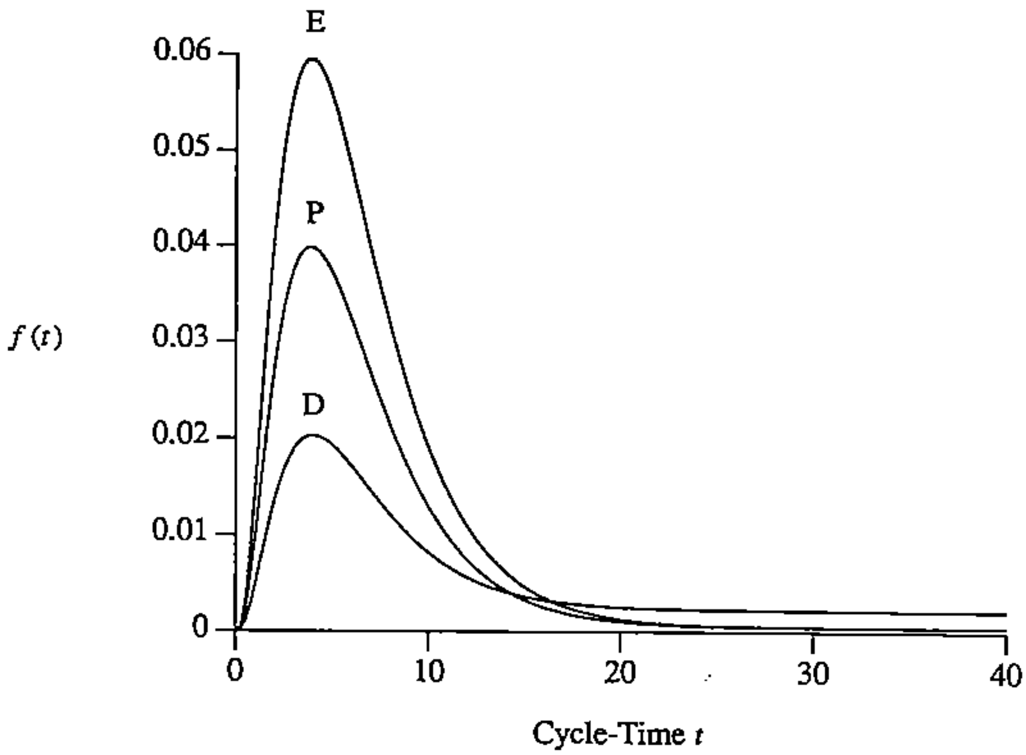

Figure 4.1a (Cycle-Time density)

Moderate Traffic ($\lambda_1=0.005$, $\lambda_2=0.006$, $\lambda_3=0.007$)

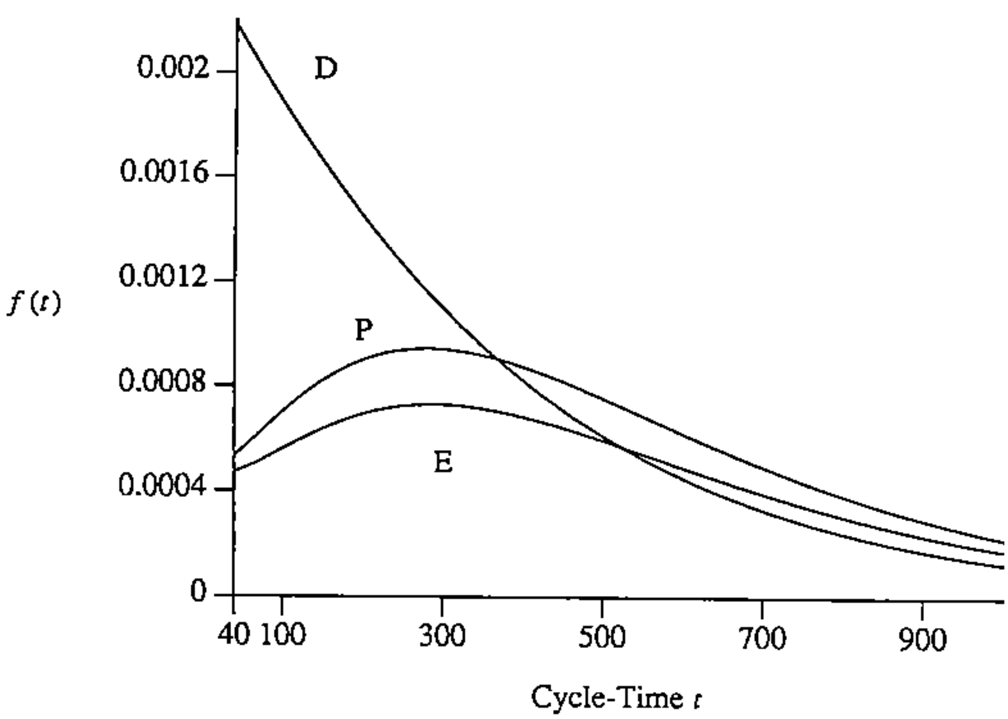

Figure 4.2a (Cycle-Time density)

Moderate Traffic ($\lambda_1=0.005$, $\lambda_2=0.006$, $\lambda_3=0.007$)

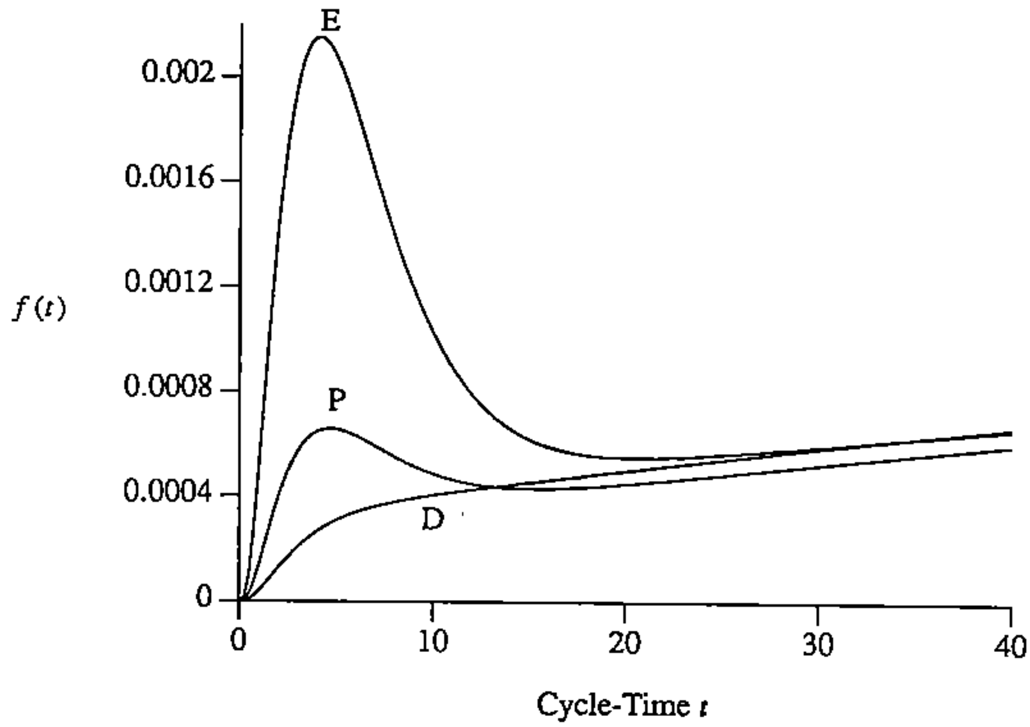

Figure 4.1b (Cycle-Time density)

Higher Traffic (λ_1=0.001 , λ_2=0.01, λ_3=0.1)

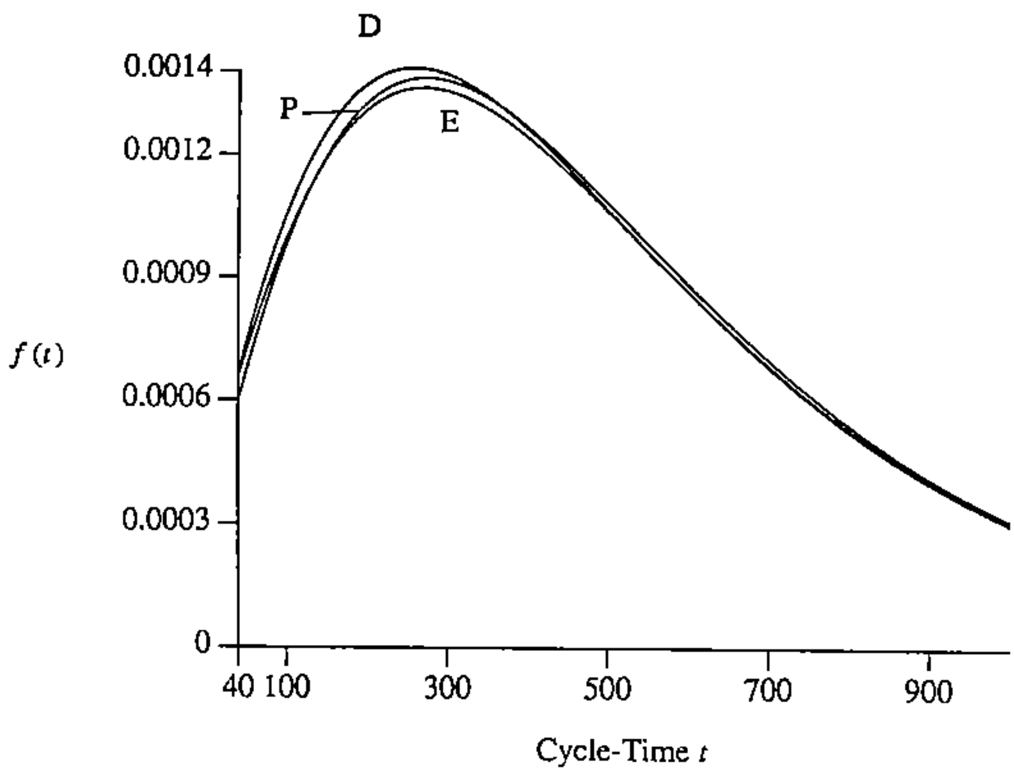

Figure 4.2b (Cycle-Time density)

Higher Traffic (λ_1=0.001 , λ_2=0.01, λ_3=0.1)

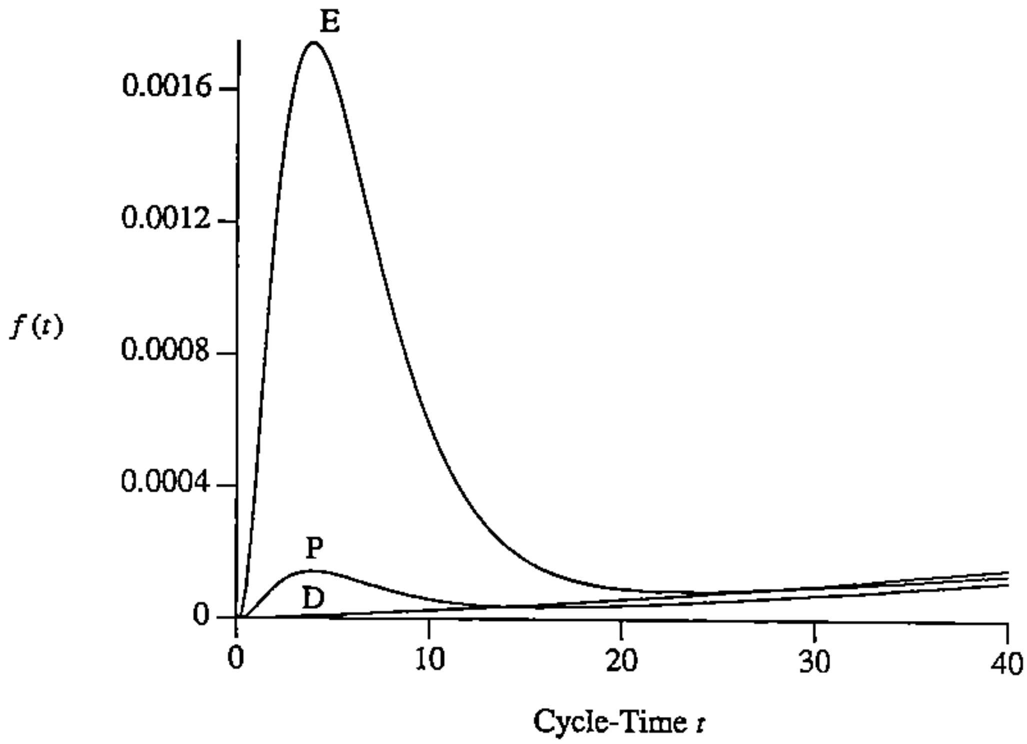

Figure 4.1c (Cycle-Time density)

Very High Traffic (λ_1=0.01 , λ_2=0.02, λ_3=0.03)

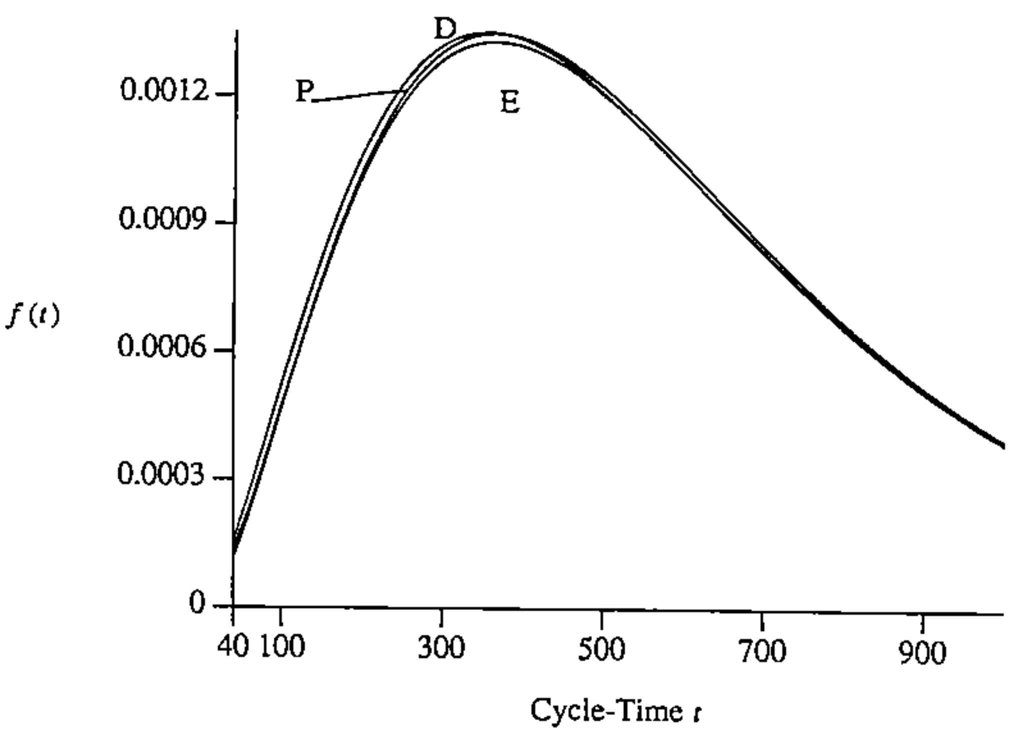

Figure 4.2c (Cycle-Time density)

Very High Traffic (λ_1=0.01 , λ_2=0.02, λ_3=0.03)

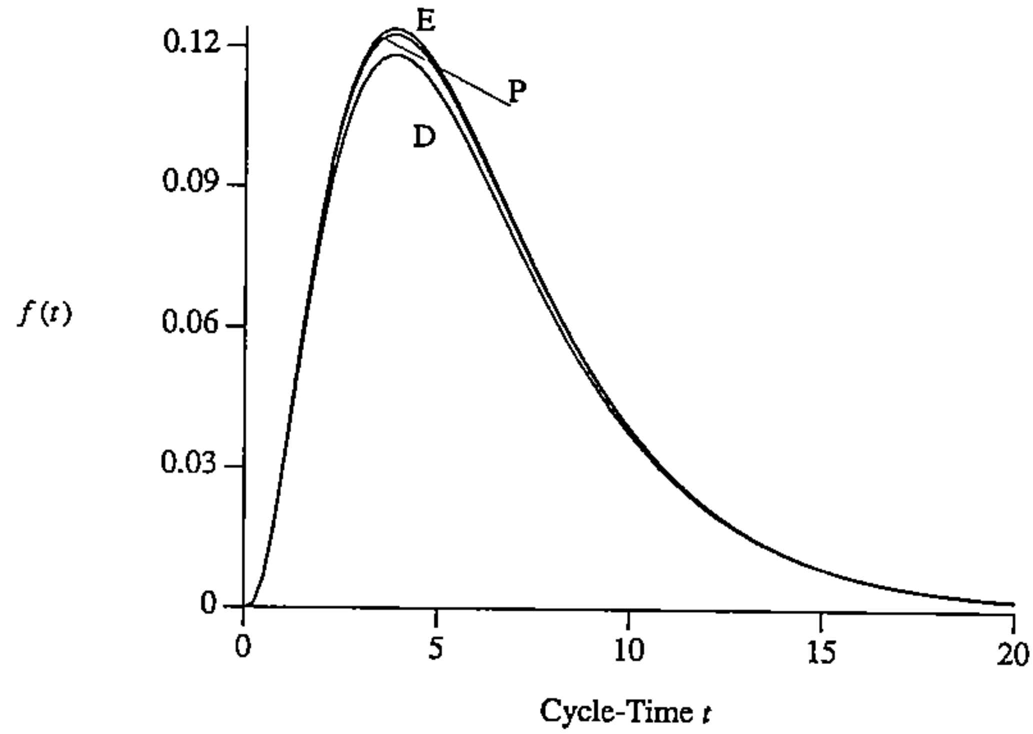

Figure 4.1d (Cycle-Time density)

Very Low Traffic ($\lambda_1=0.001$, $\lambda_2=0.002$, $\lambda_3=0.003$)

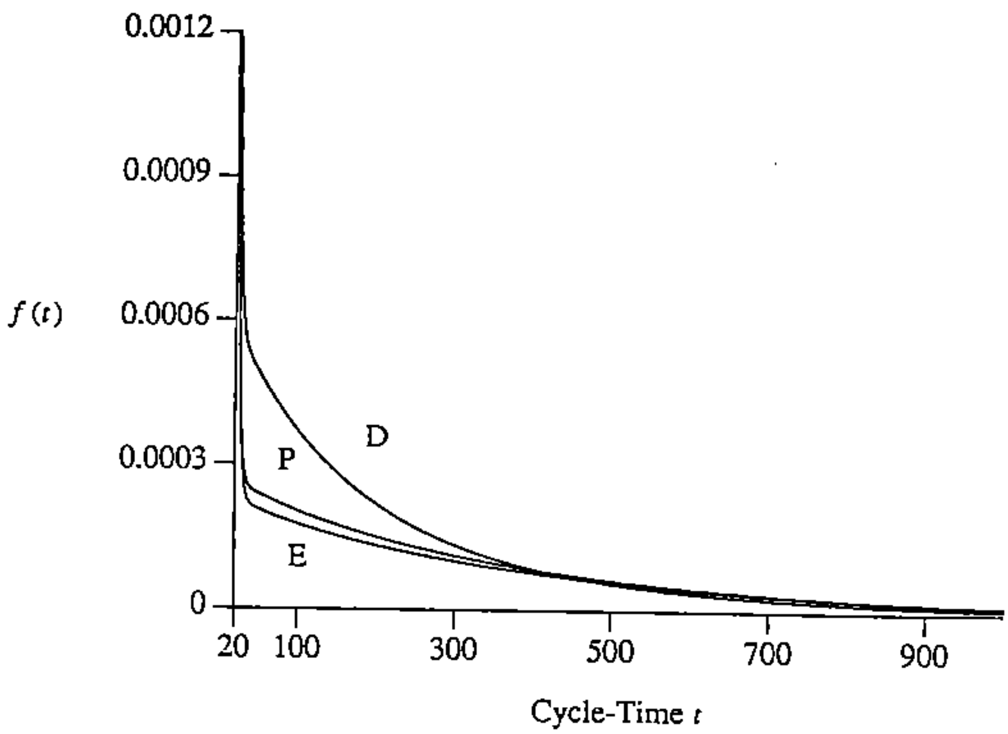

Figure 4.2d (Cycle-Time density)

Very Low Traffic ($\lambda_1=0.001$, $\lambda_2=0.002$, $\lambda_3=0.003$)

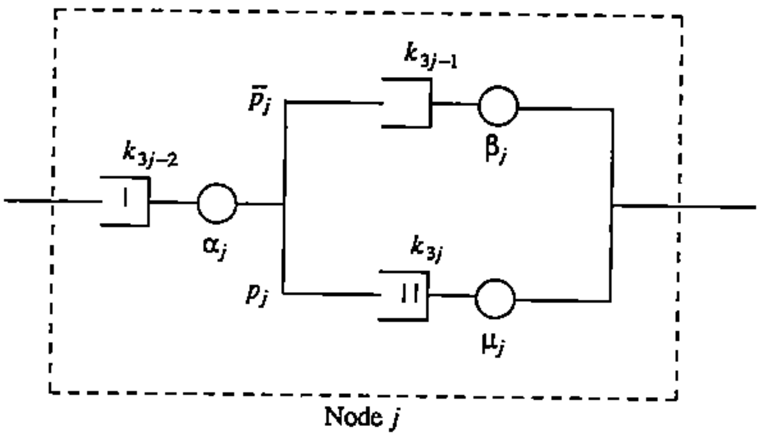

Figure 5. Description of node j in the closed-network dual.