A TUTORIAL ON THE ANALYSIS OF POLLING SYSTEMS

Hideaki Takagi
Leonard Kleinrock

February 1985
Report No. CSD-850005

A TUTORIAL ON THE ANALYSIS OF POLLING SYSTEMS[†]

by

Hideaki Takagi and Leonard Kleinrock

3732 Boelter Hall

Computer Science Department

University of California, Los Angeles

Los Angeles, California 90024

June 1, 1984

[†]This research was supported by the Defense Advanced Research Projects Agency of the Department of Defense under Contract MDA 903-82-C-0064. Hideaki Takagi is now with IBM Japan Science Institute, No. 36 Kowa Building, 5-19 Sanban-cho, Chiyoda-ku, Tokyo 102, Japan.

Abstract: A polling system is one which contains a number of queues served in cyclic order. This is a tutorial in which polling systems are analyzed to evaluate basic performance measures such as the average queue length and average waiting time. Following a taxonomy of models with reference to previous work, we consider (i) one-message buffer sustems, and (ii) infinite buffer systems with exhaustive, gated and limited service disciplines. Some examples to which the analysis of polling systems is applied are drawn from the field of computer communication networks.

Keywords:

polling systems, multiqueue, cyclic service, exhaustive service, gated service, limited service, gambler's ruin problem, renewal process, regenerative process, MSAP, Expressnet, token ring, token bus, local area networks

Table of Contents

List of Figures

1 Introduction

Polling is a way of multiplexing the service requests by several users in cyclic order with generally nonzero switch-over times. The polling scheme has been employed in computer-terminal communication systems; it is implemented in standard data link protocols such as BSC, SDLC and HDLC. From a viewpoint of queueing theory, it is a multiple queue, cyclic service system whose congestion analysis has been the subject of many papers. Its analysis is now finding a new application in the local-area computer networks (e.g., token ring). This monograph presents a reconstruction of the analysis for some types of polling system which has appeared by now in scattered publication. Our intent is to derive the basic performance measures (such as the average message delay) straightforward by skipping unnecessary (for our purpose) complications in the past literature. We also compare and contrast the models we consider from a unified approach.

To unify the terminology, let us call a user requesting service a station, and the time needed to switch service from one user to another a reply interval. The entity for service may be a variable-length message or a fixed-length packet. (A message can be viewed as consisting of several packets.) When all the statistical specifications on the message length, message arrival process and reply interval are the same for all stations, we call the system as the case of (statistically) identical stations; otherwise we have the case of nonidentical stations. The time may be continuous or discrete; in the latter case, we choose the packet service time as the (fixed-length) slot size, and let it be the unit of time. Throughout the monograph, we denote by N the number of stations. See Figure 1 for a schematic view of our model.

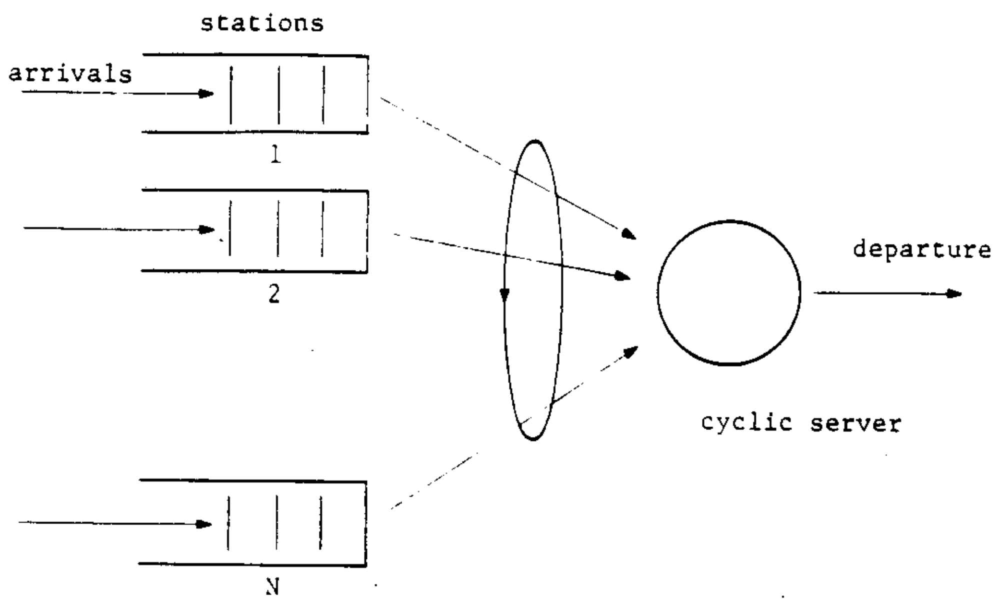

Figure 1. Polling system.

We now delineate a group of polling system models with reference to associated past work. (We do not aim at a complete list of references.) Our first classification is with respect to the buffer capacity at each station. We have (i) one-message buffer system, (ii) infinite buffer system, and (iii) generally finite buffer system (which has not been studied very much).

In one-message buffer system, at most one outstanding message can be stored at each station. We may think of this system as either such that messages arriving at a station to find the buffer occupied are lost, or such that new messages are generated at each station only after the service completion of the previous message. Analysis of the one-message buffer system can be found in [Bhar60,Sec.9.4D], [Hash81a], [Kaye72], [Mack57a], [Mack57b] and [Scho78].

In the infinite buffer system, any number of messages or packets can be stored without loss. Here, three types of service discipline have been considered: (a) exhaustive service, (b) gated service, and (c) limited service. In the exhaustive service discipline, the server serves each station until its buffer is emptied. Messages or packets arriving at a station currently in service are also served in the same service period. Such a discipline with zero reply intervals has been called an alternating priority discipline in the queueing theory. Early analysis for this case (zero reply interval) involving two stations (N=2) is in [Avi65] and [Takac68] (see also [Conw67,Sec. 9-2] and [Stid72]). The case of an arbitrary number of stations with zero reply interval is analyzed in [Coop69], and [Coop70] (see also [Coop81,Sec. 5.13]). These all assume a continuous-time model with Poisson arrival of messages. The exhaustive service with nonzero reply interval for the case N=2 is studied in [Syke70] and [Eise71]. (Their assumptions on the behavior of the server when both stations are empty are different; in [Syke70] the server keeps switching, while in [Eise71] it remains stationary.) The analysis of exhaustive service systems with an arbitrary N and nonzero reply interval is available in [Eise72], [Hash72] and [Humb78,Sec.7D] (continuous-time, nonidentical stations), in [Konh74] (discrete-time, identical stations), and in [Swar80], [DeMo81] and [Rubi83] (discrete-time, nonidentical stations). Note that the case N=1 is a queueing system with the server going on vacation [Coop81,Prob.5.12].

In the gated service discipline, the server serves a station for only those messages or packets which are in the station when it is polled. Those messages or packets which arrive during the service time are set aside to be served at the next polling. The gated service system of N identical stations with zero reply intervals is considered in [Coop69], [Coop70] and [Coop81, Sec.5.13]. An early approximate treatment for the case of nonzero reply interval was in [Leib61] and [Leib62]. Exact analyses of gated service model with an arbitary number of nonidentical stations and nonzero reply interval are given in [Hash70], [Hash72] and [Humb78,Sec.7C] (continuous-time), and [DeMo81] and [Rubi83] (discrete-time).

In the limited service system, a station is served until either 1) the buffer is emptied, or 2) the first specified number of messages or packets are served, whichever occurs first. Analysis of this system for the case where N=2 and at most one message is served from each station at a time with zero reply interval is in [Eise79] (some steps in the solution remain to be proven), where the term alternating service discipline is used; see also [Cohe83].

A similar system with nonzero reply interval is considered in [Iisa81a] and [Boxm84]. The mean message waiting time for an identical N station case with nonzero reply interval where at most one message is served from each station at a time is available in [Nomu78]. Approximate analysis for nonidentical N station case is in [Kueh79] and [Kuro81]. Another approximate treatment for a system of N stations with zero reply interval where at most k_i packets are served from each station i is in [Konh76] where the term chaining is introduced. Other papers [Iisa80] and [Iisa81b] also deal with a system of N nonidentical stations with nonzero reply interval where at most k messages are served.

For a survey of a broader class of multiplexing systems including polling, see [Chu72], [Koba77] and [Konh80]. Polling systems are also referred to in such survey articles on data communication systems as [Haye81] and [Reis82].

The following chapters of this monograph are organized as follows. Chapter 2 first presents analysis for a continuous-time, one-message buffer system with general reply intervals (my own contribution), followed by a separate analysis for the case of constant parameters (rehashment of work in [Bhar60],[Kaye 72],[Mack57a],[Scho78], and [Hash81a]). Chapters 3 through 5 provide a unified presentation of the analysis for infinite buffer systems with exhaustive and gated service. In Chapter 3 we show a detailed procedure of derivation for an exhaustive-service, discrete-time system (reconstruction of ideas in [Hash72], [Konh74], [Swar80], [DeMo81], and [Rubi83]). Chapter 4 deals with the continuous-time version of the exhaustive service system, but additionally introduces a recent development in [Ferg84]. In Chapter 5, we consider the gated service systems in both discrete and continuous time framework.

Chapter 6 considers limited service systems following the approach in [Nomu78], [Iisa80] and [Iisa81b] along with my contribution [Takag84]. In Chapter 7, we mostly reproduce work in [Coop69], [Coop70] and [Coop81] for the systems with zero reply intervals. Capitalizing on the results in preceding chapters, we have made some simplification. Chapter 8 provides several examples to which the analysis of polling systems has been applied. It includes (i) roll-call and hub polling schemes in data link networks [Schw77, Chap.12], (ii) polling in packet radio system [Toba76], (iii) MSAP (Mini-Slotted Alternating Priorities) channel access protocol [Scho80], (iv) UBS-RR (Uni-directional Broadcast System-Round Robin) and Expressnet [Toba83a], and (v) token passing in ring and bus networks [Bux81] and [Rubi83]. We conclude in Chapter 9 by summarizing the state-of-the-art and suggesting possible future research topics. Appendices A and B show the derivation and solution of a set of linear equations for the exhaustive-service, discrete-time system considered in Chapter 3. Similar presentations for other systems are found in [Hash72], [DeMo81] and [Rubi83]. Appendix C derives some equations in Chapter 4.

2 One-Message Buffer System

We consider a continuous-time polling system where each station can have at most one outstanding message; i.e., those messages which arrive at a station to find the buffer full are lost. Further, we assume an independent Poisson arrival at each station. Thus, our system can also be thought of such that each station generates a new message with an exponentially distributed time only after a previous message is completed service. Yet another view is provided by comparing our system to a machine repair system such that a repairman walks from machine to machine in an ordered fashion fixing broken machines. Here, the message arrival and its service correspond to the machine stoppage and its repair, respectively. Note that our system would be an M/G/1//N queueing system (a finite input source model, or a machine interference model; see, e.g., [Saat61, Sec.14-6]) if services were given in an FCFS order with zero reply intervals.

This chapter provides an analysis for a one-message buffer polling system consisting of statistically identical stations where the reply intervals and message service times are generally distributed. Our major performance measure is the mean message response time (the average time that an arbitrary message takes from its arrival to service completion). In the sequel, we define our system parameters and resulting performance measures. We point out some incorrectness involved in the previous work. Then we present our new analysis for the steady-state probability at the polling instant. It is shown that the mean message response time is expressed in terms of the probability that a message is found at the polling.

Let us define our system specifically. We consider a system of N stations (indexed as i = 1, 2, ... , N) served by a single server. Each station, with a single-message buffer, has an independent Poisson arrival stream of messages with a rate λ messages/second (identical for all stations). As said before, those messages which arrive to find the buffer full are lost. A message is removed from the buffer when its service is completed. Let B(x) be the distribution function for the service time of each message (again identical for all stations). Also, let $B^*(s)$ be the LST (Laplace-Stieltjes transform) of B(x), and b be the mean message service time. The server inspects the stations in the indexed order (station 1 is inspected after station N) and administers service to queued messages if any. Let R(x) be the distribution function for the reply interval from one station to another (once again, identical for all stations). Also, let $R^*(s)$ be the LST of R(x), and r be the mean reply interval. These are the parameters given to our system. Note that all stations are statistically identical; i.e., we have a symmetric system.

2.1 Performance Measures

We define a polling cycle as the time interval beginning at the polling instant of station 1 and ending with the completion of the reply interval from station N to station 1. However, if we consider the buffer occupation states at all stations as the system state, the start of each polling cycle (i.e., the instant when station 1 is polled) is not a regeneration point of the system state. Time points when station 1 is polled and all buffers are empty are regeneration points. We call the time interval between two such successive regeneration points a <u>regenerative cycle</u>. See, e.g., [Heym82, Chap.6] for the discussion of regenerative processes.

Let M be the number of polling cycles in a regenerative cycle, and C_m, m = 1, ..., M, be the duration of the m th polling cycle in a regenerative cycle. The mean (polling) cycle time E[C] is defined by

$$E[C] = \frac{E[\sum_{m=1}^{M} C_m]}{E[M]} \tag{2.1a}$$

Similarly, the mean number of messages served in a polling cycle, E[Q], is defined by

$$E[Q] = \frac{E[\sum_{m=1}^{M} Q_m]}{E[M]} \tag{2.1b}$$

where Q_m is the number of messages served in the m th polling cycle in a regenerative cycle. E[Q] can be related to [illegible], the probability that a message is found at the polling as follows:

$$\doteq \frac{E[\text{number of messages served in a regenerative cycle}]}{E[\text{number of stations polled in a regenerative cycle}]}$$

$$= \frac{E[\sum_{m=1}^{M} Q_m]}{N\ E[M]} = \frac{E[Q]}{N} \tag{2.2}$$

We proceed to derive the relationship between E[C] and E[Q]. Note that C_m consists of the sum of the reply intervals in the m th polling cycle (R_m) and the sum of message service times in the same cycle (B_m). Thus we have

$$E[\sum_{m=1}^{M} C_m] = E[\sum_{m=1}^{M} (R_m + B_m)] = E[\sum_{m=1}^{M} R_m] + E[\sum_{m=1}^{M} B_m] \tag{2.3}$$

However, since M is clearly a <u>stopping time</u> for $R_1, R_2, \ldots,$ with $E[M] < \infty$, it follows from <u>Wald's lemma</u> (see [Heym82, Chap.6]) that

$$E[\sum_{m=1}^{M} R_m] = E[M]\ E[R_1] = Nr\ E[M] \tag{2.4a}$$

By definition, we must have

$$b = \frac{E[\sum_{m=1}^{M} B_m]}{E[\sum_{m=1}^{M} Q_m]} = \frac{E[\sum_{m=1}^{M} B_m]}{E[Q]\ E[M]} \tag{2.4b}$$

Using (2.4a) and (2.4b) in (2.3), we have

$$E[\sum_{m=1}^{M} C_m] = Nr\ E[M] + bE[Q]\ E[M] \tag{2.5a}$$

Thus, from (2.1a) we get

$$E[C] = Nr + bE[Q] = N(r + b\alpha) \tag{2.5b}$$

A performance measure for our system is the mean message response time for the messages which are not lost upon arrival, which we denote by

E[T]. We now express E[T] in terms of E[Q]. Note that the buffer state of each station alternates between the 'empty' state of mean duration $1/\lambda$ and the 'full' state of mean duration E[T]. Thus, the rate of the messages served in the whole system, i.e., the throughput of the system (denoted by γ), is given by

$$\gamma = \frac{N}{E[T] + 1/\lambda} \qquad (2.6a)$$

On the other hand, the throughput must be equal to the mean number of messages served per unit time:

$$\gamma = \frac{E[\sum_{m=1}^{M} Q_m]}{E[\sum_{m=1}^{M} C_m]} = \frac{E[Q]}{E[C]} \qquad (2.6b)$$

Thus we have

$$E[T] = Nb - \frac{1}{\lambda} + \frac{N^2 r}{E[Q]} \qquad (2.7)$$

Another interesting performance measure may be the probability of message loss, P_L, when we assume that the messages arriving at nonempty stations are lost. Due to the Poisson arrival, this loss probability is equal to the long-run average fraction of time that the buffer of a station is full:

$$P_L = \frac{E[T]}{E[T] + 1/\lambda} \qquad (2.8)$$

Using P_L, we have

$$\gamma = \lambda N(1 - P_L) \quad (2.6c)$$

which implies that the throughput γ is the fraction $1 - P_L$ of the total message arrival rate λN.

From Little's result applied to those messages which are not lost upon arrival, the mean number of messages in system at an arbitrary time, $E[\tilde{N}]$, is given by

$$E[\tilde{N}] = \gamma E[T] = NP_L \quad (2.9a)$$

or, more explicitly,

$$E[\tilde{N}] = N - \frac{N(1/\lambda)}{E[T]+1/\lambda} = N - \frac{\alpha}{\lambda(r+b\alpha)} \quad (2.9b)$$

where the second terms on the central and right-hand sides of this equation represent the mean number of empty stations. The result $P_L = E[\tilde{N}]/N$ is expected from the Poisson arrival property at each station.

Note that, in the limit of zero reply intervals (i.e., $R^*(s) = 1$), $E[T]$ should be identical to the mean response time for the corresponding FCFS M/G/1//N queue (machine repairman model) given in [Saat61, Sec.14-6] (cited as L. Takács' solution):

$$E[T]_{M/G/1//N} = Nb - \frac{1}{\lambda} + \frac{1}{\lambda\left[1 + \sum_{n=1}^{N-1}\binom{N-1}{n}\prod_{j=1}^{n}\{[B^*(j\lambda)]^{-1} - 1\}\right]} \quad (2.10)$$

This claim comes from the following reasoning. First, for both cyclic and FCFS service systems, we have the same arrival process (namely, the pseudo-Poisson arrival such that the state-dependent arrival rate is $(N-i)\lambda$ when i stations are nonempty). Although the service disciplines with which the server selects a station to serve are different, the statistical characteristic of the message service time does not depend on the selected station. Thus, the behavior of unfinished work in both systems should be statistically identical. It follows from Little's principle that the mean message response times are the same for both systems.

2.2 Survey of Previous Work

We now give a short survey of related previous work. Mack, Murphy and Webb [Mack57a] considered a system (in the context of machine repairman model) of constant reply interval (r_i for station i) and constant message service time (b). They obtained an expression for the mean number of messages served in a cycle of polling as (when $r_i=r$ for all i)

$$E[Q] = \frac{N \sum_{n=0}^{N-1} \binom{N-1}{n} \prod_{j=0}^{n} [e^{\lambda(Nr+jb)} - 1]}{1 + \sum_{n=1}^{N} \binom{N}{n} \prod_{j=0}^{n-1} [e^{\lambda(Nr+jb)} - 1]} \tag{2.11}$$

They also showed the relation in (2.5b). Later, Scholl and Potier [Scho78] used (2.7) to obtain the mean message response time. Note that, in the limit of zero reply intervals (i.e., $r \to 0$), (2.7) with (2.11) reduces to (2.10) with $B^*(s) = e^{-sb}$. For the same model, Kaye [Kaye72] found the distribution for the message waiting time. (His expression for the mean waiting time can be shown to be equal to (2.7) with (2.11) minus b.) See Section 2.5 for details.

Mack [Mack57b] considered a case where the reply interval is constant and the message service times are variable. He derived a system of 2^{N-1} linear equations to find E[C]. However, he did not derive the mean response time. Bharucha-Reid [Bhar60,Sec.9.4D] also dealt with the case of constant reply interval (r) and generally distributed message service time, and showed

$$E[Q] = \frac{N\sum_{n=0}^{N-1}\binom{N-1}{n}\prod_{j=0}^{n} e^{\lambda N r}[B^*(\lambda)]^j - 1}{1+\sum_{n=1}^{N}\binom{N}{n}\prod_{j=0}^{n-1} e^{\lambda N r}[B^*(\lambda)]^j - 1} \qquad (2.12)$$

(This reduces to (2.11) when $B^*(s) = e^{-sb}$.) Hashida and Kawashima [Hash81a] derived a similar expression for the mean number of messages served in an intervisit time (defined as the time interval beginning with the start of the reply interval from station 1 to station 2 and ending with completion of the reply interval from station N to station 1).

We should note, however, that if we calculate the mean message response time with (2.7) and (2.12) and take the limit of zero reply interval (i.e., $r \to 0$), we do not get the limiting from in (2.10). Thus, the analyses in [Bhar60, Sec.9.4D] and [Hash81a] seem to be wrong. The incorrectness seems to come from their using $B^*(\lambda)$ as the probability of no arrivals during each message service time in evaluating the probability of buffer occupancy which the last N inspections have encountered. However, for example, if we find a message at station N - 1 and no message at station N, the service time at station N - 1 must have been "smaller" than what $B^*(s)$ would imply. Thus we cannot use $B^*(\lambda)$ uniformly. This error is pointed out more formally in [Coff84].

Our conclusion in this brief survey of previous work is that the mean message response time in the case of general $R^*(s)$ and $B^*(s)$ is not available. In the next section, we present a new analysis (taking into account the conditional distribution of the message length and reply interval) to give α, from which all performance measures can be calculated as shown in Section 2.1.

2.3 State Probability at Polling Instants

We begin our analysis by defining the state of station i, denoted by v_i, as

$$v_i = \begin{cases} 0 & \text{buffer of station } i \text{ is full} \\ 1 & \text{buffer of station } i \text{ is empty} \end{cases} \qquad i = 1, 2, \ldots, N \tag{2.13}$$

Obviously the system state $(v_1, v_2, \ldots, v_N)$ at each polling instant forms a Markov chain. Let $p_i(v_1, v_2, \ldots, v_N)$ be the probability that the state of station j is v_j $(j = 1, 2, \ldots, N)$ at the instant when station i is polled. We are going to express $p_{i+1}(v_1, v_2, \ldots, v_N)$ in terms of $p_i(v_1, v_2, \ldots, v_N)$ and the possible events happening between the polling instants for stations i and i + 1. Since all the stations are statistically identical, let us focus on station i = 1 below.

Since the duration of the interval between the polling instants for station 1 and 2 is equal to a reply interval when $v_1 = 1$ (no message found), and equal to a reply interval plus a message service time when $v_1 = 0$ (a message found), we have the following steady-state probability transition equation:

$$\sum_{v_1} \sum_{v_2} \cdots \sum_{v_N} p_2(v_1, v_2, \ldots, v_N) \prod_{j=1}^{N} (z_j)^{v_j}$$

$$= \sum_{v_2} \sum_{v_3} \cdots \sum_{v_N} p_1(1, v_2, \ldots, v_N)$$

$$\times \int (1-e^{-\lambda x}+z_1 e^{-\lambda x}) \prod_{j=2}^{N} (1-e^{-\lambda x}+z_j e^{-\lambda x})^{v_j} \, dR(x)$$

$$+ \sum_{v_2} \sum_{v_3} \cdots \sum_{v_N} p_1(0, v_2, \ldots, v_N)$$

$$\times \iint (1-e^{-\lambda x}+z_1 e^{-\lambda x}) \prod_{j=2}^{N} [1-e^{-\lambda(x+y)}+z_j e^{-\lambda(x+y)}]^{v_j} \, dR(x)dB(y)$$

(2.14)

Hereafter, the summation with respect to v_j covers $v_j = 0$ and 1 for all j. The range of integrals is always from 0 to ∞. We define the joint generating function for $p_1(v_1, v_2, \ldots, v_N)$ by

$$F(v_1; z_2, z_3, \ldots, z_N) \triangleq \sum_{v_2} \sum_{v_3} \cdots \sum_{v_N} p_1(v_1, v_2, \ldots, v_N) \prod_{j=2}^{N} (z_j)^{v_j}$$

$$v_1 = 0, 1 \qquad (2.15)$$

Then the right-hand side of (2.14) can be written as

$$\int F(1; 1-e^{-\lambda x} + z_2 e^{-\lambda x}, \ldots) \, dR(x)$$
$$+ (z_1 - 1) \int e^{-\lambda x} F(1; 1-e^{-\lambda x} + z_2 e^{-\lambda x}, \ldots) \, dR(x)$$
$$+ \iint F(0; 1-e^{-\lambda(x+y)} + z_2 e^{-\lambda(x+y)}, \ldots) \, dR(x)dB(y)$$
$$+ (z_1 - 1) \iint e^{-\lambda x} F(0; 1-e^{-\lambda(x+y)} + z_2 e^{-\lambda(x+y)}, \ldots) \, dR(x)dB(y)$$

(2.16a)

In order to express the left-hand side of (2.14) in terms of F, we note from the symmetry among stations that

$$p_2(v_1, v_2, \ldots, v_N) = p_1(v_2, \ldots, v_N, v_1)$$

for all values of $(v_1, v_2, \ldots, v_N)$. Thus the left-hand side of (2.14) may be expressed as

$$F(0; z_3, \ldots, z_N, z_1) + z_2 F(1; z_3, \ldots, z_N, z_1) \tag{2.16b}$$

We further introduce $f(v_1, v_2, \ldots, v_N)$ by

$$F(v_1; z_2, \ldots, z_N) = \sum_{v_2} \cdots \sum_{v_N} f(v_1, v_2, \ldots, v_N) \prod_{j=2}^{N} (z_j - 1)^{v_j} \tag{2.17}$$

Substituting this expression into (2.16a) and (2.16b), and equating the coefficients of $\Pi(z_j - 1)^{v_j}$, we obtain a system of linear equations for $f(v_1, v_2, \ldots, v_N)$:

$$\begin{aligned} & f(1, 0, v_3, \ldots, v_N)\, R^*[\lambda(\sum_{j=3}^{N} v_j)] \\ & + f(0, 0, v_3, \ldots, v_N)\, R^*[\lambda(\sum_{j=3}^{N} v_j)]\, B^*[\lambda(\sum_{j=3}^{N} v_j)] \\ & = f(0, v_3, \ldots, v_N, 0) + f(1, v_3, \ldots, v_N, 0), \end{aligned} \tag{2.18a}$$

$$\begin{aligned} & f(1, 0, v_3, \ldots, v_N)\, R^*[\lambda(1 + \sum_{j=3}^{N} v_j)] \\ & + f(0, 0, v_3, \ldots, v_N) R^*[\lambda(1 + \sum_{j=3}^{N} v_j)]\, B^*[\lambda(\sum_{j=3}^{N} v_j)] \\ & = f(0, v_3, \ldots, v_N, 1) + f(1, v_3, \ldots, v_N, 1), \end{aligned} \tag{2.18b}$$

$$\begin{aligned} & f(1, 1, v_3, \ldots, v_N)\, R^*[\lambda(1 + \sum_{j=3}^{N} v_j)] \\ & + f(0, 1, v_3, \ldots, v_N)\, R^*[\lambda(1 + \sum_{j=3}^{N} v_j)]\, B^*[\lambda(1 + \sum_{j=3}^{N} v_j)] \\ & = f(1, v_3, \ldots, v_N, 0), \end{aligned} \tag{2.18c}$$

$$f(1, 1, v_3, \ldots, v_N)\, R^*[\lambda(2 + \sum_{j=3}^{N} v_j)]$$

$$+ f(0, 1, v_3, \ldots, v_N)\, R^*[\lambda(2 + \sum_{j=3}^{N} v_j)]\, B^*[\lambda(1 + \sum_{j=3}^{N} v_j)]$$

$$= f(1, v_3, \ldots, v_N, 1)$$

$$v_3, v_4, \ldots, v_N = 0, 1 \qquad (2.18d)$$

One of the equations in (2.18a-d) is redundant (namely, the case of $v_3 = \ldots = v_N = 0$ in (2.18a)). From the normalization condition

$$F(0; 1, \ldots, 1) + F(1; 1, \ldots, 1) = 1, \qquad (2.19a)$$

we have

$$f(0, 0, \ldots, 0) + f(1, 0, \ldots, 0) = 1 \qquad (2.19b)$$

The probability that a message is found at polling is given by

$$\alpha = f(0, 0, \ldots, 0) = F(0; 1, \ldots, 1) = \sum_{v_2} \cdots \sum_{v_N} p_1(0, v_2, \ldots, v_N) \qquad (2.20)$$

We have so far been unable to find an analytical solution to the above system of equations for a general value of N. So the solution can be obtained only numerically. For the cases of $N = 2$ and 3, however, we have analytical solutions which give

$$f(0,0) = \frac{r_1 r_2 b_1}{1-r_2+r_2 b_1} \qquad (2.21a)$$

$$f(0,0,0) = \frac{1-r_3+b_2 r_3-[r_2(1-b_1)+r_1^2 b_1]\{b_2 r_2^2+b_1[r_1(1-r_3)+b_2(r_1 r_3-r_2^2)]\}}{1-r_3+b_2 r_3-r_2(1-b_1)\{b_2 r_2^2+b_1[r_1(1-r_3)+b_2(r_1 r_3-r_2^2)]\}} \qquad (2.21b)$$

where

$$r_j = R^*(j\lambda) \quad \text{and} \quad b_j = B^*(j\lambda) \quad \text{for } j = 1, 2, 3 \tag{2.22}$$

If we let $R^*(s) = e^{-sr}$ (constant reply interval) and take the limit $r \to 0$ in (2.21a,b), and use them in (2.7) with (2.2), then we recover (2.10).

In Figure 2, we plot E[T] in some cases where we see that our mean response time approaches to that for the corresponding FCFS M/G/1//N model given by (2.10).

2.4 Case of Constant Parameters

In this section, we assume that all messages are of the same length b and the reply interval between station i and station i+1 is also a constant r_i (i=1, 2, ..., N). Let

$$\overline{R} \triangleq \sum_{i=1}^{N} r_i \tag{2.23}$$

Let us first derive the probability distribution for Q, the number of messages served in a polling cycle. Contrary to (2.13), we define the state of each station u_i as 1 and 0 depending on whether it has a message or not. Let $\tau_i(m)$ be the polling instant for station i in the m th polling cycle, and $\overline{\tau}_i(m)$ be the instant at which the reply interval between station i and station i+1 starts. We denote by $u_i^{(m)}$ the state of station i at time $t = \tau_i(m)$ (i=1, 2, ..., N). Using the convention

$$U_i^{(m)} = (u_1^{(m)}, u_2^{(m)}, \ldots, u_i^{(m)}, u_{i+1}^{(m-1)}, u_{i+2}^{(m-1)}, \ldots, u_N^{(m-1)}),$$

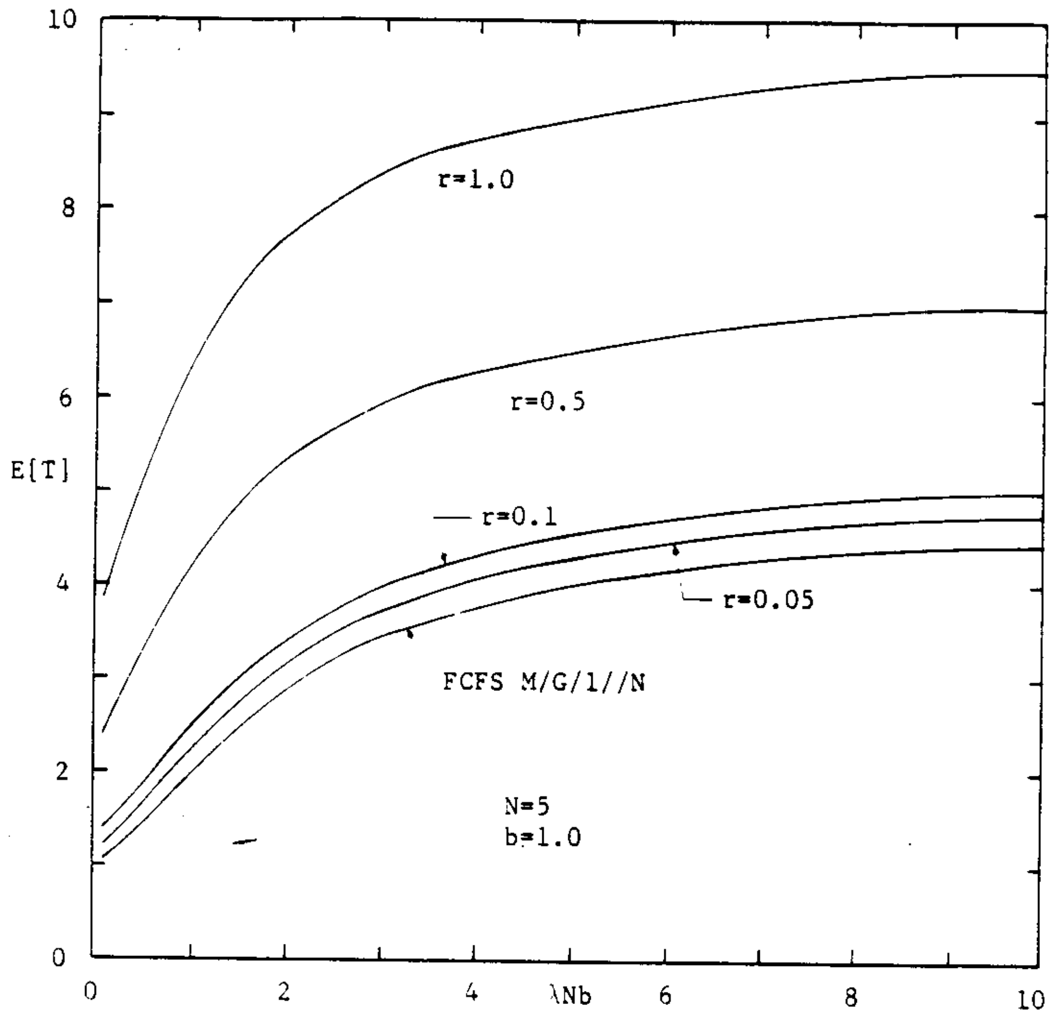

Figure 2. Mean message response time (constant reply interval and exponentially distributed message service time).

we define $P_i^{(m)}(U_i^{(m)})$ as the probability that the server observes the sequence $U_i^{(m)}$ for the states of N stations at their polled instants prior to $t = \tau_i(m)$. Note that $U_{i+1}^{(m)}$ only depends on $U_i^{(m)}$ and what happens between $t = \overline{\tau}_i(m)$ and $t = \tau_{i+1}(m)$. Therefore, a process $\{U_i^{(m)}\}$ is a Markov chain with 2^N distinct states in $[0,1]^N$.

The state transition probabilities of this Markov chain are now considered. Note that $u_i^{(m)}$ is 0 (no message found) if and only if there are no arrivals at station i after $t = \overline{\tau}_i(m-1)$. The probability of this event is given by

$$\exp[-\lambda\bar{R}-\lambda b(u_{i+1}^{(m-1)}+u_{i+2}^{(m-1)}+ \ldots +u_N^{(m-1)}+u_1^{(m)}+u_2^{(m)}+ \ldots +u_{i-1}^{(m)})] \quad (2.24)$$

Since $u_i^{(m-1)}$ does not affect $u_i^{(m)}$, we have the relation

$$P_i^{(m)}(u_1^{(m)}, \ldots, u_{i-1}^{(m)}, 1, u_{i+1}^{(m-1)}, \ldots, u_N^{(m-1)})$$
$$= \{1-\exp[-\lambda\bar{R}-\lambda b(u_{i+1}^{(m-1)}+ \ldots +u_N^{(m-1)}+u_1^{(m)}+ \ldots +u_{i-1}^{(m)})]\}$$
$$\cdot \sum_{u_i^{(m-1)}=0}^{1} P_{i-1}^{(m)}(u_1^{(m)},\ldots, u_{i-1}^{(m)}, u_i^{(m-1)}, u_{i+1}^{(m-1)},\ldots, u_N^{(m-1)}) \quad (2.25a)$$

$$P_i^{(m)}(u_1^{(m)}, \ldots, u_{i-1}^{(m)}, 0, u_{i+1}^{(m-1)}, \ldots, u_N^{(m-1)})$$
$$= \exp[-\lambda\bar{R}-\lambda b(u_{i+1}^{(m-1)}+ \ldots +u_N^{(m-1)}+u_1^{(m)}+ \ldots +u_{i-1}^{(m)})]$$
$$\cdot \sum_{u_i^{(m-1)}=0}^{1} P_{i-1}^{(m)}(u_1^{(m)},\ldots, u_{i-1}^{(m)}, u_i^{(m-1)}, u_{i+1}^{(m-1)},\ldots, u_N^{(m-1)}) \quad (2.25b)$$

Consider now the limiting probability

$$P_i(u_1, \ldots, u_N) \triangleq \lim_{m \to \infty} P_i^{(m)}(U_i^{(m)}) \tag{2.26}$$

If we take the limit $m \to \infty$, (2.25a) and (2.25b) become

$$P_i(u_1, \ldots, u_{i-1}, 1, u_{i+1}, \ldots, u_N) = \{1-\exp[-\lambda\overline{R}-\lambda b(\sum_{\substack{j=1 \\ (j\neq i)}}^{N} u_j)]\} \cdot \sum_{k=0}^{1} P_{i-1}(u_1, \ldots, u_{i-1}, k, u_{i+1}, \ldots, u_N) \tag{2.27a}$$

$$P_i(u_1, \ldots, u_{i-1}, 0, u_{i+1}, \ldots, u_N) = \exp[-\lambda\overline{R}-\lambda b(\sum_{\substack{j=1 \\ (j\neq i)}}^{N} u_j)] \cdot \sum_{k=0}^{1} P_{i-1}(u_1, \ldots, u_{i-1}, k, u_{i+1}, \ldots, u_N) \tag{2.27b}$$

It can be confirmed by direct substitution that (2.27a) and (2.27b) are satisfied by

$$P_i(u_1, \ldots, u_N) = \begin{cases} K & u_1 = \ldots = u_N = 0 \\ K \prod_{j=0}^{\sum_{k=1}^{N} u_k - 1} \{\exp[\lambda(\overline{R}+jb)]-1\} & \sum_{k=1}^{N} u_k > 0 \end{cases} \tag{2.28a}$$

where K is a constant to be determined by the normalization condition

$$\sum_{(u_1, \ldots, u_N) \in [0,1]^N} P_i(u_1, \ldots, u_N) = 1 \tag{2.28b}$$

Note that $P_i(u_1, \ldots, u_N)$ in (2.28a) is independent of i due to symmetry in the system. Since $P_i(u_1, \ldots, u_N)$ is now expressed only in terms of $Q = \sum_{k=1}^{N} u_k$, the number of messages served in a cycle, we have

$$P(n) \triangleq P\{Q=n\} = \begin{cases} K & n = 0 \\ K\binom{N}{n} \prod_{j=0}^{n-1}(e^{\lambda\tau_j} - 1) & n \geq 1 \end{cases} \tag{2.29a}$$

where

$$K^{-1} = 1 + \sum_{n=1}^{N}\binom{N}{n} \prod_{j=0}^{n-1}(e^{\lambda\tau_j} - 1) \tag{2.29b}$$

Here we have introduced the notation

$$\tau_j \equiv \overline{R} + jb \qquad j=0,\ 1,\ 2,\ \ldots \tag{2.30}$$

The average member of messages served in a cycle is given by

$$E[Q]=\sum_{n=1}^{N} nP(n)=K\sum_{n=1}^{N} n\binom{N}{n}\prod_{j=0}^{n-1}(e^{\lambda\tau_j} -1) \tag{2.31a}$$

By using $n\binom{N}{n} = N\binom{N-1}{n-1}$, this may be written as

$$E[Q]=KN\sum_{n=0}^{N-1}\binom{N-1}{n} \prod_{j=0}^{n}(e^{\lambda\tau_j} -1) \tag{2.31b}$$

Thus we get (2.11) if $r_i=r$ is assumed for all i. The probability that a message is found at the polling, α, is given by (2.2), i.e.,

$$\alpha = \frac{E[Q]}{N} = K\sum_{n=0}^{N-1}\binom{N-1}{n}\prod_{j=0}^{n}(e^{\lambda\tau_j} - 1) \tag{2.32}$$

We proceed to find the distribution of the cycle length C in the form of LST, $C^*(s)$, of its distribution function. Since

$$C = \overline{R} + \sum_{i=1}^{N} u_i b \tag{2.33}$$

it follows that

$$C^*(s)=E[e^{-sC}]=E[\exp\{-s(\overline{R}+\sum_{i=1}^{N}u_i b)\}]=\sum_{n=0}^{N}e^{-s\tau_n}P(n) \tag{2.34a}$$

Using (2.29a) for P(n) in (2.34a), we have

$$C^*(s)=K[e^{-s\overline{R}}+\sum_{n=1}^{N}\binom{N}{n}e^{-s\tau_n}\prod_{j=1}^{n-1}(e^{\lambda\tau_j}-1)] \tag{2.34b}$$

from which the mean cycle time is computed as

$$E[C]=\overline{R}+Kb\sum_{n=1}^{N}n\binom{N}{n}\prod_{j=0}^{n-1}(e^{\lambda\tau_j}-1)$$

$$=\overline{R}+NKb\sum_{n=0}^{N-1}\binom{N-1}{n}\prod_{j=0}^{n}(e^{\lambda\tau_j}-1) \tag{2.35a}$$

By (2.32), we have

$$E[C]=\overline{R}+Nb\alpha \tag{2.35b}$$

which gives (2.5b) when $r_i = r$ for all i.

2.4.1 Waiting Time

We derive the distribution of message waiting time W (the period from the time of arrival to the service start). Consider a tagged message whose service is just being started in station, say, i. Suppose that n messages (including the tagged message) are served in a cycle ending with this service. The (arbitrarily chosen) tagged message falls in such a cycle with probability

$$\pi(n)=\frac{n}{E[Q]}P(n) \tag{2.36}$$

(This comes from the reasoning that an arbitrary message falls in a longer cycle with greater probability; see, e.g., [Coop81, Sec.2.1]). Conditioned on the fact that our tagged message is the n th message served in the cycle, it must have arrived sometime during the interval of length τ_{n-1} before the service starts.

Let t be the arrival time of the message in the preceding cycle. Then the conditional distribution of W is given by

$$P\{W > w|n\} = \begin{cases} \dfrac{P\{t < \tau_{n-1} - w\}}{P\{t \le \tau_{n-1}\}} & 0 \le w \le \tau_{n-1} \\ 0 & w \ge \tau_{n-1} \end{cases} \qquad (2.37)$$

Since t is distributed exponentially with mean $1/\lambda$, we have

$$P\{W > w|n\} = \begin{cases} \dfrac{1 - e^{-\lambda(\tau_{n-1} - w)}}{1 - e^{-\lambda\tau_{n-1}}} & 0 \le w \le \tau_{n-1} \\ 0 & w \ge \tau_{n-1} \end{cases} \qquad (2.38a)$$

Taking the complement, we get

$$P\{W \le w|n\} = 1 - P\{W > w|n\} = \begin{cases} \dfrac{e^{\lambda w} - 1}{e^{\lambda\tau_{n-1}} - 1} & 0 \le w \le \tau_{n-1} \\ 1 & w \ge \tau_{n-1} \end{cases} \qquad (2.38b)$$

We now uncondition (2.38b) on n by using (2.36). Defining

$$n(w) \triangleq \lfloor 1 + (w - \overline{R})/b \rfloor \qquad (2.39)$$

we have

$$P\{W \le w\} = \sum_{n=1}^{n(w)} 1 \cdot \frac{n}{E[Q]} P(n) + \sum_{n=n(w)+1}^{N} \frac{e^{\lambda w} - 1}{e^{\lambda \tau_{n-1}} - 1} \cdot \frac{n}{E[Q]} P(n) \tag{2.40}$$

Using a recursive relation (which comes from (2.29a))

$$\frac{P(n)}{P(n-1)} = \left(\frac{N-n+1}{n}\right)(e^{\lambda \tau_{n-1}} - 1) \qquad 1 \le n \le N \tag{2.41}$$

we obtain the distribution function for W:

$$P\{W \le w\} = \frac{1}{E[Q]} \sum_{n=1}^{n(w)} nP(n) + (e^{\lambda w} - 1) \sum_{n=n(w)+1}^{N} (N-n+1)P(n-1) \tag{2.42}$$

The mean of W can be calculated from its conditional distribution in (2.38a). The conditional mean is given by

$$E[W|n] = \int_0^{\tau_{n-1}} P\{W>w|n\}dw = \frac{\tau_{n-1}}{1-e^{-\lambda \tau_{n-1}}} - \frac{1}{\lambda} \tag{2.43a}$$

Unconditioning this expression on n by using (2.36), ~~and making use of~~ (2.41), we get

$$E[W] = \frac{1}{E[Q]} \sum_{n=0}^{N-1} \tau_n e^{\lambda \tau_n} (N-n)P(n) - \frac{1}{\lambda} \tag{2.43b}$$

However, this expression is shown to be identical to

$$E[W] = (N-1)b - \frac{1}{\lambda} + \frac{N\overline{R}}{E[Q]} \tag{2.44a}$$

To prove (2.44a), let us substitute (2.29a) into (2.43b) to have

$$E[W] = \frac{1}{E[Q]}[\bar{R}e^{\lambda\bar{R}}NK + \sum_{n=1}^{N-1} \tau_n e^{\lambda\tau_n}(N-n)K\binom{N}{n}\prod_{j=0}^{n-1}(e^{\lambda\tau_j}-1)] - \frac{1}{\lambda}$$

$$= \frac{NK}{E[Q]}[\bar{R}e^{\lambda\bar{R}} + \sum_{n=1}^{N-1}(\bar{R}+nb)e^{\lambda\tau_n}\binom{N-1}{n}\prod_{j=0}^{n-1}(e^{\lambda\tau_j}-1)] - \frac{1}{\lambda}$$

$$= \frac{NK\bar{R}}{E[Q]}[e^{\lambda\bar{R}} + \sum_{n=1}^{N-1}\binom{N-1}{n}e^{\lambda\tau_n}\prod_{j=0}^{n-1}(e^{\lambda\tau_j}-1)]$$

$$+ \frac{NKb}{E[Q]}\sum_{n=1}^{N-1}n\binom{N-1}{n}e^{\lambda\tau_n}\prod_{j=0}^{n-1}(e^{\lambda\tau_j}-1) - \frac{1}{\lambda} \tag{2.44b}$$

However, we can show that

$$K^{-1} = e^{\lambda\bar{R}} + \sum_{n=1}^{N-1}\binom{N-1}{n} e^{\lambda\tau_n}\prod_{j=0}^{n-1}(e^{\lambda\tau_j}-1) \tag{2.45a}$$

and

$$K\sum_{n=1}^{N-1}n\binom{N-1}{n}e^{\lambda\tau_n}\prod_{j=0}^{n-1}(e^{\lambda\tau_j}-1) = (N-1)\alpha \tag{2.45b}$$

Substituting (2.45a) and (2.45b) into (2.44b), we get (2.44a). To prove (2.45a), we write the r.h.s. as

$$e^{\lambda\bar{R}} + \sum_{n=1}^{N-1}\binom{N-1}{n}\prod_{j=0}^{n}(e^{\lambda\tau_j}-1) + \sum_{n=1}^{N-1}\binom{N-1}{n}\prod_{j=0}^{n-1}(e^{\lambda\tau_j}-1)$$

$$= 1 + \sum_{n=1}^{N}\binom{N-1}{n-1}\prod_{j=0}^{n-1}(e^{\lambda\tau_j}-1) + \sum_{n=1}^{N-1}\binom{N-1}{n}\prod_{j=0}^{n-1}(e^{\lambda\tau_j}-1)$$

$$= 1 + \sum_{n=1}^{N}[\binom{N-1}{n-1} + \binom{N-1}{n}]\prod_{j=0}^{n-1}(e^{\lambda\tau_j}-1)$$

$$= 1 + \sum_{n=1}^{N}\binom{N}{n}\prod_{j=0}^{n-1}(e^{\lambda\tau_j}-1) = K^{-1}$$

due to (2.29b). We can shown (2.45b) as

$$\sum_{n=1}^{N-1} n\binom{N-1}{n} e^{\lambda\tau_n} \prod_{j=0}^{n-1}(e^{\lambda\tau_j} - 1)$$

$$= (N-1)\sum_{n=1}^{N-1}\binom{N-2}{n-1} e^{\lambda\tau_n} \prod_{j=0}^{n-1}(e^{\lambda\tau_j} - 1)$$

$$= (N-1)\left[\sum_{n=1}^{N-1}\binom{N-2}{n-1} \prod_{j=0}^{n}(e^{\lambda\tau_j} - 1) + \sum_{n=1}^{N-1}\binom{N-2}{n-1}\prod_{j=0}^{n-1}(e^{\lambda\tau_j} - 1)\right]$$

$$= (N-1)\sum_{n=0}^{N-1}\left[\binom{N-2}{n-1} + \binom{N-2}{n}\right] \prod_{j=0}^{n}(e^{\lambda\tau_j} - 1)$$

$$= (N-1)\sum_{n=0}^{N-1}\binom{N-1}{n} \prod_{j=0}^{n}(e^{\lambda\tau_j} - 1) = (N-1)\frac{\lambda}{K}$$

due to (2.32).

From (2.44a), we have the mean response time

$$E[T] = E[W] + b = Nb - \frac{1}{\lambda} + \frac{N\overline{R}}{E[Q]} \tag{2.46}$$

which yields (2.7) when $r_i = r$ for all i.

As a digression, let us compare our average response time given in (2.46) in the limit $\overline{R} \to 0$ to the average response time in the M/D/1//N queue (or the machine interference model). Note that in the M/D/1//N queue, the service is given to messages in the order of arrival, while in the polling system the service is administered according to the cyclic scan of stations.

We first consider the limit $\overline{R} \to 0$ in (2.46). Note that the expression for K^{-1} in (2.29b) can be written as

$$K^{-1} = 1 + (e^{\lambda\overline{R}} - 1)[N + \sum_{n=2}^{N} \binom{N}{n} \prod_{j=1}^{n-1} (e^{\lambda\tau_j} - 1)] \tag{2.47}$$

Thus, we have $K \to 1$ as $\overline{R} \to 0$. Similarly, since the expression for E[Q] in (2.31b) can be written as

$$E[Q] = NK(e^{\lambda\overline{R}} - 1)[1 + \sum_{n=1}^{N-1} \binom{N-1}{n} \prod_{j=1}^{n} (e^{\lambda\tau_j} - 1)]$$

we have

$$\lim_{\overline{R}\to 0} \frac{E[Q]}{\overline{R}} = N\lambda[1 + \sum_{n=1}^{N-1} \binom{N-1}{n} \prod_{j=1}^{n} (e^{\lambda jb} - 1)] \tag{2.48}$$

Using this limit in (2.46), we get

$$E[T]_{\text{polling}}\Big|_{\overline{R}\to 0} = Nb - \frac{1}{\lambda} + \frac{1}{\lambda[1 + \sum_{n=1}^{N-1} \binom{N-1}{n} \prod_{j=1}^{n} (e^{\lambda jb} - 1)]} \tag{2.49}$$

It is clear that $E[T]_{M/G/1//N}$ in (2.10) with $B^*(s) = e^{-sb}$ is identical to (2.49).

2.4.2 Intervisit Time

The intervisit time for station i is defined as the time interval between $t = \overline{\tau}_i(m)$ and $t = \tau_i(m+1)$. The probability that the server observes a sequence of states $(u_{i+1}, \ldots, u_N, u_1, \ldots, u_{i-1})$ during the intervisit time is given by the marginal distribution

$$\sum_{u_i=0}^{1} P_i(u_1, \ldots, u_{i-1}, u_i, u_{i+1}, \ldots, u_N)$$

$$= \begin{cases} Ke^{\lambda\overline{R}} & u_1 = \ldots = u_{i-1} = u_{i+1} = \ldots = u_N = 0 \\ K\cdot\exp[\lambda(\overline{R} + b\sum_{\substack{k=1\\(k\neq i)}}^{N} u_k)] \prod_{j=0}^{\sum_{\substack{k=1\\(k\neq i)}}^{N} u_k - 1} (e^{\lambda\tau_j} - 1) & \sum_{\substack{k=1\\(k\neq i)}}^{N} u_k > 0 \end{cases} \tag{2.50}$$

from (2.28a). Therefore, the distribution of the number of messages served in an intervisit time, $Q' = \sum_{\substack{k=1\\(k\neq i)}}^{N} u_k$, is given by

$$P'(n) = P\{Q'=n\} = \begin{cases} Ke^{\lambda\overline{R}} & n = 0 \\ K\binom{N-1}{n} e^{\lambda\tau_n} \prod_{j=0}^{n-1} (e^{\lambda\tau_j} - 1) & 1 \leq n \leq N-1 \end{cases} \tag{2.51}$$

The normalization for (2.51) is provided by (2.45a).

The LST $I^*(s)$ for the duration of an intervisit time I is obtained from (2.51). Since

$$I = \overline{R} + \sum_{\substack{k=1\\(k\neq i)}}^{N} u_k b \tag{2.52}$$

we have

$$I^*(s) = E[e^{-sI}] = E[\exp\{-s(\overline{R} + \sum_{\substack{k=1\\(k\neq i)}}^{N} u_k b)\}] = \sum_{n=0}^{N-1} e^{-s\tau_n} P'(n) \tag{2.53a}$$

Using (2.51) we get

$$I^*(s)=K[e^{(\lambda-s)\overline{R}}+\sum_{n=1}^{N-1}\binom{N-1}{n}\,e^{(\lambda-s)\tau_n}\prod_{j=0}^{n-1}(e^{\lambda\tau_j}-1)] \tag{2.53b}$$

from which the mean intervisit time is given by

$$E[I]=\overline{R}+Kb\sum_{n=1}^{N-1}n\binom{N-1}{n}\,e^{\lambda\tau_n}\prod_{j=0}^{n-1}(e^{\lambda\tau_j}-1) \tag{2.54a}$$

By use of (2.45b), however, this reduces to

$$E[I]=\overline{R}+(N-1)b\alpha \tag{2.54b}$$

From (2.32) and (2.53b), it can be shown that

$$\alpha = 1 - I^*(\lambda) \tag{2.55a}$$

which implies the fact that a message is found at the polling when there is an arrival during the intervisit time. As a matter of fact, (2.55a) means

$$K\sum_{n=0}^{N-1}\binom{N-1}{n}\prod_{j=0}^{n}(e^{\lambda\tau_j}-1)$$

$$=1-K[1+\sum_{n=1}^{N-1}\binom{N-1}{n}\prod_{j=0}^{n-1}(e^{\lambda\tau_j}-1)] \tag{2.55b}$$

The l.h.s. is equal to (by replacing n by n-1)

$$K\sum_{n=1}^{N}\binom{N-1}{n-1}\prod_{j=0}^{n-1}(e^{\lambda\tau_j}-1)$$

Using $\binom{N-1}{n}+\binom{N-1}{n-1}=\binom{N}{n}$ and (2.29b), it is clear that (2.55b) holds.

Lastly we express the LST $W^*(s)$ of the waiting time distribution function in terms of $I^*(s)$. Let $w(y)$ be the pdf for the waiting time.

The waiting time is y when there is an arrival at time t-y after the start of the intervisit time whose duration is t. Thus,

$$w(y) = \frac{\int_y^\infty \lambda e^{-\lambda(t-y)} dI(t)}{\int_0^\infty (1-e^{-\lambda t})\, dI(t)} \tag{2.56a}$$

where I(t) is the distribution function of the intervisit time. The denominator of (2.56a) is the probability of arrival in the intervisit time, and equal to $1-I^*(\lambda)=\alpha$. The Laplace transform of the numerator of (2.56a) is

$$\int_0^\infty e^{-sy}dy\int_y^\infty \lambda e^{-\lambda(t-y)}dI(t) = \int_0^\infty e^{-st}dI(t)\int_0^t \lambda e^{-(\lambda-s)(t-y)}dy$$

$$= \frac{\lambda}{\lambda-s}\int_0^\infty (e^{-st} - e^{-\lambda t})dI(t) = \frac{\lambda[I^*(s) - I^*(\lambda)]}{\lambda - s}$$

Thus we have

$$W^*(s) = \frac{\lambda[I^*(s) - I^*(\lambda)]}{(\lambda-s)[1 - I^*(\lambda)]} \tag{2.56b}$$

From (2.56b), the mean waiting time is given by

$$E[W] = \frac{E[I]}{\alpha} - \frac{1}{\lambda} \tag{2.57}$$

From (2.54b) and (2.57), we recover (2.44a).

3 Exhaustive-Service, Discrete-Time System

We consider a discrete-time, exhaustive-service polling system where each station has an infinite buffer capacity to store outstanding packets. The time is slotted with slot size equal to the service time of a fixed-length packet, and is measured by slots. We call the time interval [t, t+1] the t th slot.

The packet arrival process at each station is assumed to be independent of those at other stations. Let

$X_i(t) \triangleq$ number of packets arriving at station i in the t th slot

For each i, $\{X_i(t);\ t = 0, 1, 2, \ldots\}$ is assumed to be an independent and identically distributed sequence of random variables. The GF, mean and variance of $X_i(t)$ are given by

$$P_i(z) \triangleq E[z^{X_i(t)}] \qquad \mu_i \triangleq E[X_i(t)] = P_i^{(1)}(1)$$

$$\sigma_i^2 \triangleq \mathrm{Var}[X_i(t)] = P_i^{(2)}(1) + P_i^{(1)}(1) - [P_i^{(1)}(1)]^2 \qquad (3.1)$$

We assume that N stations are polled in the order of their indices $i = 1, 2, \ldots, N$. A polling cycle is defined as the time interval beginning at the polling instant of station 1 and ending with the completion of reply interval from station N to station 1. Let us denote by $\tau_i(m)$ the polling instant of station i in the m th cycle, and by $\bar{\tau}_i(m)$ the instant when the reply interval between station i and station i + 1 starts. In other words, if there is any outstanding service request at station i at time $t = \tau_i(m)$, the difference $\bar{\tau}_i(m) - \tau_i(m)$ represents the time to serve station i. If there is no outstanding request, then $\bar{\tau}_i(m) = \tau_i(m)$.

Using these notations, we have $\tau_{i+1}(m) - \bar{\tau}_i(m)$ = length of reply interval between station i and station i + 1 in the m th cycle. For each i, it is assumed that $\{\tau_{i+1}(m) - \bar{\tau}_i(m)\ ;\ m = 1, 2, \ldots\}$ is a sequence of independent and identically distributed random variables. The GF, mean and variance of $\tau_{i+1}(m) - \bar{\tau}_i(m)$ are given by

$$R_i(z) \triangleq E[z^{\tau_{i+1}(m) - \bar{\tau}_i(m)}] \qquad r_i \triangleq E[\tau_{i+1}(m) - \bar{\tau}_i(m)] = R_i^{(1)}(1)$$

$$\delta_i^2 \triangleq \mathrm{Var}[\tau_{i+1}(m) - \bar{\tau}_i(m)] = R_i^{(2)}(1) + R_i^{(1)}(1) - [R_i^{(1)}(1)]^2 \qquad (3.2)$$

3.1 Bulk Arrival Process

In the above, the arrival process at each station is defined in terms of the number of (fixed-size) packets. However, one may think of the case where each station gets arrivals of variable-length messages such that each message consists of several packets. To consider such a case, let

$\alpha_i(t) \triangleq$ number of messages arriving at station i in the t th slot

$\beta_i \triangleq$ number of packets included in a message at station i

$\{\alpha_i(t)\}$ and $\{\beta_i\}$ are assumed to be mutually independent. Define the GF and the moments of these random variables as

$$A_i(z) \triangleq E[z^{\alpha_i(t)}] \qquad \lambda_i \triangleq E[\alpha_i(t)] = A_i^{(1)}(1)$$

$$\sigma_i^2 \triangleq \mathrm{Var}[\alpha_i(t)] = A_i^{(2)}(1) + A_i^{(1)}(1) - [A_i^{(1)}(1)]^2 \qquad (3.3)$$

and

$$B_i(z) \triangleq E[z^{\beta_i}] \qquad b_i \triangleq E[\beta_i] = B_i^{(1)}(1)$$

$$b_i^{(2)} \triangleq E[\{\beta_i\}^2] = B_i^{(2)}(1) + B_i^{(1)}(1) \tag{3.4}$$

Note that $B_i(0) = 0$ by definition.

The distribution for the number of packets arriving at station i in a slot is then given by the compound distribution

$$P_i(z) = A_i[B_i(z)] \tag{3.5a}$$

from which we get the relationship

$$\mu_i = \lambda_i b_i, \qquad \sigma_i^2 = (\gamma_i^2 - \lambda_i) b_i^2 + \lambda_i b_i^{(2)} \tag{3.5b}$$

In the following, all analysis is carried out in terms of $P_i(z)$, μ_i, and σ_i^2. However, by use of (3.5a) and (3.5b), one can readily convert the results into the form expressed in terms of message-related notation. On the other hand, one has only to let $B_i(z) = z$ (which implies that $b_i = b_i^{(2)} = 1$, $A_i(z) = P_i(z)$, $\lambda_i = \mu_i$ and $\gamma_i^2 = \sigma_i^2$) when he wants to change the message-related notation to the packet-related notation.

Let us now call all the packets arriving at each station in a slot a supermessage. In any slot, a supermessage does not arrive at station i with probability $P_i(0)$, and arrives with probability $1 - P_i(0)$ (Bernoulli arrival process). We are interested in the position of an arbitrarily chosen packet (called the tagged packet) within a supermessage. Note that this is the delay from the instant when the first packet in the supermessage is started service to the instant when the service for our tagged packet is started.

For each i, let M_i be the number of packets included in a supermessage, and Y_i be the number of packets before the tagged packet within the supermessage. See Figure 3. Since the distribution for M_i is given by

$$P\{M_i = k\} = \frac{P\{X_i(t) = k\}}{1 - P_i(0)} \qquad k = 1, 2, \ldots \tag{3.6a}$$

we have the GF for M_i, $M_i(z)$, as

$$M_i(z) = \frac{P_i(z) - P_i(0)}{1 - P_i(0)} \tag{3.6b}$$

Note that Y_i is the backward recurrence time in the (discrete-time) renewal process where the interevent time distribution is given by (3.6a). Thus, the distribution for Y_i is given by

$$P\{Y_i = k\} = \frac{P\{M_i > k\}}{E[M_i]} \qquad k = 0, 1, 2, \ldots \tag{3.7}$$

The GF for Y_i, $Y_i(z)$, is now found as

$$Y_i(z) = \frac{1}{E[M_i]} \sum_{k=0}^{\infty} P\{M_i > k\} z^k = \frac{1}{E[M_i]} \sum_{k=0}^{\infty} z^k \sum_{j=k+1}^{\infty} P\{M_i = j\}$$

$$= \frac{1}{E[M_i]} \sum_{j=1}^{\infty} P\{M_i = j\} \sum_{k=0}^{j-1} z^k = \frac{1}{E[M_i]} \sum_{j=1}^{\infty} \frac{1 - z^j}{1 - z} P\{M_i = j\}$$

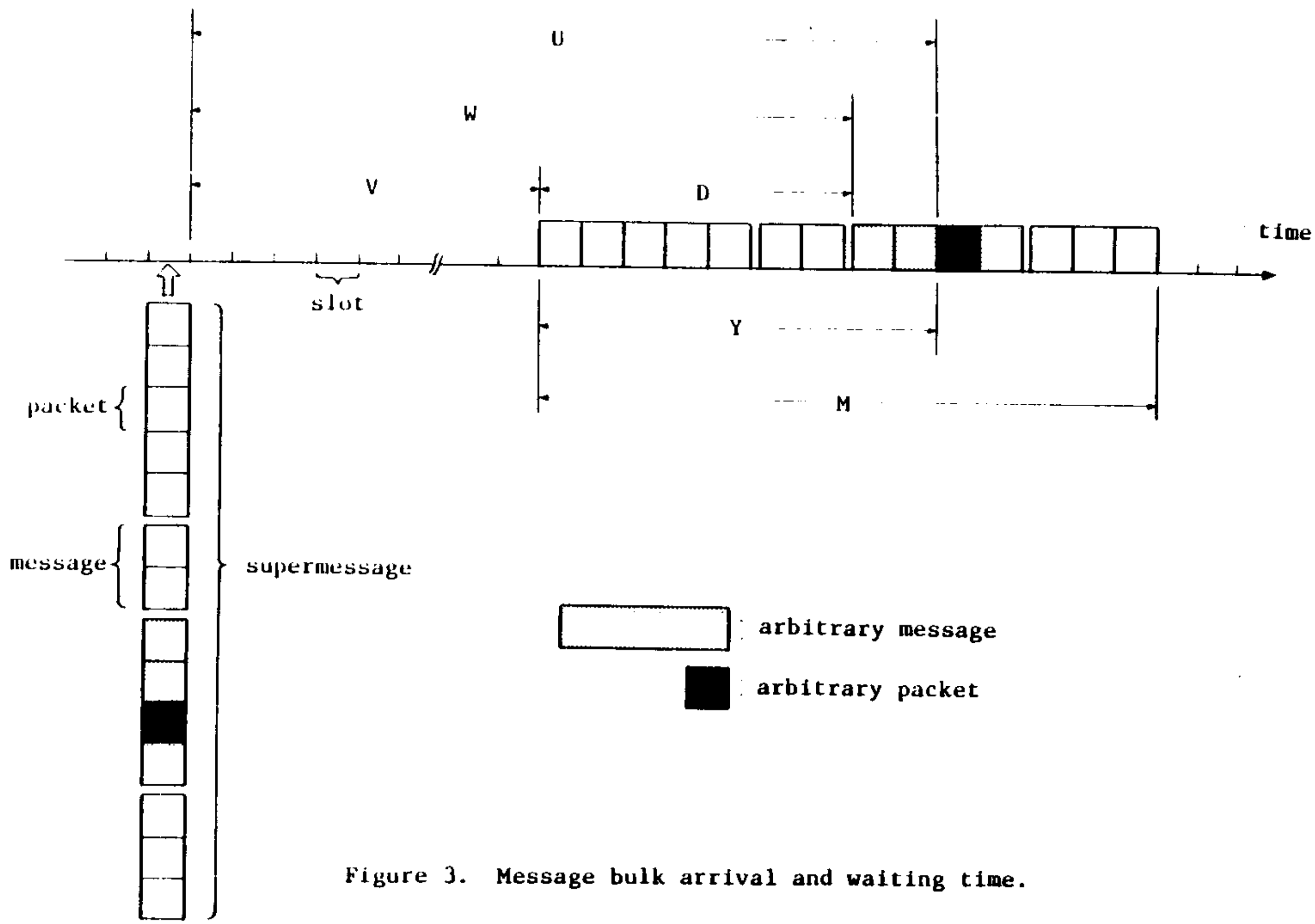

Figure 3. Message bulk arrival and waiting time.

or

$$Y_i(z) = \frac{1 - M_i(z)}{E[M_i](1-z)} = \frac{1 - P_i(z)}{\mu_i(1-z)} \tag{3.8a}$$

From (3.6b) and (3.8a), we get

$$E[Y_i] = \frac{M_i^{(2)}(1)}{2M_i^{(1)}(1)} = \frac{P_i^{(2)}(1)}{2P_i^{(1)}(1)} = \frac{\sigma_i^2 + \mu_i^2 - \mu_i}{2\mu_i} \tag{3.8b}$$

This is the average packet delay within a supermessage.

We next concider the position of an arbitrarily chosen message (called the tagged message) within a supermessage. We are interested in the number of packets included in the messages before the tagged message within a supermessage, which we denote by D_i. Note that D_i is the delay in service from the instant when the first packet in the supermessage is started service to the instant when the service for the tagged message is started. See Figure 3.

To find the distribution of D_i, let Z_i be the number of messages before the tagged message within a supermessage. Recalling that $\alpha_i(t)$ is the number of messages in the supermessage, we have the distribution of Z_i as

$$P\{Z_i = k\} = \frac{P\{\alpha_i(t) > k\}}{E[\alpha_i(t)]} = \frac{1}{\alpha_i} P\{\alpha_i(t) > k\} \qquad k = 0, 1, 2, \ldots \tag{3.9}$$

Under the condition that $Z_i = k$, the GF for D_i is given by $[B_i(z)]^k$. Thus, by unconditioning we have the GF for D_i, $D_i(z)$, as

$$D_i(z) = \sum_{k=0}^{\infty} [B_i(z)]^k \, P\{Z_i = k\} \tag{3.10a}$$

Substituting (3.9) and manipulating the expression in the same way as that leads to (3.8a), we get

$$D_i(z) = \frac{1 - A_i[B_i(z)]}{\lambda_i[1 - B_i(z)]} \tag{3.10b}$$

From (3.10b), the average delay for the tagged message within the supermessage is given by

$$E[D_i] = \frac{(\gamma_i^2 - \lambda_i + \lambda_i^2)\, b_i}{2\lambda_i} \tag{3.10c}$$

In Section 3.6, we first find the mean waiting time for the supermessage (denoted by $E[V_i]$). Then we have the mean packet waiting time ($E[U_i]$) and mean message waiting time ($E[W_i]$) by adding $E[Y_i]$ and $E[D_i]$, respectively, to $E[V_i]$.

3.2 Gambler's Ruin Problem

In this section, we study the gambler's ruin problem which is essentially the busy period analysis for a discrete-time queueing system. Let us consider a gambler who starts with an initial capital L_0 (≥ 0) and plays a sequence of independent and identical games. The gain on the n th game is X_n from which is subtracted a playing fee of 1 unit. Define

$$F(z) \triangleq E[z^{L_0}], \quad P(z) \triangleq E[z^{X_n}], \quad \mu \triangleq E[X_n] < 1, \quad \sigma^2 \triangleq Var[X_n] \tag{3.11}$$

If L_n denotes the remaining capital after the n th game, then

$$L_n = L_0 + X_1 + X_2 \dots + X_n - n \qquad n \geq 1 \tag{3.12}$$

The gambler's ruin occurs after the T th game where

$$T = \min\{n: L_n = 0\} \tag{3.13}$$

Note that T = 0 if $L_0 = 0$. See Figure 4 for a sample path of L_n. Note that, in the context of busy period analysis, T is the length of busy period beginning with L_0 packets with arrival process given by P(z).

Let $g_{n,k}$ be the probability of the event {no ruin before the n th game, and a capital equal to k after the n th game}. That is,

$$g_{n,k} = P\{L_j > 0 \ (0 \leq j < n), \ L_n = k\} \tag{3.14}$$

This event can occur in any of the k + 1 (mutually exclusive) ways: {no ruin before the (n - 1)st game, and a capital equal to j after the (n - 1)st game} followed by {a gain of k - j + 1 on the n th game} ($1 \leq j \leq k + 1$). Thus,

$$g_{n,k} = \sum_{j=1}^{k+1} g_{n-1,j} P\{X_n = k - j + 1\} \qquad n = 1, 2, \dots; \quad k = 0, 1, 2, \dots \tag{3.15a}$$

and

$$g_{0,k} = P\{L_0 = k\} \tag{3.15b}$$

We introduce the GF's

$$G_n(z) \triangleq \sum_{k=0}^{\infty} g_{n,k} z^k \qquad n = 0, 1, 2, \dots; \qquad G(z,\omega) \triangleq \sum_{n=0}^{\infty} G_n(z)\omega^n \tag{3.16a}$$

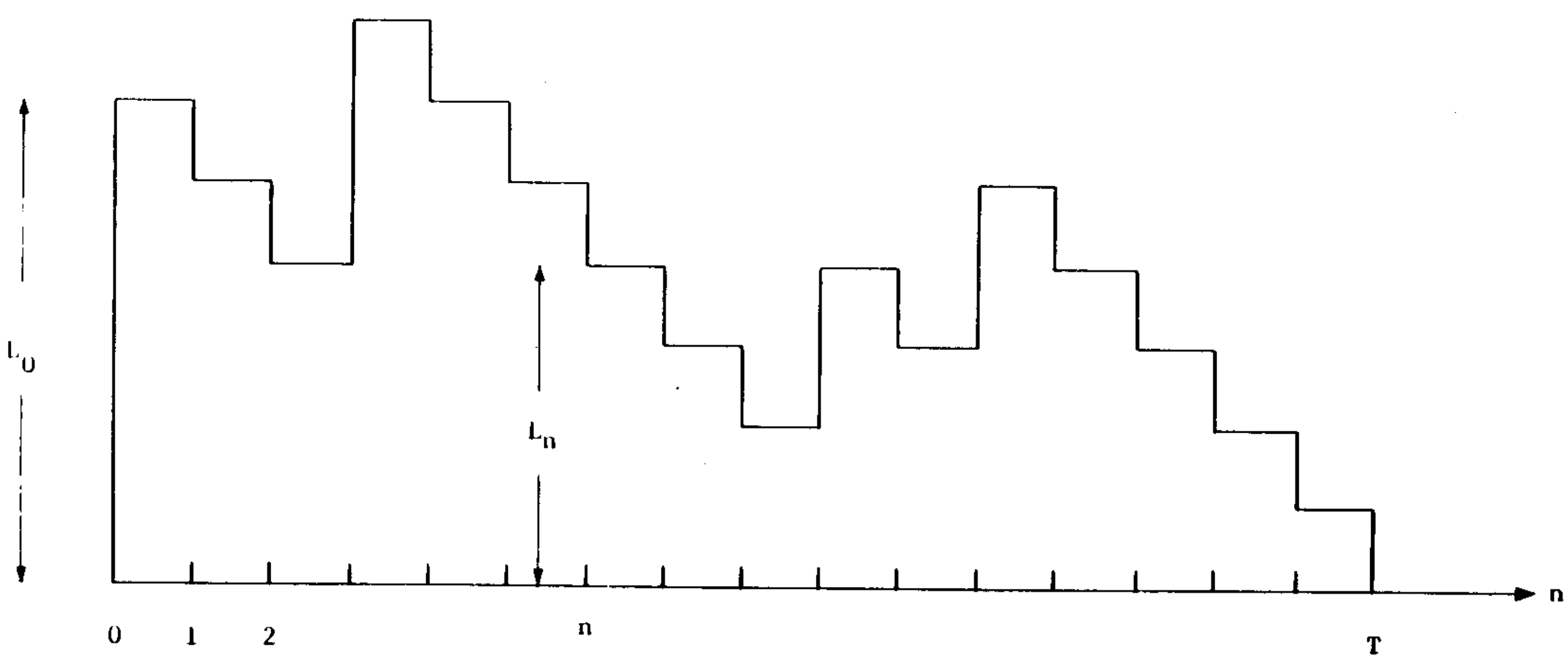

Figure 4. A sample path of L_n.

Note that

$$g_{n,0} = G_n(0) = P\{L_j > 0 \ (0 \le j < n),\ L_n = 0\} = P\{T = n\}$$

and so

$$G(0,\omega) = \sum_{n=0}^{\infty} G_n(0)\omega^n = \sum_{n=0}^{\infty} P\{T = n\}\omega^n = E[\omega^T] \qquad (3.16b)$$

is the GF for the ruin time T.

Let us express $G(z,\omega)$ in terms of $F(z)$, $P(z)$ and $G(0,\omega)$. To do so, first multiply (3.15a) and (3.15b) by z^k, and sum over k to obtain

$$G_n(z) = \sum_{k=0}^{\infty} z^k \sum_{j=1}^{k+1} g_{n-1,j} P\{X_n = k-j+1\} = \sum_{j=1}^{\infty} g_{n-1,j} z^{j-1} \sum_{k=j-1}^{\infty} P\{X_n = k-j+1\} z^{k-j+1}$$

$$= \frac{1}{z}[G_{n-1}(z) - G_{n-1}(0)]P(z) \qquad n = 1, 2, \ldots$$

and

$$G_0(z) = \sum_{k=0}^{\infty} g_{0,k} z^k = \sum_{k=0}^{\infty} P\{L_0 = k\} z^k = F(z)$$

Therefore

$$G(z,\omega) = G_0(z) + \sum_{n=1}^{\infty} G_n(z)\omega^n = F(z) + \frac{\omega}{z} P(z) \sum_{n=0}^{\infty} [G_n(z) - G_n(0)]\omega^n$$

$$= F(z) + \frac{\omega}{z} P(z)[G(z,\omega) - G(0,\omega)]$$

Solving the last equation for $G(z,\omega)$, we get

$$G(z,\omega) = \frac{zF(z) - \omega P(z)G(0,\omega)}{z - \omega P(z)} \tag{3.17}$$

In (3.17), $G(0,\omega)$ is unknown. We determine $G(z,\omega)$ from (3.17) by an appeal to the analytic property of $G(z,\omega)$ within a unit circle $|z| = 1$ on the z-plane. Let us use Rouche's theorem to show that there is a unique zero for the function in the denominator of (3.17). Since $P(z)$ is a GF, $|P(z)| \leq 1$ for $|z| \leq 1$. So, for each ω, $|\omega| < 1$, we have $|\omega P(z)| < |z|$ on $|z| = 1$. Hence, by Rouche's theorem, the functions z and $z - \omega P(z)$ have the same number of zeros within $|z| = 1$. Thus, there exists a unique root, $z = \Theta(\omega)$, of the equation $z - \omega P(z) = 0$ within $|z| = 1$. That is,

$$\Theta(\omega) - \omega P[\Theta(\omega)] = 0 \qquad |\Theta(\omega)| < 1 \tag{3.18a}$$

Note that $\Theta(\omega)$ is the GF for the ruin time when $L_0 = 1$. From (3.18a), we may find

$$\Theta(1) = 1, \quad \Theta^{(1)}(1) = \frac{1}{1 - \mu}, \quad \Theta^{(2)}(1) = \frac{\mu}{(1 - \mu)^2} + \frac{\sigma^2}{(1 - \mu)^3} \tag{3.18b}$$

From the analyticity of $G(z,\omega)$ within $|z| = 1$, the numerator of (3.17) must also be zero at $z = \Theta(\omega)$. So, by use of (3.18a),

$$0 = \Theta(\omega)F[\Theta(\omega)] - \omega P[\Theta(\omega)]G(0,\omega)$$

$$= \Theta(\omega)F[\Theta(\omega)] - \Theta(\omega)G(0,\omega) = \Theta(\omega)\{F[\Theta(\omega)] - G(0,\omega)\}$$

Since $|\omega| < 1$, (3.18a) implies $\Theta(\omega) \neq 0$. (If $\Theta(\omega) = 0$, then $P(0) = 0$ which implies $|\omega| \geq 1$.) Thus we have

$$E[\omega^T] = G(0,\omega) = F[\Theta(\omega)] \tag{3.19a}$$

From (3.18b) and (3.19a), the mean and variance of the ruin time T are found as

$$E[T] = \frac{E[L_0]}{1 - \mu}, \quad Var[T] = \frac{Var[L_0]}{(1-\mu)^2} + \frac{\sigma^2 E[L_0]}{(1-\mu)^3} \tag{3.19b}$$

Later we need the following formula:

$$E\left\{\sum_{n=0}^{T-1} z^{L_n}\right\} = z\,\frac{F(z) - 1}{z - P(z)} \tag{3.20a}$$

Proof. Noting that T and L_n are random variables (and that $L_n \geq 1$ for $n \leq T - 1$), we have

$$E\left\{\sum_{n=0}^{T-1} z^{L_n}\right\} = \sum_{j=1}^{\infty} \sum_{n=0}^{j-1} \sum_{k=1}^{\infty} z^k P\{T=j,\ L_n=k\} = \sum_{n=0}^{\infty} \sum_{k=1}^{\infty} z^k \sum_{j=n+1}^{\infty} P\{T=j,\ L_n=k\}.$$

However, for $k \geq 1$,

$$\sum_{j=n+1}^{\infty} P\{T=j,\ L_n=k\} = P\{T>n,\ L_n=k\} = g_{n,k}$$

Therefore,

$$E\left\{\sum_{n=0}^{T-1} z^{L_n}\right\} = \sum_{n=0}^{\infty} \sum_{k=1}^{\infty} z^k g_{n,k} = \sum_{n=0}^{\infty} [G_n(z) - G_n(0)] = G(z,1) - G(0,1)$$

From (3.19a), $G(0,1) = F[\Theta(1)] = F(1) = 1$. Using (3.17) for $G(z,1)$, we get

$$E\left\{\sum_{n=0}^{T-1} z^{L_n}\right\} = \frac{zF(z) - P(z)}{z - P(z)} - 1 = z\,\frac{F(z) - 1}{z - P(z)} \qquad \text{q.e.d.}$$

Similarly, we also have

$$E\left\{\sum_{n=1}^{T} z^{L_n}\right\} = P(z)\,\frac{F(z) - 1}{z - P(z)} \qquad (3.20b)$$

3.3 Number of Packets at Polling Instants

Our first analysis is for the number of packets found in each station at the time when some station is polled. Let

$L_i(t) \triangleq$ number of packets at station i at time t

(this does not include the arrivals in [t, t + 1]) and define the joint GF for $[L_1(t), L_2(t), \ldots, L_N(t)]$ at time $t = \tau_i(m)$, i.e., at the instant when the server becomes available to station i by

$$F_i(z_1, z_2, \ldots, z_N) \triangleq E\left\{\prod_{j=1}^{N} z_j^{L_j(\tau_i(m))}\right\} \qquad (3.21a)$$

Similarly, we define the marginal GF for $L_i(t)$ at time $t = \tau_i(m)$ by

$$F_i(z) \triangleq E[z^{L_i(\tau_i(m))}] = F_i(1, \ldots, 1, z, 1, \ldots, 1) \qquad (3.21b)$$

where z is the i th argument in $F_i(1, \ldots, 1, z, 1, \ldots, 1)$. We will express $F_{i+1}(z_1, z_2, \ldots, z_N)$ in terms of $F_i(z_1, z_2, \ldots, z_N)$.

First, consider the <u>service time</u> for station i: $[\tau_i(m), \bar{\tau}_i(m)]$. During this period, the number of packets at station i changes just as the capital in the gambler's ruin problem (Section 3.2) where the initial capital corresponds to $L_i(\tau_i(m))$, the number of packets found at the polling instants.

In this analogy, the service time $\bar{\tau}_i(m) - \tau_i(m)$ corresponds to the gambler's ruin time. Thus, according to (3.19a), we have the GF for the service time:

$$S_i(z) \triangleq E[z^{\bar{\tau}_i(m)-\tau_i(m)}] = F_i[\Theta_i(z)] \tag{3.22}$$

where $F_i(z)$ is given by (3.21b), and $\Theta_i(\omega)$ satisfies (see (3.18a))

$$\Theta_i(z) - zP_i[\Theta_i(z)] = 0 \tag{3.23a}$$

from which we have (see (3.18b))

$$\Theta_i(1) = 1, \quad \Theta_i^{(1)}(1) = \frac{1}{1-\mu_i}, \quad \Theta_i^{(2)}(1) = \frac{\mu_i}{(1-\mu_i)^2} + \frac{\sigma_i^2}{(1-\mu_i)^3} \tag{3.23b}$$

We now consider the joint GF for $[L_1(t), L_2(t), \ldots, L_N(t)]$ at time $t = \bar{\tau}_i(m)$, i.e., at the time when the service at station i is completed. Since $L_i(\bar{\tau}_i(m)) = 0$,

$$E\left\{\prod_{j=1}^{N} z_j^{L_j(\bar{\tau}_i(m))}\right\} = E\left\{\prod_{\substack{j=1\\(j\neq i)}}^{N} z_j^{L_j(\bar{\tau}_i(m))}\right\}$$

$$= \sum_{k_1=0}^{\infty} \cdots \sum_{k_{i-1}=0}^{\infty} \sum_{k_{i+1}=0}^{\infty} \cdots \sum_{k_N=0}^{\infty} P\{L_j(\tau_i(m))=k_j (1\le j\le N, j\neq i)\}$$

$$\cdot E\left\{\prod_{\substack{j=1\\(j\neq i)}}^{N} z_j^{L_j(\bar{\tau}_i(m))} \middle| L_j(\tau_i(m))=k_j (1\le j\le N, j\neq i)\right\}$$

However, for station j ($\neq$i), the number of packets at $t = \bar{\tau}_i(m)$ is the number of packets at $t = \tau_i(m)$ plus the number of arrivals during $[\tau_i(m), \bar{\tau}_i(m)]$:

$$L_j(\bar{\tau}_i(m)) = L_j(\tau_i(m)) + \text{number of arrivals in } [\tau_i(m), \bar{\tau}_i(m)] \qquad j \neq i$$

Also, by the notion of compound distribution, the joint GF for the number of packets arriving at stations 1, 2, ..., i-1, i+1, ..., N during $[\tau_i(m), \bar{\tau}_i(m)]$, given that $L_j(\tau_i(m)) = k_j (1 \leq j \leq N,\ j \neq i)$, is given by

$$E\left\{ \{ \prod_{\substack{j=1 \\ (j \neq i)}}^{N} P_j(z_j) \}^{\bar{\tau}_i(m) - \tau_i(m)} \middle| L_j(\tau_i(m)) = k_j (1 \leq j \leq N,\ j \neq i) \right\}$$

It follows that

$$E\left\{ \prod_{\substack{j=1 \\ (j \neq i)}}^{N} z_j^{L_j(\bar{\tau}_i(m))} \middle| L_j(\tau_i(m)) = k_j (1 \leq j \leq N,\ j \neq i) \right\}$$

$$= E\left\{ \prod_{\substack{j=1 \\ (j \neq i)}}^{N} z_j^{k_j} \{P_j(z_j)\}^{\bar{\tau}_i(m) - \tau_i(m)} \middle| L_j(\tau_i(m)) = k_j (1 \leq j \leq N,\ j \neq i) \right\}$$

$$= \prod_{\substack{j=1 \\ (j \neq i)}}^{N} z_j^{k_j} \, E\left\{ \{ \prod_{\substack{j=1 \\ (j \neq i)}}^{N} P_j(z_j) \}^{\bar{\tau}_i(m) - \tau_i(m)} \middle| L_j(\tau_i(m)) = k_j (1 \leq j \leq N,\ j \neq i) \right\}$$

$$= \prod_{\substack{j=1 \\ (j \neq i)}}^{N} z_j^{k_j} \, E\left\{ \{ \prod_{\substack{j=1 \\ (j \neq i)}}^{N} P_j(z_j) \}^{\bar{\tau}_i(m) - \tau_i(m)} \right\}$$

$$= \prod_{\substack{j=1 \\ (j \neq i)}}^{N} z_j^{k_j} \, E\left\{ \text{[illegible]}_i[\prod_{\substack{j=1 \\ (j \neq i)}}^{N} P_j(z_j)]^{L_i(\tau_i(m))} \right\}$$

where (3.21b) and (3.22) have been used to get the last expression. Thus, by unconditioning, we obtain

$$E\left\{\prod_{j=1}^{N} z_j^{L_j(\bar{\tau}_i(m))}\right\} = E\left\{\prod_{\substack{j=1\\(j\neq i)}}^{N} z_j^{L_j(\tau_i(m))} \{P_j(z_j)\}^{\bar{\tau}_i(m)-\tau_i(m)}\right\}$$

$$= E\left\{\{\prod_{\substack{j=1\\(j\neq i)}}^{N} z_j^{L_j(\tau_i(m))}\}\cdot\{\Theta_i[\prod_{\substack{j=1\\(j\neq i)}}^{N} P_j(z_j)]\}^{L_i(\tau_i(m))}\right\}$$

$$= F_i(z_1,z_2,\ldots,z_{i-1},\ \Theta_i[\prod_{\substack{j=1\\(j\neq i)}}^{N} P_j(z_j)],\ z_{i+1},\ldots,z_N) \tag{3.24}$$

Next, consider the reply interval between station i and station i + 1 : $[\bar{\tau}_i(m),\ \tau_{i+1}(m)]$. During this period, only the packet arrivals change $[L_1(t),\ L_2(t),\ \ldots,\ L_N(t)]$. The joint GF for the number of packets arriving at all stations during the reply interval $[\bar{\tau}_i(m),\ \tau_{i+1}(m)]$ whose distribution is specified by $R_i(z)$ is given as a compound distribution:

$$R_i[\prod_{j=1}^{N} P_j(z_j)] \tag{3.25}$$

Since the event in $[\tau_i(m),\ \bar{\tau}_i(m)]$ and the event in $[\bar{\tau}_i(m),\ \tau_{i+1}(m)]$ are independent, the joint GF for $[L_1(t),\ L_2(t),\ \ldots,\ L_N(t)]$ at time $t = \tau_{i+1}(m)$, i.e., at the time of polling to station i + 1 is given by the product of (3.24) and (3.25):

$$F_{i+1}(z_1,z_2,\ldots,z_N)=R_i[\prod_{j=1}^{N} P_j(z_j)]\cdot F_i(z_1,z_2,\ldots z_{i-1},\Theta_i[\prod_{\substack{j=1\\(j\neq i)}}^{N} P_j(z_j)],z_{i+1},\ldots,z_N) \tag{3.26}$$

3.3.1 Mean and Variance

We do not solve the equation (3.26). (See [Konh74] and [Swar80] for the solution.) Instead, we use (3.26) to find the means and covariances for the number of packets at each station when some station is polled. To write down the recursive relations among them, let

$$f_i(j) \triangleq \left. \frac{\partial F_i(z_1, z_2, \ldots, z_N)}{\partial z_j} \right|_{z_1 = z_2 = \ldots = z_N = 1}$$

$$f_i(j,k) \triangleq \left. \frac{\partial^2 F_i(z_1, z_2, \ldots, z_N)}{\partial z_j \partial z_k} \right|_{z_1 = z_2 = \ldots = z_N = 1}$$

$$i,\ j,\ k = 1,\ 2,\ \ldots,\ N \qquad (3.27)$$

Note that

$$F_i^{(1)}(1) = f_i(i), \quad F_i^{(2)}(1) = f_i(i,i) \qquad (3.28a)$$

Note also that, if L_i^* denotes the number of packets at station i when it is polled, then

$$E[L_i^*] = f_i(i), \quad Var[L_i^*] = f_i(i,i) + E[L_i^*] - \{E[L_i^*]\}^2 \qquad (3.28b)$$

In Appendix A, a set of N^2 equations for $\{f_i(j); i,j=1,2,\ldots,N\}$ are shown to be

$$f_{i+1}(i) = r_i \lambda_i \qquad (3.29a)$$

$$f_{i+1}(j) = r_i \lambda_j + f_i(j) + \frac{f_i(i) \lambda_j}{1 - \rho_i} \qquad j \neq i \qquad (3.29b)$$

It is also shown there that the solution to the equations (3.29a) and (3.29b) is given by

$$E[L_i^*] = f_i(i) = \frac{\mu_i(1-\mu_i)\sum_{k=1}^{N} r_k}{1-\sum_{k=1}^{N}\mu_k} \tag{3.30a}$$

$$f_i(j) = \mu_j\left\{\sum_{k=j}^{i-1} r_k + \frac{\left[\sum_{k=j+1}^{i-1}\mu_k\right]\left[\sum_{k=1}^{N} r_k\right]}{1-\sum_{k=1}^{N}\mu_k}\right\} \qquad j \neq i \tag{3.30b}$$

In Appendix B, a set of N^3 equations for $\{f_i(j,k); i,j,k = 1,2,\ldots,N\}$ are derived to be

$$f_{i+1}(j,k) = \mu_j\mu_k(\sigma_i^2 + r_i^2) + r_i\mu_k f_i(j) + r_i\mu_j f_i(k) + f_i(i)\mu_j\mu_k\left\{\frac{2r_i}{1-\mu_i} + \frac{1}{(1-\mu_i)^2} + \frac{\sigma_i^2}{(1-\mu_i)^3}\right\}$$

$$+ \frac{f_i(i,j)\mu_k + f_i(i,k)\mu_j}{1-\mu_i} + f_i(j,k) + \frac{f_i(i,i)\mu_j\mu_k}{(1-\mu_i)^2} \qquad i\neq j,\ i\neq k,\ j\neq k \tag{3.31a}$$

$$f_{i+1}(j,j) = \mu_j^2(\sigma_i^2 + r_i^2) + r_i(\sigma_j^2 - \mu_j) + 2r_i\mu_j f_i(j) + f_i(j,j) + \frac{2f_i(i,j)\mu_j}{1-\mu_i} + \frac{f_i(i,i)\mu_j^2}{(1-\mu_i)^2}$$

$$+ f_i(i)\left\{\frac{\sigma_j^2 - \mu_j}{1-\mu_i} + \mu_j^2\left(\frac{2r_i}{1-\mu_i} + \frac{1}{(1-\mu_i)^2} + \frac{\sigma_i^2}{(1-\mu_i)^3}\right)\right\} \qquad i\neq j \tag{3.31b}$$

$$f_{i+1}(i,k) = \mu_i\mu_k(\delta_i^2 + r_i^2) + r_i\mu_i\left(f_i(k) + \frac{f_i(i)\mu_k}{1 - \mu_i}\right) \quad i \neq k \tag{3.31c}$$

$$f_{i+1}(i,i) = \mu_i^2(\delta_i^2 + r_i^2) + r_i(\sigma_i^2 - \mu_i) \tag{3.31d}$$

No closed-form solution to the equations (3.31a-d) seems available. They may be solved numerically as a system of N^3 linear equations. In Appendix B, $f_i(i,i)$ in the special case of identical stations is derived. In such a case, dropping the subscripts from the parameters, we evaluate (3.28b) as

$$E[L^*] = \frac{Nr\mu(1 - \mu)}{1 - N\mu} \tag{3.32a}$$

$$\mathrm{Var}[L^*] = \frac{\delta^2\mu^2N(1-\mu)}{1 - N\mu} + \frac{\sigma^2rN[1-(N+1)\mu+(2N-1)\mu^2]}{(1 - N\mu)^2} \tag{3.32b}$$

3.4 Service Time, Intervisit Time and Cycle Time

Let us define the <u>cycle time</u> for station i, C_i, as the period begining at the time of its polling in a cycle and ending at the time of its polling in the next cycle; its duration is given by $\tau_i(m+1) - \tau_i(m)$. Note that the cycle time may also be defined as $\bar{\tau}_i(m+1) - \bar{\tau}_i(m)$ since we are assuming the steady state. The <u>intervisit time</u> for station i, I_i, is defined as the period beginning at the time of its service completion in a cycle and ending at the time when it is polled in the next cycle; its duration is given by $\tau_i(m+1) - \bar{\tau}_i(m)$. Note that the service time duration for station i is given by $\bar{\tau}_i(m) - \tau_i(m)$. In this section, we derive the means and variances for the service time, cycle time and intervisit time.

First, recall that the distribution for the service time at station i is given by (3.22), which is similar to (3.19a). Thus, the mean and the variance for the service time at station i are given by (see (3.19b))

$$E[S_i] = \frac{E[L_i^*]}{1-\mu_i} = \frac{\mu_i \sum_{k=1}^{N} r_k}{1 - \sum_{k=1}^{N} \mu_k} \tag{3.33a}$$

$$Var[S_i] = \frac{Var[L_i^*]}{(1-\mu_i)^2} + \frac{\sigma_i^2 E[L_i^*]}{(1-\mu_i)^3} = \frac{1}{(1-\mu_i)^2}\left(Var[L_i^*] + \frac{\mu_i \sigma_i^2 \sum_{k=1}^{N} r_k}{1 - \sum_{k=1}^{N} \mu_k} \right) \tag{3.33b}$$

where we have used (3.30a) for $E[L_i^*]$. In the case of identical stations, by using (3.32a) and (3.32b) in (3.33a) and (3.33b), we have

$$E[S] = \frac{Nr\mu}{1-N\mu} \tag{3.34a}$$

$$Var[S] = \frac{\delta^2\mu^2 N}{(1-\mu)(1-N\mu)} + \frac{\sigma^2 rN(1+\mu-N\mu)}{(1-\mu)(1-N\mu)^2} \tag{3.34b}$$

Now, we relate the intervisit time duration $\tau_i(m+1) - \bar{\tau}_i(m)$ to the number of packets at the polling instant $t = \tau_i(m+1)$. Noting that the number of packets found at station i at time $t = \tau_i(m+1)$ is equal to the number of arrivals during the preceding intervisit time $[\bar{\tau}_i(m), \tau_i(m+1)]$, we have

$$E[z^{L_i(\tau_i(m+1))}] = E[\{P_i(z)\}^{\tau_i(m+1)-\bar{\tau}_i(m)}] \tag{3.35a}$$

Thus, if we introduce the GF for the intervisit time for station i by

$$I_i(z) \triangleq E[z^{\tau_i(m+1)-\bar{\tau}_i(m)}] \tag{3.35b}$$

then (3.35a) is equivalent to

$$F_i(z) = I_i[P_i(z)] \tag{3.36a}$$

from which we easily get

$$E[L_i^*] = \mu_i E[I_i], \quad Var[L_i^*] = \mu_i^2 Var[I_i] + \sigma_i^2 E[I_i] \tag{3.36b}$$

Solving these equations for $E[I_i]$ and $Var[I_i]$, and using (3.30a), we have the mean and variance of the intervisit time for station i as

$$E[I_i] = \frac{(1-\mu_i)\sum_{k=1}^{N} r_k}{1-\sum_{k=1}^{N}\mu_k}, \quad Var[I_i] = \frac{Var[L_i^*]}{\mu_i^2} - \frac{\sigma_i^2(1-\mu_i)\sum_{k=1}^{N} r_k}{\mu_i^2(1-\sum_{k=1}^{N}\mu_k)} \tag{3.37a}$$

The meaning of (3.36a) is that the number of packets found at the polling instant is equal to the number of arrivals during the intervisit time. In the case of identical stations, these reduce to

$$E[I] = \frac{Nr(1-\mu)}{1-N\mu}, \quad Var[I] = \frac{\delta^2 N(1-\mu)}{1-N\mu} + \frac{\sigma^2 rN(N-1)}{(1-N\mu)^2} \tag{3.37b}$$

We next find the relation between the cycle time and the intervisit time. By conditioning on the length of the intervisit time $\tau_i(m+1) - \bar{\tau}_i(m)$, the distribution for the length of the cycle time $\bar{\tau}_i(m+1) - \bar{\tau}_i(m)$ is given as

$$E[z^{\bar{\tau}_i(m+1)-\bar{\tau}_i(m)}] = E[z^{\bar{\tau}_i(m+1)-\tau_i(m+1)} z^{\tau_i(m+1)-\bar{\tau}_i(m)}]$$

$$= \sum_{n=0}^{\infty} E[z^{\bar{\tau}_i(m+1)-\tau_i(m+1)} z^n | \tau_i(m+1)-\bar{\tau}_i(m)=n] P\{\tau_i(m+1)-\bar{\tau}_i(m)=n\}$$

However, by (3.22), (3.35a) and (3.23a), we see that

$$E[z^{\bar{\tau}_i(m+1)-\tau_i(m+1)} z^n | \tau_i(m+1)-\bar{\tau}_i(m)=n]$$

$$= E[\{\Theta_i(z)\}^{L_i(\tau_i(m+1))} z^n | \tau_i(m+1)-\bar{\tau}_i(m)=n]$$

$$= E[\{P_i[\Theta_i(z)]\}^{\tau_i(m+1)-\bar{\tau}_i(m)} z^n | \tau_i(m+1)-\bar{\tau}_i(m)=n] = \{zP_i[\Theta_i(z)]\}^n = [\Theta_i(z)]^n$$

Thus we have

$$E[z^{\bar{\tau}_i(m+1)-\bar{\tau}_i(m)}] = \sum_{n=0}^{\infty} [\Theta_i(z)]^n P\{\tau_i(m+1)-\bar{\tau}_i(m)=n\} = E[\{\Theta_i(z)\}^{\tau_i(m+1)-\bar{\tau}_i(m)}] \quad (3.38)$$

If we introduce the GF for the cycle time for station i by

$$C_i(z) \triangleq E[z^{\bar{\tau}_i(m+1)-\bar{\tau}_i(m)}] \quad (3.39a)$$

then (3.38) is equivalent to

$$C_i(z) = I_i[\Theta_i(z)] \quad (3.39b)$$

From (3.39b) and (3.37a) the mean and variance of the cycle time for station i can be found as

$$E[C_i] = \frac{\sum_{k=1}^{N} r_k}{1 - \sum_{k=1}^{N} \mu_k} \tag{3.40a}$$

$$Var[C_i] = \frac{1}{\mu_i^2(1-\mu_i)^2}\left\{ Var[L_i^*] + \frac{\sigma_i^2(\mu_i^2+\mu_i-1)\sum_{k=1}^{N} r_k}{1 - \sum_{k=1}^{N} \mu_k} \right\} \tag{3.40b}$$

Note that $E[C_i]$ in (3.40a) is independent of i. In the case of identical stations, these become

$$E[C] = \frac{Nr}{1-N\mu}, \quad Var[C] = \frac{\delta^2 N}{(1-\mu)(1-N\mu)} + \frac{\sigma^2 r N^2}{(1-\mu)(1-N\mu)^2} \tag{3.41}$$

3.4.1 Mean Cycle Time and Stability

An intuitive argument to derive the expression for the mean cycle time in (3.40a) is as follows. Assuming stationarity, we conjecture that the mean cycle time consists of the mean total reply intervals and the mean time to serve the packets arriving during the cycle time:

$$E[C_i] = \sum_{k=1}^{N} r_k + [\sum_{k=1}^{N} \mu_k] E[C_i] \tag{3.42}$$

which certainly leads to (3.40a). However, we must make clear the meaning of stationarity.

In order to consider the stationarity, let us identify a series of <u>regeneration points</u> in the process $[L_1(t), L_2(t), \ldots, L_N(t)]$. Note that the polling instant of station i is not a regeneration point. Time points when station i is polled and all $L_j(\tau_i(m))$, $1 \le j \le N$, are zero are regeneration points. We call the interval between two such successive regeneration points a <u>regenerative cycle</u>. See, for example, [Heym82, Chap.6] for discussion of regenerative processes.

Let M_i be the number of polling cycles in a regenerative cycle, and $C_i^{(m)}$, m = 1, 2, ..., M_i be the duration of the m th polling cycle in a regenerative cycle. The mean polling cycle time $E[C_i]$ is defined by

$$E[C_i] = \frac{E[\sum_{m=1}^{M_i} C_i^{(m)}]}{E[M_i]} \qquad (3.43)$$

As shown in [Stid72] and [Heym82, Chap.10], a sufficient condition for the <u>stationary version</u> of this regenerative process to be a <u>stationary process</u> is that the mean regenerative cycle time is finite:

$$E[\sum_{m=1}^{M_i} C_i^{(m)}] < \infty \qquad (3.44a)$$

This is satisfied if

$$E[C_i] < \infty \qquad (3.44b)$$

Note that (3.44a) and (3.44b) implies that $E[M_i] < \infty$, since $E[C_i] > 0$. Thus, from (3.40a) and (3.44b), a sufficient condition for the existence of the regenerative process is given by

$$\sum_{k=1}^{N} \mu_k < 1 \tag{3.45}$$

Since there is only one server, stationarity means <u>stability</u>. Thus we may claim that (3.45) is the condition for stability. This condition is seen to be independent of the reply intervals, which is to be expected since at near-saturation load the proportion of the server's time spent on reply intervals becomes negligibly small.

We note that [Konh74] also discusses the existence of stationary solution in a different way.

3.5 Number of Packets at Arbitrary Times

We proceed to find the (marginal) distribution for the number of packets at station i at arbitrary times which we denote by L_i. The GF for L_i, denoted by $Q_i(z)$, is given by the time average of $z^{L_i(t)}$ over the regenerative cycle defined in Section 3.4.1:

$$Q_i(z) \triangleq E[z^{L_i(t)}] = \frac{E\left\{\sum_{m=1}^{M_i} \sum_{t=\tau_i(m)}^{\tau_i(m+1)-1} z^{L_i(t)}\right\}}{E\left\{\sum_{m=1}^{M_i} \{\tau_i(m+1)-\tau_i(m)\}\right\}} \tag{3.46}$$

Note here that M_i is a <u>stopping time</u> in the regenerative process with $E[M_i] < \infty$. It follows from <u>Wald's lemma</u> (see [Heym82, Chap.6]) that

$$E\left[\sum_{m=1}^{M_i}\sum_{t=\tau_i(m)}^{\tau_i(m+1)-1} z^{L_i(t)}\right] = E[M_i]\, E\left[\sum_{t=\tau_i(m)}^{\tau_i(m+1)-1} z^{L_i(t)}\right] \tag{3.47a}$$

$$E\left[\sum_{m=1}^{M_i}\{\tau_i(m+1)-\tau_i(m)\}\right] = E[M_i]\, E[\tau_i(m+1)-\tau_i(m)] \tag{3.47b}$$

Thus we have

$$Q_i(z) = \frac{E\left[\sum_{t=\tau_i(m)}^{\tau_i(m+1)-1} z^{L_i(t)}\right]}{E[\tau_i(m+1) - \tau_i(m)]} \tag{3.48}$$

Note that the denominator in (3.48) is simply the mean cycle time duration given by (3.40a).

To evaluate the numerator in (3.48), let us divide the cycle time into the service time and the intervisit time for station i:

$$\sum_{t=\tau_i(m)}^{\tau_i(m+1)-1} = \sum_{t=\tau_i(m)}^{\bar{\tau}_i(m)-1} + \sum_{t=\bar{\tau}_i(m)}^{\tau_i(m+1)-1} \tag{3.49}$$

Note that during the service time for station i, $L_i(t)$ behaves like the capital in the gambler's ruin problem where the initial capital corresponds to the number of packets found at the polling instant. Thus, the sum over the service time can be evaluated by applying (3.20a) with substitution

$$n \to t-\tau_i(m), \quad T \to \bar{\tau}_i(m)-\tau_i(m), \quad L_n \to L_i(t)$$

$$F(z) = E[z^{L_0}] \rightarrow F_i(z) = E[z^{L_i(\tau_i(m))}]$$

We get

$$E\left\{\sum_{t=\tau_i(m)}^{\bar{\tau}_i(m)-1} z^{L_i(t)}\right\} = z\,\frac{F_i(z) - 1}{z - P_i(z)} \tag{3.50a}$$

During the intervisit time for station i, $L_i(t)$ only increases in such a way that the increase per slot is given by the GF $P_i(z)$. Thus, the sum over the intervisit time may be evaluated as

$$E\left\{\sum_{t=\bar{\tau}_i(m)}^{\tau_i(m+1)-1} z^{L_i(t)}\right\} = E\left\{\sum_{t=\bar{\tau}_i(m)}^{\tau_i(m+1)-1} \{P_i(z)\}^{t-\bar{\tau}_i(m)}\right\}$$

$$= \frac{1 - E[\{P_i(z)\}^{\tau_i(m+1)-\bar{\tau}_i(m)}]}{1 - P_i(z)} = \frac{1 - I_i[P_i(z)]}{1 - P_i(z)}$$

By (3.36a), we get

$$E\left\{\sum_{t=\bar{\tau}_i(m)}^{\tau_i(m+1)-1} z^{L_i(t)}\right\} = \frac{1 - F_i(z)}{1 - P_i(z)} \tag{3.50b}$$

Combining (3.40a), (3.50a) and (3.50b), we obtain the GF for L_i, the number of packets at station i at arbitrary instants:

$$Q_i(z) = \frac{1 - \sum_{k=1}^{N} \mu_k}{\sum_{k=1}^{N} r_k} \cdot \frac{1 - F_i(z)}{P_i(z) - z} \cdot \frac{(1-z)P_i(z)}{1 - P_i(z)} \tag{3.51}$$

It is straightforward to calculate $E[L_i] = Q_i^{(1)}(1)$ from (3.51). We get

$$E[L_i] = \frac{E[\{L_i^*\}^2]}{2E[L_i^*]} + \frac{\sigma_i^2}{2}\left(\frac{1}{1 - \mu_i} - \frac{1}{\mu_i}\right) \tag{3.52a}$$

In the case of identical stations, by using (3.32a) and (3.32b), we get

$$E[L] = \frac{\sigma^2 \mu}{2r} + \frac{\sigma^2}{2(1-N\mu)} + \frac{Nr\mu(1-\mu)}{2(1-N\mu)} \tag{3.52b}$$

3.6 Packet Waiting Time

Finally, we derive the GF, $V_i(z)$, for the waiting time (excluding the service time) for the first packet in an arbitrary arrival group (i.e., a supermessage) at station i, which we denote by V_i. See Figure 3. Since the waiting time for the first packet in a group of packets which arrive in the same slot is independent of the number of packets before an arbitrarily chosen packet within the group (whose GF, $Y_i(z)$, is given in (3.8a)), the GF for the waiting time for an arbitrary packet arriving at station i, U_i, is given by

$$U_i(z) = V_i(z)Y_i(z) \tag{3.53a}$$

The mean packet waiting time at station i is then given by

$$E[U_i] = E[V_i] + E[Y_i] \tag{3.53b}$$

We note that V_i is called the <u>virtual</u> waiting time in [Rubi83].

Let us denote by $V_i(t)$ the waiting time for the first packet in a supermessage arriving at station i in slot [t, t+1]; we count the waiting time after time t+1 (see Figure 3). Then, $V_i(z)$ is given by the time average of $z^{V_i(t)}$ over the regenerative cycle as before:

$$V_i(z) \triangleq E[z^{V_i(t)}] = \frac{E\left\{\sum_{m=1}^{M_i} \sum_{t=\tau_i(m)}^{\tau_i(m+1)-1} z^{V_i(t)}\right\}}{E\left\{\sum_{m=1}^{M_i} [\tau_i(m+1)-\tau_i(m)]\right\}} \tag{3.54a}$$

By an argument similar to (3.47a) and (3.47b), we may express $V_i(z)$ as the time average of $z^{V_i(t)}$ over a polling cycle as before.

$$V_i(z) = \frac{E\left\{\sum_{t=\tau_i(m)}^{\tau_i(m+1)-1} z^{V_i(t)}\right\}}{E[\tau_i(m+1) - \tau_i(m)]} \tag{3.54b}$$

The numerator in (3.54b) may be evaluated by dividing the summation domain into the service time and the intervisit time for station i as (3.49).

For an arrival during the service time for station i, the waiting time is only a delay due to the packets already in the station:

$$V_i(t) = L_i(t) - 1 \qquad \tau_i(m) \le t \le \bar{\tau}_i(m) - 1 \tag{3.55a}$$

Thus, by (3.50a), we immediately have

$$E\left\{\sum_{t=\tau_i(m)}^{\bar{\tau}_i(m)-1} z^{V_i(t)}\right\} = \frac{F_i(z) - 1}{z - P_i(z)} \tag{3.55b}$$

For an arrival in the intervisit time, the waiting time consists of the delay due to the packets already in the station, and the delay until the start of the next polling instant for station i:

$$V_i(t) = L_i(t) + \tau_i(m+1) - t - 1 \quad \bar{\tau}_i(m) \le t \le \tau_i(m+1) - 1 \tag{3.56a}$$

Note that $L_i(t)$ is the number of packets which have arrived during $[\bar{\tau}_i(m), t]$ at station i. Therefore,

$$E[z^{L_i(t)}] = \{P_i(z)\}^{t-\bar{\tau}_i(m)}$$

Thus, we get

$$E\left\{\sum_{t=\bar{\tau}_i(m)}^{\tau_i(m+1)-1} z^{V_i(t)}\right\} = E\left\{\sum_{t=\bar{\tau}_i(m)}^{\tau_i(m+1)-1} z^{L_i(t)+\tau_i(m+1)-t-1}\right\}$$

$$= E\left\{\sum_{t=\bar{\tau}_i(m)}^{\tau_i(m+1)-1} \{P_i(z)\}^{t-\bar{\tau}_i(m)} z^{\tau_i(m+1)-t-1}\right\}$$

$$= E\left\{\frac{z^{\tau_i(m+1)-\bar{\tau}_i(m)} - \{P_i(z)\}^{\tau_i(m+1)-\bar{\tau}_i(m)}}{z - P_i(z)}\right\}$$

$$= \frac{E[z^{\tau_i(m+1)-\bar{\tau}_i(m)}] - E[\{P_i(z)\}^{\tau_i(m+1)-\bar{\tau}_i(m)}]}{z - P_i(z)}$$

Using (3.35a) and (3.36a), we get

$$E\left\{ \sum_{t=\bar{\tau}_i(m)}^{\tau_i(m+1)-1} z^{V_i(t)} \right\} = \frac{I_i(z) - F_i(z)}{z - P_i(z)} \tag{3.56b}$$

Combining (3.40a), (3.55b) and (3.56b) for (3.54b), we obtain the GF for V_i, the waiting time for the first packet in an arbitrary arrival at station i:

$$V_i(z) = \frac{1 - \sum_{k=1}^{N} \mu_k}{\sum_{k=1}^{N} r_k} \cdot \frac{I_i(z) - 1}{z - P_i(z)} \tag{3.57a}$$

which may also be written as (see (3.37a))

$$V_i(z) = \frac{I_i(z) - 1}{E[I_i]} \cdot \frac{1 - \mu_i}{z - P_i(z)} \tag{3.57b}$$

From (3.57b), the mean virtual wait $E[V_i] = V_i^{(1)}(1)$ is calculated as

$$E[V_i] = \frac{E[\{I_i\}^2]}{2E[I_i]} + \frac{\sigma_i^2}{2(1 - \mu_i)} - \frac{1 + \mu_i}{2} \tag{3.58a}$$

or, in terms of L_i^*,

$$E[V_i] = \frac{E[\{L_i^*\}^2]}{2\mu_i E[L_i^*]} + \frac{\sigma_i^2}{2\mu_i}\left(\frac{\mu}{1-\mu_i} - \frac{1}{\mu_i}\right) - \frac{1+\mu_i}{2} \tag{3.58b}$$

The GF and mean for the waiting time of an arbitrary packet are respectively given by

$$U_i(z) = \frac{1 - \sum_{k=1}^{N} \mu_k}{\mu_i \sum_{k=1}^{N} r_k} \cdot \frac{1 - P_i(z)}{z - P_i(z)} \cdot \frac{I_i(z) - 1}{1 - z} \tag{3.59a}$$

and

$$E[U_i] = \frac{E[\{L_i^*\}^2]}{2\mu_i E[L_i^*]} + \frac{\sigma_i^2}{2\mu_i}\left(\frac{1}{1-\mu_i} - \frac{1}{\mu_i}\right) - 1 \tag{3.59b}$$

Comparing (3.52a) and (3.59b), we see the relation

$$E[L_i] = \mu_i E[U_i] + \mu_i \tag{3.60}$$

which confirms Little's formula [Litt61].

In the case of identical stations, we have

$$E[V] = \frac{\delta^2}{2r} + \frac{\sigma^2 N}{2(1 - N\mu)} - \frac{1+\mu}{2} + \frac{Nr(1-\mu)}{2(1-N\mu)} \tag{3.61a}$$

and

$$E[U] = \frac{\delta^2}{2r} + \frac{\sigma^2}{2\mu(1 - N\mu)} + \frac{Nr(1-\mu)}{2(1-N\mu)} - 1 \tag{3.61b}$$

We note that our V_i is always less than the "queueing delay" in [Konh74] and [Swar80] by 1 since they count the waiting time from the slot in which a packet arrives while we begin counting them from the next slot.

3.6.1 Mean Message Waiting Time

We denote by W_i the waiting time for an arbitrary message at station i. W_i consists of the waiting time for a supermessage, V_i, and the delay within the supermessage D_i. See Figure 3. Thus we have

$$E[W_i] = E[V_i] + E[D_i] \qquad (3.62)$$

where $E[D_i]$ is given by (3.10c).

We want $E[W_i]$ to be expressed in terms of message-related parameters. [Rubi83] shows the procedure to calculate $E[V_i]$ by analyzing the number of supermessages at polling instants. We here consider only the case of identical stations for which we have (3.61a). Using the conversion (3.5b), we can write (3.61a) as

$$E[V] = \frac{\delta^2}{2r} + \frac{N[(\sigma^2-\lambda)b^2+\lambda b^{(2)}]}{2(1-N\lambda b)} - \frac{1+\lambda b}{2} + \frac{Nr(1-\lambda b)}{2(1-N\lambda b)} \qquad (3.63a)$$

Thus, by adding (3.10c) and (3.63a), we have the mean message waiting time in the case of identical stations:

$$E[W] = \frac{\delta^2}{2r} + \frac{N[\lambda b^{(2)}+r(1-\lambda b)]}{2(1-N\lambda b)} + \frac{(\sigma^2-\lambda)b}{2\lambda(1-N\lambda b)} - \frac{1}{2} \qquad (3.63b)$$

If we let $b = b^{(2)} = 1$ (single-packet messages) and replace λ and γ^2 by μ and σ^2, respectively, in (3.63b), we recover (3.61b).

We may use (3.63b) to derive the mean message waiting time at station 1, $E[W_1]$, when all traffic is generated only at staion 1 (the GF for the number of message arrivals at station 1 is given by $[A(z)]^N$). This is obtained by simply letting $N = 1$ and then replacing r, δ^2, λ and γ^2 by Nr, $N\delta^2$, $N\lambda$ and $N\gamma^2$, respectively, in (3.63b). We thus have

$$E[W_1] = \frac{\delta^2}{2r} + \frac{Nr}{2} + \frac{N\lambda b^{(2)}}{2(1-N\lambda b)} + \frac{(\gamma^2-\lambda)b}{2\lambda(1-N\lambda b)} - \frac{1}{2} \tag{3.64a}$$

(3.64a) can also be obtained by letting $I(z) = [R(z)]^N$ in (3.58a), adding $E[Y_i]$ in (3.8b), and using the conversion in (3.5b). We note that [Rubi83] obtains $E[W_1]$ in the case of Poisson distribution for the number of message arrivals ($\gamma^2=\lambda$). From (3.63b) and (3.64a), we have

$$E[W] - E[W_1] = \frac{N(N-1)\lambda rb}{2(1-N\lambda b)} \geq 0 \tag{3.64b}$$

This means that, for the same total traffic, the mean waiting time at station 1 in the case where all traffic is concentrated there is smaller than that in the case of balanced traffic. Note that we have the reverse relationship in the gated service system as we see in Section 5.1.

4 Exhaustive-Service, Continuous-Time System

We discuss the continuous-time version of the exhaustive service system. We consider the case of Poisson message arrival, and only in Section 4.4 we refer to the mean message waiting time in a Poisson bulk arrival system.

Our model is as follows.

(i) An entity for service is a variable-length message. The length of each message at station i is assumed to be distributed according to a general distribution function whose LST is denoted by $B_i^*(s)$. Let b_i and $b_i^{(2)}$ be the mean and the second moment, respectively, for the message length at station i:

$$b_i = -B_i^{*(1)}(0), \qquad B_i^{(2)} = B_i^{*(2)}(0) \tag{4.1}$$

(ii) The message arrival process at station i is assumed to be Poisson arrival with rate λ_i.

(iii) The reply interval between station i and station i+1 is assumed to be distributed according to a general distribution function whose LST is denoted by $R_i^*(s)$. Denote by r_i and δ_i^2 the mean and variance, respectively, for the reply interval between station i and station i+1:

$$r_i = -R_i^{*(1)}(0), \qquad \delta_i^2 = R_i^{*(2)}(0) - r_i^2. \tag{4.2}$$

In this chapter we consider the case of exhaustive service. Note that the gambler's ruin problem in the case of discrete-time system (in Section 3.2) corresponds to the busy period analysis in an M/G/1 queue with arrival rate λ_i and service time distribution $B_i^*(s)$. Define $\Theta_i^*(s)$ as the LST of the distribution function for the busy period in such an M/G/1 queue. As shown in [Klei75,Sec.5.8] and [Coop81,Sec.5.8], it satisfies

$$\Theta_i^{*}(s) = B_i^{*}[s + \lambda_i - \lambda_i\Theta_i^{*}(s)] \tag{4.3a}$$

from which we have the first and second moments, θ_i and $\theta_i^{(2)}$, for the busy period as

$$\theta_i = -\Theta_i^{*(1)}(0) = \frac{b_i}{1-\rho_i}, \quad \theta_i^{(2)} = \Theta_i^{*(2)}(0) = \frac{b_i^{(2)}}{(1-\rho_i)^3} \tag{4.3b}$$

where

$$\rho_i \triangleq \lambda_i b_i \tag{4.4}$$

Also, define $\Gamma_i(z)$ as the GF for the number of messages served in a busy period in the same M/G/1 queue. From [Klei75,Sec.5.9], we have

$$\Gamma_i(z) = zB_i^{*}[\lambda_i - \lambda_i\Gamma_i(z)] \tag{4.5a}$$

from which

$$E[\Gamma_i] = \frac{1}{1-\rho_i}, \quad \mathrm{Var}\,[\Gamma_i] = \frac{\rho_i(1-\rho_i) + \lambda_i^2 b_i^{(2)}}{(1-\rho_i)^3} \tag{4.5b}$$

4.1 Number of Messages at Polling Instants

Let

$L_i(t) \triangleq$ number of messages at station i at time t

and define the joint and marginal GF's for $[L_1(t),L_2(t),\ldots,L_N(t)]$ at time $t=\tau_i(m)$, i.e., at the time when station i is polled as in (3.21a) and (3.21b), respectively.

In order to relate $F_i(z_1,z_2,\ldots, z_N)$ to $F_{i+1}(z_1,z_2,\ldots, z_N)$, first note that the service time for station i, $\bar{\tau}_i(m) - \tau_i(m)$, is distributed as the sum of $L_i(\tau_i(m))$ busy periods where each busy period is distributed as given in (4.3a). Thus we have

$$E[e^{-s\{\bar{\tau}_i(m)-\tau_i(m)\}}] = [\Theta_i^*(s)]^{L_i(\tau_i(m))} \tag{4.6a}$$

Therefore the joint GF (except for station i) for the number of arrivals during the service time for station i is given by

$$\left\{\Theta_i^*\left[\sum_{\substack{j=1\\(j\neq i)}}^{N} (\lambda_j - \lambda_j z_j)\right]\right\}^{L_i(\tau_i(m))} \tag{4.6b}$$

Similarly, the joint GF for the number of message arrivals during the reply interval between station i and station i+1 is given by

$$R_i^*\left[\sum_{j=1}^{N} (\lambda_j - \lambda_j z_j)\right] \tag{4.7}$$

Hence, by an argument similar to Section 3.3, we get

$$F_{i+1}(z_1,z_2,\ldots,z_N)=R_i^*\left[\sum_{j=1}^{N}(\lambda_j-\lambda_j z_j)\right]$$

$$\cdot\ F_i(z_1,z_2,\ldots,z_{i-1},\Theta_i^*\left[\sum_{\substack{j=1\\(j\neq i)}}^{N}(\lambda_j-\lambda_j z_j)\right],\ z_{i+1},\ldots,z_N) \tag{4.8}$$

Now, define the moments for $[L_1(t),L_2(t),\ldots,L_N(t)]$ at $t=\tau_i(m)$ as in (3.27), (3.28a) and (3.28b). From (4.8), a set of N^2 equations for $\{f_i(j);\ i,j = 1,2,\ldots,N\}$ are given by

$$f_{i+1}(i) = r_i\lambda_i \tag{4.9a}$$

$$f_{i+1}(j) = r_i\lambda_j + f_i(j) + \frac{f_i(i)\lambda_j b_i}{1-\rho_i} \qquad j\neq i \tag{4.9b}$$

(Note that (4.9a) and (4.9b) have the same form as (3.29a) and (3.29b).) The solution to (4.9a) and (4.9b) is given by

$$E[L_i^*] = f_i(i) = \frac{\lambda_i(1-\rho_i)\sum_{k=1}^{N} r_k}{1-\sum_{k=1}^{N}\rho_k} \tag{4.10a}$$

$$f_i(j) = \lambda_j \left[\sum_{k=j}^{i-1} r_k + \frac{[\sum_{k=j+1}^{i-1}\rho_k][\sum_{k=1}^{N} r_k]}{1-\sum_{k=1}^{N}\rho_k} \right] \qquad j\neq i \tag{4.10b}$$

A set of N^3 equations for $\{f_i(j,k);\ i,j,k = 1,2,\ldots,N\}$ are given by

$$f_{i+1}(j,k) = \lambda_j\lambda_k(\delta_i^2+r_i^2) + r_i\lambda_k f_i(j) + r_i\lambda_j f_i(k) + f_i(i)\lambda_j\lambda_k\left[\frac{2\lambda_i b_i}{1-\rho_i} + \frac{b_i^{(2)}}{(1-\rho_i)^3}\right]$$

$$+ \frac{b_i}{1-\rho_i}[f_i(i,j)\lambda_k+f_i(i,k)\lambda_j]+f_i(j,k)+\frac{f_i(i,i)\lambda_j\lambda_k b_i^2}{(1-\rho_i)^2} \qquad i\neq j,\ j\neq k \tag{4.11a}$$

$$f_{i+1}(i,k) = \lambda_i\lambda_k(\delta_i^2+r_i^2) + r_i\lambda_i\left[f_i(k) + \frac{f_i(i)\lambda_k b_i}{1-\rho_i}\right] \qquad i\neq k \tag{4.11b}$$

$$f_{i+1}(i,i) = \lambda_i^2(\delta_i^2 + r_i^2) \tag{4.11c}$$

In the case of identical stations, we have

$$E[L^*] = \frac{Nr\lambda(1-\rho)}{1-N\rho} \tag{4.12a}$$

$$f_i(i,i) = \frac{\delta^2\lambda^2 N(1-\rho)}{1-N\rho} + \frac{N(N-1)\lambda^3 rb^{(2)}}{(1-N\rho)^2} + \frac{N^2r^2\lambda^2(1-\rho)^2}{(1-N\rho)^2} \tag{4.12b}$$

4.2 Service Time, Intervisit Time and Cycle Time

The LST of the distribution function for the service time at station i,

S_i, is given by (4.6a); i.e.,

$$S_i^*(s) = F_i[\Theta_i^*(s)] \tag{4.13}$$

from which we have

$$E[S_i] = E[L_i^*]\theta_i = \frac{\rho_i \sum_{k=1}^{N} r_k}{1-\sum_{k=1}^{N} \rho_k} \tag{4.14a}$$

$$\mathrm{Var}[S_i] = \theta_i^2\,\mathrm{Var}[L_i^*] + (\theta_i^{(2)} - \theta_i^2)\,E[L_i^*]$$

$$= \frac{1}{(1-\rho_i)^2}\left[b_i^2\,\mathrm{Var}[L_i^*] + \frac{\lambda_i[b_i^{(2)}-b_i^2(1-\rho_i)]\sum_{k=1}^{N} r_k}{1-\sum_{k=1}^{N}\rho_k} \right] \tag{4.14b}$$

where we have used (4.3b) and (4.10a). In the case of identical stations, by using (4.12b) we have

$$E[S] = \frac{Nr\rho}{1-N\rho} \tag{4.15a}$$

$$\mathrm{Var}[S] = \frac{r^2N\rho^2}{(1-\rho)(1-N\rho)} + \frac{N\lambda rb^{(2)}(1+\rho-N\rho)}{(1-\rho)(1-N\rho)^2} \tag{4.15b}$$

Note that (4.15a) and (4.15b) correspond to (3.34a) and (3.34b), respectively, in the discrete-time system.

We next give the GF for the number of messages served in a service time for station i, denoted by $T_i(z)$. Since these messages are those served in a busy period initiated by $L_i(\Pi_i(m))$ messages, it follows that

$$T_i(z) = F_i[\Pi_i(z)] \tag{4.16}$$

where $\Gamma_i(z)$ is given by (4.5a). Using (4.5b) and (4.10a), we have

$$E[T_i(m)] = E[\Gamma_i]E[L_i^*] = \frac{\lambda_i \sum_{k=1}^{N} r_k}{1-\sum_{k=1}^{N} \rho_k} \tag{4.17a}$$

$$Var[T_i(m)] = Var[L_i^*](E[\Gamma_i])^2 + E[L_i^*]\ Var\ [\Gamma_i]$$

$$= \frac{1}{(1-\rho_i)^2}\left[Var[L_i^*] + \frac{\lambda_i[\rho_i(1-\rho_i)+\lambda_i^2 b_i^{(2)}] \sum_{k=1}^{N} r_k}{1-\sum_{k=1}^{N}\rho_k}\right] \tag{4.17b}$$

In the case of identical stations, by using (4.12b), we get

$$E[T(m)] = \frac{N\lambda r}{1-N\rho} \tag{4.18a}$$

$$Var[T(m)] = \frac{\delta^2 N\lambda^2}{(1-\rho)(1-N\rho)} + \frac{N^2 r\lambda^3 b^{(2)}}{(1-\rho)(1-N\rho)^2} + \frac{N\lambda r(1+\rho)}{(1-\rho)(1-N\rho)} \tag{4.18b}$$

We proceed to consider $I_i^*(s)$, the LST for distribution function of the intervisit time for station i. Since the number of message arrivals during the intervisit time is the number of messages found at the polling instant, we have the relation

$$I_i^*(\lambda_i - \lambda_i z) = F_i(z) \tag{4.19}$$

or, by $s=\lambda_i - \lambda_i z$,

$$I_i^*(s) = F_i(1 - s/\lambda_i) \tag{4.20a}$$

from which we have

$$E[I_i] = \frac{f_i(i)}{\lambda_i} = \frac{(1-\rho_i)\sum_{k=1}^{N} r_k}{1-\sum_{k=1}^{N}\rho_k}, \quad E[(I_i)^2] = \frac{f_i(i,i)}{\lambda_i^{\,2}} \tag{4.20b}$$

In the case of identical stations, by using (4.12b), we have

$$E[I] = \frac{Nr(1-\rho)}{1-N\rho}, \quad Var[I] = \frac{\delta^2 N(1-\rho)}{1-N\rho} + \frac{N(N-1)\lambda r b^{(2)}}{(1-N\rho)^2} \tag{4.21}$$

which correspond to (3.37b) in the discrete-time system.

The LST of the distribution function for the cycle time for station i, C_i, may be obtained in the same way as that has led to (3.39b). If we denote by $I_i(t)$ the distribution function for I_i, then

$$C_i^*(s) \triangleq E[e^{-s\{\bar{\tau}_i(m+1) - \bar{\tau}_i(m)\}}]$$

$$= E[e^{-s\{\bar{\tau}_i(m+1)-\tau_i(m+1)\}} e^{-s\{\tau_i(m+1)-\bar{\tau}_i(m)\}}]$$

$$= \int_0^{\infty} E[e^{-s\{\bar{\tau}_i(m+1)-\tau_i(m+1)\}} e^{-st} \mid \tau_i(m+1) - \bar{\tau}_i(m)=t]\, dI(t)$$

However, by applying (4.13) and (4.19), we have

$$E[e^{-s\{\bar{\tau}_i(m+1)-\tau_i(m+1)\}} \mid \tau_i(m+1) - \bar{\tau}_i(m)=t]$$

$$= E[\cdot_i^*(s)^{\cdot_i(m+1)} \mid \tau_i(m+1)-\bar{\tau}_i(m)=t]$$

$$= E[e^{-\{\cdot_i - \cdot_i \cdot_i^*(s)\}\{\tau_i(m+1)-\bar{\tau}_i(m)\}} \mid \tau_i(m+1)-\bar{\tau}_i(m)=t]$$

$$= E[e^{-\{\cdot_i - \cdot_i \cdot_i^*(s)\}t}]$$

It follows that

$$C_i^*(s) = I_i^*[s - \cdot_i - \cdot_i \cdot_i^*(s)] \tag{4.22}$$

which corresponds to (3.39b) in the discrete time system. From (4.22), we have

$$E[C_i] = (1+\lambda_i\theta_i)E[I_i] = \frac{\sum_{k=1}^{N} r_k}{1-\sum_{k=1}^{N}\rho_k} \tag{4.23a}$$

$$\mathrm{Var}[C_i] = (1+\lambda_i\theta_i)^2\,\mathrm{Var}[I_i] + \lambda_i\theta_i^{(2)}\,E[I_i]$$

$$= \frac{1}{\lambda_i^2(1-\rho_i)^2}\left[\mathrm{Var}[L_i^*] + \frac{[-(1-\rho_i)+\lambda_i^2 b_i^{(2)}]\sum_{k=1}^{N} r_k}{1-\sum_{k=1}^{N}\rho_k}\right] \tag{4.23b}$$

where we have used (4.3b) and (4.20b). In the case of identical stations, by using (4.12b), we obtain

$$E[C] = \frac{Nr}{1-N\rho},\quad \mathrm{Var}[C] = \frac{\delta^2 N}{(1-\rho)(1-N\rho)} + \frac{N^2\lambda r b^{(2)}}{(1-\rho)(1-N\rho)^2} \tag{4.24}$$

which corresponds to (3.41) in the discrete time system. From (4.23a), the stability condition is given by

$$\sum_{k=1}^{N}\rho_k < 1 \tag{4.25}$$

4.3 Number of Messages at Departure Instants

Let us define a sequence of times $\{\tau_i^{(n)}(m); n=1,2,\ldots,T_i(m)\}$, where $\tau_i^{(n)}(m)$ is the service completion instant for the n th message served in the m th cycle at station i. $T_i(m)$ is the number of messages served in the m th cycle at station i. We now consider the GF, $Q_i(z)$, for the number of messages just after the moment of message service completion at station i, which is denoted by L_i. It is given by the average of $z^{L_i(\tau_i^{(n)}(m))}$ over the average number of messages served in the m th cycle, where the averages are taken over the regenerative cycle $\{C_i^{(m)}, m=1,2,\ldots,M_i\}$. The regeneration points are those when station i is polled and all $L_j(\tau_i(m)), 1 \le j \le N$ are zero (see Section 3.4.1 for a similar discussion in the discrete-time system). Thus we have

$$Q_i(z) \triangleq E[z^{L_i}] = \frac{E\left[\sum_{m=1}^{M_i} \sum_{n=1}^{T_i(m)} z^{L_i(\tau_i^{(n)}(m))}\right]}{E\left[\sum_{m=1}^{M_i} T_i(m)\right]} \tag{4.26}$$

However, since M_i is a stopping time for the regenerative process with $E[M_i] < \infty$, it follows from Wald's lemma that

$$E\left[\sum_{m=1}^{M_i} \sum_{n=1}^{T_i(m)} z^{L_i(\tau_i^{(n)}(m))}\right] = E[M_i]\, E\left[\sum_{n=1}^{T_i(m)} z^{L_i(\tau_i^{(n)}(m))}\right] \tag{4.27a}$$

$$E\left[\sum_{m=1}^{M_i} T_i(m)\right] = [M_i]\, E[T_i(m)] \tag{4.27b}$$

Thus we get

$$Q_i(z) \triangleq E[z^{L_i}] = \frac{E\left[\sum_{n=1}^{T_i(m)} z^{L_i(\tau_i^{(n)}(m))}\right]}{E[T_i(m)]} \tag{4.28}$$

The denominator in (4.28) is given in (4.17a).

To evaluate the sum in the numerator of (4.28), we apply the formula (3.20b) with substitution

$$T \to T_i(m), \quad L_n \to L_i(\tau_i^{(n)}(m)), \quad P(z) \to B_i^*(\lambda_i - \lambda_i z)$$

$$F(z) = E[z^{L_0}] \to F_i(z) = E[z^{L_i(\tau_i(m))}]$$

Then we get

$$E\left[\sum_{n=1}^{T_i(m)} z^{L_i(\tau_i^{(n)}(m))}\right] = \frac{B_i^*(\lambda_i-\lambda_i z)}{z-B_i^*(\lambda_i-\lambda_i z)}\,[F_i(z)-1]$$

Substituting this expression and (4.17a) into (4.28), we have

$$Q_i(z) = \frac{1-\sum_{k=1}^{N}\rho_k}{\lambda_i\sum_{k=1}^{N} r_k}\cdot\frac{B_i^*(\lambda_i-\lambda_i z)}{z-B_i^*(\lambda_i-\lambda_i z)}\,[F_i(z)-1] \tag{4.29}$$

From (4.29) we easily get the average number of messages at station i left behind the service completion at station i:

$$E[L_i] = \rho_i + \frac{\lambda_i^2 b_i^{(2)}}{2(1-\rho_i)} + \frac{[1-\sum_{k=1}^{N}\rho_k]f_i(i,i)}{2\lambda_i(1-\rho_i)\sum_{k=1}^{N} r_k} \tag{4.30a}$$

In the case of identical stations, we have

$$E[L] = \rho + \frac{\lambda(\delta^2+r^2)}{2r} + \frac{(N-1)\lambda r}{2(1-N\rho)} + \frac{N\lambda^2 b^{(2)}}{2(1-N\rho)} \tag{4.30b}$$

Although (4.29), (4.30a) and (4.30b) have been obtained for the number of messages left by departing messages, we here claim that they hold for the number of messages at an arbitrary instant. This claim comes from the assumption of Poisson arrival and the fact that the state changes by unit step values (one by one) only. In such a case, as shown in [Klei75,Sec.5.3 and

Prob.5.6], the arrival-time, arbitrary-time and departure-time probabilities have the same limiting distribution.

4.4 Waiting Time

Finally we may derive $W_i^*(s)$, the LST of the distribution function for the waiting time W_i (excluding the service time) of each message at station i. Note that the number of message arrivals at station i during the period in which a message stays at station i (the LST of the distribution function for this period is given by $W_i^*(s)B_i^*(s)$) is equal to the number of messages left behind at the service completion time for that message. Hence, we have

$$W_i^*(\lambda_i - \lambda_i z)B_i^*(\lambda_i - \lambda_i z) = Q_i(z) \tag{4.31}$$

Letting $s=\lambda_i - \lambda_i z$, we get

$$W_i^*(s) = \frac{Q_i(1-s/\lambda_i)}{B_i^*(s)} = \frac{1-\sum_{k=1}^{N}\rho_k}{\sum_{k=1}^{N} r_k} \cdot \frac{1-F_i(1-s/\lambda_i)}{s-\lambda_i+\lambda_i B_i^*(s)} \tag{4.32}$$

From (4.32) we have

$$E[W_i] = \frac{\lambda_i b_i^{(2)}}{2(1-\rho_i)} + \frac{[1-\sum_{k=1}^{N}\rho_k]f_i(i,i)}{2\lambda_i^2(1-\rho_i)\sum_{k=1}^{N} r_k} \tag{4.33a}$$

Comparing (4.30a) and (4.33a), we may confirm Little's relation

$$E[L_i] = \lambda_i E[W_i] + \rho_i \tag{4.34}$$

We note that Little's relation applies to the mean number of messages at an arbitrary time.

In the case of identical stations, we have

$$E[W] = \frac{\delta^2}{2r} + \frac{Nr(1-\rho)}{2(1-N\rho)} + \frac{N\lambda b^{(2)}}{2(1-N\rho)} \tag{4.33b}$$

Note that this has the same form as (3.63b) when Poisson arrival is assumed ($\gamma^2 = \lambda$) and "a half slot" is made shrink to zero.

It may also be interesting to relate $W_i^*(s)$ to $I_i^*(s)$. Using (4.20a) and (4.20b) in (4.32), we have

$$W_i^*(s) = \frac{1-I_i^*(s)}{E[I_i]} \cdot \frac{1-\rho_i}{s-\lambda_i+\lambda_i B_i^*(s)} \tag{4.35a}$$

from which we get

$$E[W_i] = \frac{E[\{I_i\}^2]}{2E[I_i]} + \frac{\lambda_i b_i^{(2)}}{2(1-\rho_i)} \tag{4.35b}$$

Note the analogy of this expression to that for the mean waiting time in the M/G/1 queue with the server going on vacation [Coop81,Prob.5.12(d)]. [Bux83a] attempts to approximate $E[\{I_i\}^2]$ to get around the numerical complexity in solving N^3 linear equations.

In [Eise72], the explicit expression for $E[\{I_i\}^2]$ in the case of $N=2$ nonidentical stations is given. In our notation, it is

$$E[\{I_1\}^2] = \delta_1^2 + r_1^2 + \frac{2\rho_1\rho_2[\rho_2 r_1+(1-\rho_1)r_2]^2+(r_1+r_2)[\lambda_1\rho_2^2 b_1^{(2)}+\lambda_2(1-\rho_1)^2 b_2^{(2)}]}{(1-\rho_1-\rho_2)^2(1-\rho_1-\rho_2+2\rho_1\rho_2)}$$

$$+ \frac{2r_1 r_2(1-\rho_1)\rho_2+\rho_2^2(\delta_1^2+r_1^2)+(1-\rho_1)^2(\delta_2^2+r_2^2)}{(1-\rho_1-\rho_2)(1-\rho_1-\rho_2+2\rho_1\rho_2)}$$

$$+ \frac{2r_1[\rho_2 r_1+(1-\rho_1)r_2]}{1-\rho_1-\rho_2} \tag{4.36a}$$

Let us use (4.36a) to have the mean waiting time in the case of zero reply intervals. For this purpose, we let $\delta_1^2 = \delta_2^2 = 0$ and $r_1 = r_2 = r$ and then take the limit of $r \to 0$ in the ratio $E[\{I_i\}^2]/E[I_i]$. Thus we easily have

$$E[W_1] = \frac{\lambda_1\rho_2^2 b_1^{(2)} + \lambda_2(1-\rho_1)^2 b_2^{(2)}}{2(1-\rho_1)(1-\rho_1-\rho_2)(1-\rho_1-\rho_2+2\rho_1\rho_2)} + \frac{\lambda_1 b_1^{(2)}}{2(1-\rho_1)} \tag{4.36b}$$

which is in agreement with old results in [Avi65], [Conw67,Sec.9-2] and [Takac68]. This is also given in [Coop70] as a special case of N=2. From (4.36b) and an expression for $E[W_2]$ (obtained by exchanging indices 1 and 2 in (4.36b)), we can confirm a relation

$$\sum_{i=1}^{2} \rho_i E[W_i] = \frac{\rho_0 W_0}{1-\rho_0} \tag{4.37a}$$

where

$$\rho_0 = \sum_{i=1}^{2} \rho_i, \quad W_0 = \sum_{i=1}^{2} \frac{1}{2}\lambda_i b_i^{(2)} \tag{4.37b}$$

Note that (4.37a) and (4.37b) provide an instance of M/G/1 conservation law for non-preemptive work-conserving service systems [Klei76,Sec.3.4].

We may play with (4.33b) as in Section 3.6.1. First, the mean waiting time at station 1, $E[W_1]$, when all traffic is concentrated on station 1 is obtained by letting N=1 and then replacing r, δ^2 and λ by Nr, $N\delta^2$ and $N\lambda$, respectively, in (4.33b). Thus we have (not using ρ)

$$E[W_1] = \frac{\delta^2}{2r} + \frac{Nr}{2} + \frac{N\lambda b^{(2)}}{2(1-N\lambda b)} \tag{4.38}$$

From (4.33b) and (4.38), we get the relationship in the same form as (3.64b).

Next, let us derive the mean message waiting time in a Poisson bulk arrival system where the bulk arrival rate is λ and GF for the number of messages in a bulk is given by $G(z)$ (we assume the case of identical stations). If $B^*(s)$ is the LST of the distribution function for each message length, we have equivalently a system of Poisson supermessage arrivals of rate λ and supermessage length distribution given by $G[B^*(s)]$ (see [Klei75,Prob.5.12]). Note that the mean and second moment of the supermessage length are given by gb and $(g^{(2)}-g)b^2+gb^{(2)}$, respectively, where g and $g^{(2)}$ are the mean and second moment of the bulk size, respectively. Thus, by the analogy with (4.33b), we have the mean supermessage waiting time

$$E[W_s] = \frac{\delta^2}{2r} + \frac{Nr(1-\lambda gb)}{2(1-N\lambda gb)} + \frac{N\lambda[(g^{(2)}-g)b^2+gb^{(2)}]}{2(1-N\lambda gb)} \tag{4.39a}$$

However, according to (3.10c), the average length before an arbitrary message within a supermessage is given by

$$E[D_s] = \frac{(g^{(2)}-g)b}{2g} \tag{4.39b}$$

Thus the mean message waiting time in a Poisson bulk arrival system is given by

$$\begin{aligned} E[W] &= E[W_s] + E[D_s] \\ &= \frac{\delta^2}{2r} + \frac{Nr(1-\lambda gb)}{2(1-N\lambda gb)} + \frac{N\lambda gb^{(2)}}{2(1-N\lambda gb)} + \frac{(g^{(2)}-g)b}{2g(1-N\lambda gb)} \end{aligned} \tag{4.40}$$

4.5 Correlation of Station Times

Recently, [Ferg84] (based on [Amin75] and [Humb78,Chap.7]) has shown that $E[(I_i)^2]$ can be computed by solving N^2 linear equations. To present this result, let us introduce the <u>station time</u> (called <u>terminal service time</u> in [Ferg84]) $\tilde{x}_k$ at the k th visit (by the server) to a station, defined as the reply interval from the last station plus the service time for the current station. Thus, if the k th visit is to station i, then we have

$$\tilde{x}_k = \bar{\tau}_i(m) - \bar{\tau}_{i-1}(m) \tag{4.41}$$

See Figure 5. Note that the visit number k is sequentially incremented at every visit to all stations. The intervisit time before the k th visit, I_k, may be expressed as

$$I_k = \sum_{j=1}^{N-1} \tilde{x}_{k-j} + \tilde{r}_{k-1} \tag{4.42}$$

where $\tilde{r}_{k-1}$ is the reply interval before the k th visit.

The N successive station times are expressed in a vector $\underline{x}_k = [\tilde{x}_{k-(N-1)}, \ldots, \tilde{x}_{k-1}, \tilde{x}_k]$ whose probability density function $P_k(x_{k-(N-1)}, \ldots, x_{k-1}, x_k)$ and its Laplace transform $X_k^*(s_1, s_2, \ldots, s_N)$ are defined by

$$X_k^*(s_1, s_2, \ldots, s_N) \triangleq E[\exp(-\sum_{j=1}^{N} s_j \tilde{x}_{k-N+j})]$$

$$= \int_0^\infty \int_0^\infty \cdots \int_0^\infty \exp(-\sum_{j=1}^{N} s_j\, x_{k-N+j})\, P_k(x_{k-(N-1)}, \ldots, x_k) \prod_{j=1}^{N} dx_{k-N+j} \tag{4.43}$$

Let us denote by $\theta_k^{*(m)}(t)$ the m fold convolution of $\theta_k(t)$, the probability density function for the busy period at the k th visited station. Considering the events happening during I_k, we can express $P_k(\underline{x}_k)$ in terms of $P_{k-1}(\underline{x}_{k-1})$:

$$P_k(x_{k-(N-1)}, \ldots, x_{k-1}, x_k) = \int_0^\infty P_{k-1}(x_{k-N}, x_{k-N+1}, \ldots, x_{k-1})\, dx_{k-N}$$

$$\cdot \int_0^{x_k} dR_{k-1}(t) \sum_{m=0}^{\infty} \theta_k^{*(m)}(x_k - t)\, \frac{1}{m!}[\lambda_k(t + \sum_{j=1}^{N-1} x_{k-j})]^m \exp[-\lambda_k(t + \sum_{j=1}^{N-1} x_{k-j})] \tag{4.44}$$

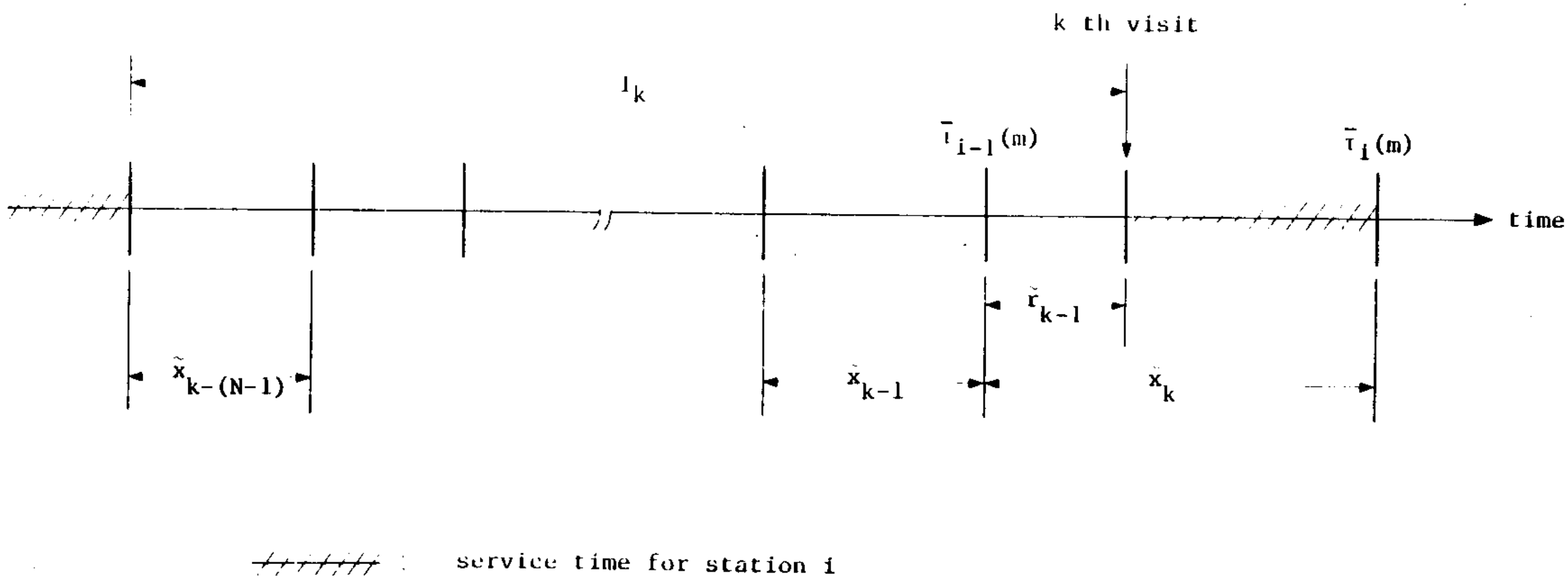

Figure 5. Intervisit time.

where $R_{k-1}(t)$ is the distribution function of $\tilde{r}_{k-1}$ and λ_k is the Poisson arrival rate at the k th visited station. Note that the first integral represents the fact that $\tilde{x}_{k-N}$ can take any value. The factor $\frac{1}{m!}[\lambda_k(t+\sum_{j=1}^{N-1} x_{k-j})]^m \exp[-\lambda_k(t+\sum_{j=1}^{N-1} x_{k-j})]$ is the probability of m arrivals during the intervisit time of duration $t+\sum_{j=1}^{N-1} x_{k-j}$. $\theta_k^{*(m)}(x_k - t)$ is the probability density that the busy period initiated by the m messages has a length of $x_k - t$.

We now take the Laplace transform of (4.44) by multiplying $\exp(-\sum_{j=1}^{N} s_j x_{k-N+j})$ on both sides and then integrating them over $[x_{k-N+1}, \ldots, x_{k-1}, x_k]$. By definition, the left-hand side yields $X_k^*(s_1, s_2, \ldots, s_N)$. As for the right-hand side, let us first multiply $\exp(-s_N x_k)$ and integrate over x_k to have (let $f(t)$ be a function of t)

$$\int_0^\infty \exp(-s_N x_k)dx_k \int_0^{x_k} f(t)\theta_k^{*(m)}(x_k-t)dt$$

$$= \int_0^\infty f(t)\exp(-s_N t)dt \int_t^\infty \theta_k^{*(m)}(x_k-t)\exp[-s_N(x_k-t)]dx_k$$

$$= \int_0^\infty f(t)\exp(-s_N t)dt.[\Theta_k^*(s_N)]^m \qquad (4.45a)$$

where $\Theta_k^*(s)$ is the Laplace transform of $\theta_k(t)$. Next, replacing f(t) by an appropriate function, take the summation over m to have

$$\sum_{m=0}^{\infty} \int_0^{\infty} dR_{k-1}(t)\frac{1}{m!}[\lambda_k \ (t+\sum_{j=1}^{N-1} x_{k-j})]^m \exp[-\lambda_k(t+\sum_{j=1}^{N-1} x_{k-j})] \exp(-s_N t)[\Theta_k^*(s_N)]^m$$

$$= \int_0^{\infty} \exp\{-s_N t-\lambda_k[1-\Theta_k^*(s_N)](t+\sum_{j=1}^{N-1} x_{k-j})\}\, dR_{k-1}(t)$$

$$= R_{k-1}^*(s_N+\lambda_k[1-\Theta_k^*(s_N)]).\exp\{-\lambda_k[1-\Theta_k^*(s_N)]\sum_{j=1}^{N-1} x_{k-j}\} \quad (4.45b)$$

Finally, the multiplication of $\exp(-\sum_{j=1}^{N-1} s_j x_{k-N+j}) = \exp(-\sum_{j=1}^{N-1} s_{N-j}\, x_{k-j})$

and integration gives

$$\int_0^{\infty} \cdots \int_0^{\infty} \int_0^{\infty} P_{k-1}(x_{k-N},\ldots, x_{k-2}, x_{k-1})\exp\left[-\sum_{j=1}^{N-1} x_{k-j}\{s_{N-j}+\lambda_k[1-\Theta_k^*(s_N)]\}\right] \prod_{j=1}^{N} dx_{k-j}$$

$$= X_{k-1}^*(0, s_1+\lambda_k[1-\Theta_k^*(s_N)], s_2+\lambda_k[1-\Theta_k^*(s_N)],\ldots, s_{N-1}+\lambda_k[1-\Theta_k^*(s_N)])$$

(4.45c)

Thus we get the Laplace transform of (4.44) as

$$X_k^*(s_1, s_2, \ldots, s_N) = R_{k-1}^*(s_N+\lambda_k [1-\Theta_k^* (s_N)]).$$

$$. X_{k-1}^*(0, s_1+\lambda_k[1-\Theta_k^*(s_N)], s_2+\lambda_k[1-\Theta_k^*(s_N)],\ldots,s_{N-1}+\lambda_k[1-\Theta_k^*(s_N)])$$

(4.46)

From (4.46), we first calculate the mean station time. Differentiating (4.46) with respect to s_N and then setting $s_1 = \ldots = s_N = 0$, we have

$$E[\tilde{x}_k] = (1+\lambda_k\, \vartheta_k)\, r_{k-1} + \lambda_k\, \vartheta_k \sum_{j=1}^{N-1} E[\tilde{x}_{k-j}] \quad (4.47a)$$

where $r_{k-1} = E[\tilde{r}_{k-1}]$, and $\theta_k = -\Theta_k^{*(1)}(0)$ is the mean busy period at the k th visited station. From (4.3b), we already have $\theta_k = b_k/(1-\rho_k)$, where $\rho_k = \lambda_k b_k$. It follows from (4.47a) that

$$E[\tilde{x}_k] = r_{k-1} + \rho_k \cdot \sum_{j=0}^{N-1} E[\tilde{x}_{k-j}] \tag{4.47b}$$

However, clearly the summation on the right-hand side is the mean cycle time for the k th visited station. Thus the mean station time for station i is given by

$$E[\tilde{x}_i] = r_{i-1} + \rho_i \, E[C_i] \tag{4.48a}$$

where $E[C_i]$ is given by (4.23a). From (4.42) we also have

$$E[I_i] = E[C_i] - E[\tilde{x}_i] + r_{i-1} \tag{4.48b}$$

Using (4.48a) and (4.23a), we get $E[I_i]$ in (4.20b).

We next calculate the covariance of $\tilde{x}_k$ and $\tilde{x}_{k-j}$, $E[(\tilde{x}_k-\bar{x}_k)(\tilde{x}_{k-j}-\bar{x}_{k-j})]$, where $\bar{x}_k \triangleq E[\tilde{x}_k]$. In Appendix C, we derive from (4.46) that

$$E[(\tilde{x}_k-\bar{x}_k)^2] = \frac{1+\rho_k}{1-\rho_k} \delta_{k-1}^2 + \frac{\lambda_k b_k^{(2)}}{(1-\rho_k)^3} E[I_k] + \frac{\rho_k^2}{(1-\rho_k)^2} \mathrm{Var}[I_k] \tag{4.49a}$$

$$E[(\tilde{x}_k-\bar{x}_k)(\tilde{x}_{k-j}-\bar{x}_{k-j})] = \frac{\rho_k}{1-\rho_k} \sum_{m=1}^{N-1} E[(\tilde{x}_{k-m}-\bar{x}_{k-m})(\tilde{x}_{k-j}-\bar{x}_{k-j})] \tag{4.49b}$$

where (from (4.42))

$$\mathrm{Var}[I_k] = \delta_{k-1}^2 + \sum_{n=1}^{N-1} \sum_{m=1}^{N-1} E[(\tilde{x}_{k-n}-\bar{x}_{k-n})(\tilde{x}_{k-m}-\bar{x}_{k-m})] \tag{4.49c}$$

We now write (4.49a-c) in terms of the covariances of the station times for stations i and j. Since they depend on which station is visited first, we define r_{ij} as the covariance of $\tilde{x}_i$ and $\tilde{x}_j$ when station j is visited before station i. Thus

$$r_{ij} \triangleq \begin{cases} E[(\tilde{x}_i-\bar{x}_i)(\tilde{x}_j-\bar{x}_j)] & j < i \\ E[(\tilde{x}_{i+N}-\bar{x}_{i+N})(\tilde{x}_j-\bar{x}_j)] & j > i \end{cases} \tag{4.50}$$

In (4.49b), the covariance of $\tilde{x}_i$ and $\tilde{x}_j$ is expressed in terms of the covariances of each of $\tilde{x}_{i-(N-1)}, \tilde{x}_{i-(N-2)}, \ldots, \tilde{x}_{i-1}$ and $\tilde{x}_j$. Suppose that the k th visit is to station i. Then (4.49a-c) are converted to

$$r_{ii} = \frac{1+\rho_i}{1-\rho_i}\,\delta_{i-1}^{\;2} + \frac{\lambda_i b_i^{(2)}}{(1-\rho_i)^3}\,E[I_i] + \frac{\rho_i^{\;2}}{(1-\rho_i)^2}\,\mathrm{Var}[I_i] \tag{4.51a}$$

$$r_{ij} = \frac{\rho_i}{1-\rho_i}\left(\sum_{m=i+1}^{N} r_{jm} + \sum_{m=1}^{j-1} r_{jm} + \sum_{m=j}^{i-1} r_{mj}\right) \qquad j < i \tag{4.51b}$$

$$r_{ij} = \frac{\rho_i}{1-\rho_i}\left(\sum_{m=i+1}^{j-1} r_{jm} + \sum_{m=j}^{N} r_{mj} + \sum_{m=1}^{i-1} r_{mj}\right) \qquad j > i \tag{4.51c}$$

where

$$\mathrm{Var}[I_i] = \delta_{i-1}^{\;2} + \sum_{n=1}^{i-1}\left(\sum_{m=i+1}^{N} r_{nm} + \sum_{m=1}^{n-1} r_{nm} + \sum_{m=n}^{i-1} r_{mn}\right)$$

$$+ \sum_{n=i+1}^{N}\left(\sum_{m=i+1}^{n-1} r_{nm} + \sum_{m=n}^{N} r_{mn} + \sum_{m=1}^{i-1} r_{mn}\right) \tag{4.51d}$$

However, using (4.51b) and (4.51c), we may write (4.51d) in a simpler form

$$\mathrm{Var}[I_i] = \delta_{i-1}^{\;2} + \frac{1-\rho_i}{\rho_i}\sum_{\substack{j=1\\(j\neq i)}}^{N} r_{ij} \tag{4.52a}$$

Substituting (4.52a) into (4.51a), we have

$$r_{ii} = \frac{1}{(1-\rho_i)^2}\,\delta_{i-1}^{\;2} + \frac{\lambda_i b_i^{(2)}}{(1-\rho_i)^3}\,E[I_i] + \frac{\rho_i}{1-\rho_i}\sum_{\substack{j=1\\(j\neq i)}}^{N} r_{ij} \tag{4.52b}$$

Note that (4.51b), (4.51c) and (4.52b) constitute a system of N^2 linear equations for $\{r_{ij};\ i,j=1,\ 2,\ \ldots,\ N\}$. Once they have been solved, we may use them in (4.52a) to compute $Var[I_i]$, which is then used in (4.35b) to evaluate $E[W_i]$.

In the case of identical stations [Humb78,Sec.7.D] we have

$$r_{ij} = \frac{\rho[\delta^2(1-N\rho)+N\lambda rb^{(2)}]}{(1-N\rho)^2(1-\rho)} \qquad i \neq j \qquad (4.53a)$$

$$r_{ii} = \frac{(1+\rho-N\rho)[\delta^2(1-N\rho)+N\lambda rb^{(2)}]}{(1-N\rho)^2(1-\rho)} \qquad (4.53b)$$

which leads to (4.21).

5 Gated Service Systems

5.1 Discrete-Time System

We employ the same assumptions and parameters for the packet arrival process and reply interval as in Chapter 3 (they are given by (3.1) through (3.10c)). In the case of gated service, there is no need for the gambler's ruin problem. Thus, we start with the distribution for $[L_1(t), L_2(t), \ldots, L_N(t)]$ at time $t=\tau_i(m)$, i.e. the polling instant for station i, where $L_j(t)$ is the number of packets at station j at time t. Define the joint and marginal GF's as in (3.21a) and (3.21b), respectively.

For the gated service system, the service time duration for station i, $\bar{\tau}_i(m) - \tau_i(m)$, is simply equal to $L_i(\tau_i(m))$, the number of packets found at the polling instant:

$$\bar{\tau}_i(m) - \tau_i(m) = L_i(\tau_i(m)) \tag{5.1}$$

Thus the joint GF for the number of packets arriving during the service time is given by

$$E[\prod_{j=1}^{N} P_j(z_j)^{\bar{\tau}_i(m)-\tau_i(m)}] = E[\prod_{j=1}^{N} P_j(z_j)^{L_i(\tau_i(m))}]$$

Therefore, (instead of (3.26)), we get

$$F_{i+1}(z_1, z_2, \ldots, z_N) = R_i(\prod_{j=1}^{N} P_j(z_j)) \cdot$$
$$\cdot F_i(z_1, z_2, \ldots, z_{i-1}, \prod_{j=1}^{N} P_j(z_j), z_{i+1}, \ldots, z_N) \tag{5.2}$$

If we denote by L_i^* the number of packets at station i when it is polled, we can define the moments for L_i^* as in (3.27), (3.28a) and (3.28b).

A set of N^2 equations for $\{f_i(j);\ i,\ j = 1, 2, \ldots, N\}$ are now given by

$$f_{i+1}(i) = r_i\mu_i + \mu_i f_i(i) \tag{5.3a}$$

$$f_{i+1}(j) = r_i\mu_j + \mu_j f_i(i) + f_i(j) \qquad j \neq i \tag{5.3b}$$

The solution to the equations (5.3a) and (5.3b) is

$$E[L_i^*] = f_i(i) = \frac{\mu_i \sum_{k=1}^{N} r_k}{1 - \sum_{k=1}^{N} \mu_k} \tag{5.4a}$$

$$f_i(j) = \mu_j \left\{ \sum_{k=j}^{i-1} r_k + \frac{\left(\sum_{k=j}^{i-1}\mu_k\right)\left(\sum_{k=1}^{N} r_k\right)}{1 - \sum_{k=1}^{N}\mu_k} \right\} \qquad j \neq i \tag{5.4b}$$

A set of N^3 equations for $\{f_i(j,k);\ i,\ j,\ k = 1, 2, \ldots, N\}$ are given by

$$f_{i+1}(j,k)=\mu_j\mu_k(\sigma_i^2+r_i^2)+r_i\mu_k f_i(j)+r_i\mu_j f_i(k)+f_i(i)\mu_j\mu_k(2r_i+1)+f_i(j,k)$$
$$+\mu_j f_i(i,k)+\mu_k f_i(i,j)+\mu_j\mu_k f_i(i,i) \qquad i\neq j,\ i\neq k,\ j\neq k$$

(5.5a)

$$f_{i+1}(j,j)=\mu_j^2(\sigma_i^2+r_i^2)+r_i(\sigma_j^2-\mu_j)+2r_i\mu_j f_i(j)+f_i(i)[\sigma_j^2-\mu_j+\mu_j^2(2r_i+1)]$$
$$+f_i(j,j)+2\mu_j f_i(i,j)+\mu_j^2 f_i(i,i) \qquad i\neq j \tag{5.5b}$$

$$f_{i+1}(i,k)=\mu_i\mu_k(\sigma_i^2+r_i^2)+r_i\mu_i f_i(k)+f_i(i)\mu_i\mu_k(2r_i+1)+\mu_i f_i(i,k)$$
$$+\mu_i\mu_k f_i(i,i) \qquad i\neq k \tag{5.5c}$$

$$f_{i+1}(i,i)=\mu_i^2(\sigma_i^2+r_i^2)+r_i(\sigma_i^2-\mu_i)+f_i(i)[\sigma_i^2-\mu_i+\mu_i^2(2r_i+1)]+\mu_i^2 f_i(i,i)$$

(5.5d)

In the case of identical stations, we have

$$E[L^*] = \frac{Nr\mu}{1-N\mu} \qquad Var[L^*] = \frac{\delta^2\mu^2 N}{(1+\mu)(1-N\mu)} + \frac{\sigma^2 rN[1-(N-1)\mu]}{(1+\mu)(1-N\mu)^2} \tag{5.6}$$

We next consider the cycle time $\tau_i(m+1)-\tau_i(m)$ for station i (its GF is denoted by $C_i(z)$). Since the number of packets found at station i at time $t=\tau_i(m+1)$ is equal to the number of packets arriving at station i during $[\tau_i(m), \tau_i(m+1)]$, it immediately follows that

$$F_i(z) = C_i[P_i(z)] \tag{5.7}$$

From (5.7), the mean and variance for the cycle time are calculated as

$$E[C_i] = E[\tau_i(m+1) - \tau_i(m)] = \frac{\sum_{k=1}^{N} r_k}{1-\sum_{k=1}^{N} \mu_k} \tag{5.8a}$$

$$Var[C_i] = \frac{Var[L_i^*]}{\mu_i^2} - \frac{\sigma_i^2 \sum_{k=1}^{N} r_k}{\mu_i^2(1-\sum_{k=1}^{N} \mu_k)} \tag{5.8b}$$

Comparing (3.40a) and (5.8a), we see that the mean cycle times in the exhaustive service and gated service systems are the same. We may define the same regenerative process as in Section 3.4.1. Thus the stability condition in the gated service system is also given by (3.45). In the case of identical stations, (5.8a) and (5.8b) become

$$E[C]=\frac{Nr}{1-N\mu}, \quad Var[C]=\frac{\delta^2 N}{(1+\mu)(1-N\mu)} + \frac{\sigma^2 rN^2}{(1+\mu)(1-N\mu)^2} \tag{5.9}$$

This may be compared to (3.41).

Note that the service time $\bar{\tau}_i(m)-\tau_i(m)$ is the number of packets found at the polling as in (5.1). Thus, its GF, $S_i(z)$, is given by

$$S_i(z) = F_i(z) \tag{5.10}$$

As for the intervisit time, we only know its mean:

$$E[I_i] = E[C_i] - E[S_i] = \frac{(1-\mu_i)\sum_{k=1}^{N} r_k}{1-\sum_{k=1}^{N} \mu_k} = (1-\mu_i)E[C_i] \tag{5.11}$$

By comparing (3.33a) with (5.4a), we see that the mean service times in the exhaustive and gated service systems are identical. So is the mean intervisit time from (3.37a) and (5.11).

We now proceed to find the GF, $Q_i(z)$, for the number of packets at station i at an arbitrary time. By the same regenerative arguments as in Section 3.5, $Q_i(z)$ is given by (3.48). The summation domain in the numerator of (3.48) is divided into the service time $[\tau_i(m), \bar{\tau}_i(m)]$ and the intervisit time $[\bar{\tau}_i(m), \tau_i(m+1)]$ as in (3.49). To evaluate the sum over the service time, recall (5.1) to see that

$$L_i(t)=\bar{\tau}_i(m) - t + \text{number of arrivals in } [\tau_i(m),t] \qquad \tau_i(m)\le t\le\bar{\tau}_i(m)-1 \tag{5.12a}$$

Thus we have

$$E\left[\sum_{t=\tau_i(m)}^{\bar{\tau}_i(m)-1} z^{L_i(t)}\right] = E\left[\sum_{t=\tau_i(m)}^{\bar{\tau}_i(m)-1} z^{\bar{\tau}_i(m)-t} (P_i(z))^{t-\tau_i(m)}\right]$$

$$=E\left[z\frac{z^{\bar{\tau}_i(m)-\tau_i(m)} - \{P_i(z)\}^{\bar{\tau}_i(m)-\tau_i(m)}}{z - P_i(z)} \right] = \frac{z\{F_i(z) - F_i[P_i(z)]\}}{z - P_i(z)} \tag{5.12b}$$

For the sum over the intervisit time, note that

$$L_i(t) = \text{number of arrivals in } [\bar{\tau}_i(m),\ t] \quad \bar{\tau}_i(m) \le t \le \tau_i(m+1)-1 \tag{5.13a}$$

Thus we have

$$E\left\{ \sum_{t=\bar{\tau}_i(m)}^{\tau_i(m+1)-1} z^{L_i(t)} \right\} = \sum_{t=\bar{\tau}_i(m)}^{\tau_i(m+1)-1} \{P_i(z)\}^{t-\tau_i(m)}$$

$$= \frac{E[\{P_i(z)\}^{\bar{\tau}_i(m)-\tau_i(m)}] - E[\{P_i(z)\}^{\tau_i(m+1)-\tau_i(m)}]}{1 - P_i(z)}$$

$$= \frac{F_i[P_i(z)] - C_i[P_i(z)]}{1 - P_i(z)} = \frac{F_i[P_i(z)] - F_i(z)}{1 - P_i(z)} \tag{5.13b}$$

where we have used (5.7). Using (5.8a), (5.12b) and (5.13b) in (3.48), we get

$$Q_i(z) = \frac{1 - \sum_{k=1}^{N} \rho_k}{\sum_{k=1}^{N} r_k} \cdot \frac{F_i[P_i(z)] - F_i(z)}{P_i(z) - z} \cdot \frac{(1 - z)P_i(z)}{1 - P_i(z)} \tag{5.14}$$

From (5.14) we have the number of packets at station i at an arbitrary time as

$$E[L_i] = \frac{(1+\mu_i)E[(L_i^*)^2]}{2E[L_i^*]} - \frac{\sigma_i^2}{2\mu_i} \tag{5.15a}$$

In the case of identical stations, we get

$$E[L] = \frac{\delta^2\mu}{2r} + \frac{\sigma^2}{2(1-N\mu)} + \frac{Nr\mu(1+\mu)}{2(1-N\mu)} \tag{5.15b}$$

Lastly we derive the GF, $V_i(z)$, for the waiting time for the first packet in an arbitrary arrival at station i. As before, the waiting time for an arbitrary packet arriving at station i is given by $V_i(z)Y_i(z)$ with $Y_i(z)$ in (3.8a). See (3.53a) and (3.53b). $V_i(z)$ is expressed as in (3.54b) where its denominator is the mean cycle time given by (5.8a). The summation domain in the numerator of (3.54b) is divided into the service time $[\tau_i(m), \overline{\tau}_i(m)]$ and the intervisit time $[\overline{\tau}_i(m), \tau_i(m+1)]$ as in (3.49). We have

$$W_i(t) = \begin{cases} L_i(t) + \tau_i(m+1) - \overline{\tau}_i(m) - 1 & \tau_i(m) \le t \le \overline{\tau}_i(m)-1 \\ L_i(t) + \tau_i(m+1) - t - 1 & \overline{\tau}_i(m) \le t \le \tau_i(m+1)-1 \end{cases} \tag{5.16}$$

Since $L_i(t)$, $t \in [\tau_i(m), \overline{\tau}_i(m)]$, is given by (5.12a), we have

$$E\left[\sum_{t=\tau_i(m)}^{\overline{\tau}_i(m)-1} z^{W_i(t)}\right] = E\left[\sum_{t=\tau_i(m)}^{\overline{\tau}_i(m)-1} z^{\tau_i(m+1)-t-1} \{P_i(z)\}^{t-\tau_i(m)}\right]$$

$$= E\left[\frac{z^{\tau_i(m+1)-\tau_i(m)} - z^{\tau_i(m+1)-\overline{\tau}_i(m)}\{P_i(z)\}^{\overline{\tau}_i(m)-\tau_i(m)}}{z - P_i(z)}\right] \tag{5.17a}$$

Since $L_i(t)$, $t\in[\bar{\tau}_i(m),\tau_i(m+1)]$, is given by (5.13a), we have

$$E\left[\sum_{t=\bar{\tau}_i(m)}^{\tau_i(m+1)-1} z^{W_i(t)}\right] = E\left[\sum_{t=\bar{\tau}_i(m)}^{\tau_i(m+1)-1} \{P_i(z)\}^{t-\tau_i(m)} z^{\tau_i(m+1)-t-1}\right]$$

$$= E\left[\frac{z^{\tau_i(m+1)-\bar{\tau}_i(m)}\,\{P_i(z)\}^{\bar{\tau}_i(m)-\tau_i(m)} - \{P_i(z)\}^{\tau_i(m+1)-\tau_i(m)}}{z - P_i(z)}\right] \tag{5.17b}$$

Using (5.8a), (5.17a) and (5.17b) in (3.54b), we get

$$V_i(z) = \frac{1}{E[C_i]}\, E\left[\frac{z^{\tau_i(m+1)-\tau_i(m)} - \{P_i(z)\}^{\tau_i(m+1)-\tau_i(m)}}{z - P_i(z)}\right] \tag{5.17c}$$

By use of (5.7), we obtain

$$V_i(z) = \frac{1}{E[C_i]} \cdot \frac{C_i(z) - F_i(z)}{z - P_i(z)} \tag{5.18}$$

From (5.18) we have the mean waiting time for the first packet in an arbitrary arrival at station i as

$$E[V_i] = \frac{(1+\mu_i)E[C_i^2]}{2E[C_i]} - \frac{1+\mu_i}{2} \tag{5.19a}$$

or, in terms of L_i^*,

$$E[V_i] = \frac{(1+\mu_i)E[L_i^{*2}]}{2\mu_i E[L_i^*]} - \frac{(1+\mu_i)\sigma_i^2}{2\mu_i^2} - \frac{1+\mu_i}{2} \tag{5.19b}$$

Using (3.53a) and (3.53b) together with (3.8a) and (3.8b), the GF and mean for the waiting time of an arbitrary packet are respectively given by

$$U_i(z) = \frac{1}{\mu_i E[C_i]} \cdot \frac{1-P_i(z)}{z-P_i(z)} \cdot \frac{C_i(z) - F_i(z)}{1 - z} \qquad (5.20a)$$

and

$$E[U_i] = \frac{(1+\mu_i)E[\{L_i^*\}^2]}{2\mu_i E[L_i^*]} - \frac{\sigma_i^2}{2\mu_i^2} - 1 \qquad (5.20b)$$

Comparing (5.15a) and (5.20b), we again confirm Little's relation as in (3.60).

In the case of identical stations, we have

$$E[V] = \frac{\delta^2}{2r} + \frac{\sigma^2 N}{2(1-N\mu)} - \frac{1+\mu}{2} + \frac{Nr(1+\mu)}{2(1-N\mu)} \qquad (5.21a)$$

and

$$E[U] = \frac{\delta^2}{2r} + \frac{\sigma^2}{2\mu(1-N\mu)} + \frac{Nr(1+\mu)}{2(1-N\mu)} - 1 \qquad (5.21b)$$

From (3.61b) and (5.21b) we see that

$$E[U]\Big|_{\text{gated}} - E[U]\Big|_{\text{exhaustive}} = \frac{Nr\mu}{1-N\mu} = \mu E[C] \qquad (5.22)$$

Thus, for identical stations, the mean wait for the gated service system is always greater than that for the exhaustive service system.

As in Section 3.6.1, we may use (5.21a) to get the mean message waiting time

$$E[W] = \frac{\delta^2}{2r} + \frac{N[\lambda b^{(2)} + r(1+\lambda b)]}{2(1 - N\lambda b)} + \frac{(\gamma^2 - \lambda)b}{2\lambda(1-N\lambda b)} - \frac{1}{2} \tag{5.23}$$

From (3.63b) and (5.23) we again see that

$$E[W]\Big|_{gated} - E[W]\Big|_{exhaustive} = \frac{Nr\lambda b}{1 - N\lambda b} = \lambda b E[C] \tag{5.24}$$

From (5.23), we can get the mean message waiting time at station 1, $E[W_1]$, when all traffic is generated at station 1 (see Section 3.6.1 for derivation):

$$E[W_1] = \frac{\delta^2}{2r} + \frac{N[\lambda b^{(2)} + r(1+N\lambda b)]}{2(1 - N\lambda b)} + \frac{(\gamma^2 - \lambda)b}{2\lambda(1-N\lambda b)} - \frac{1}{2} \tag{5.25a}$$

which leads to

$$E[W_1] - E[W] = \frac{N(N-1)\lambda rb}{2(1 - N\lambda b)} \geq 0 \tag{5.25b}$$

Note that the relationship between $E[W]$ and $E[W_1]$ here is exactly opposite to the case of exhaustive service system as shown in (3.64b). Also, from (3.64b), (5.24) and (5.25b), we have the inequalities [Rubi83]:

$$E[W_1]\Big|_{exhaustive} \leq E[W]\Big|_{exhaustive} \leq E[W]\Big|_{gated} \leq E[W_1]\Big|_{gated} \tag{5.26}$$

5.2 Continuous-Time System

This section deals with the continuous-time system; the assumptions and notation for the message length, message arrival process and reply interval are carried over from Chapter 4.

We first find the relation between $F_i(z_1, z_2, \ldots, z_N)$ and $F_{i+1}(z_1, z_2, \ldots, z_N)$. In the case of gated service system, the number of messages served in the service time for station i, $\bar{\tau}_i(m) - \tau_i(m)$, is simply $L_i(\tau_i(m))$, and each message service takes a time given by $B_i^*(s)$. Thus, we have

$$E[e^{-s\{\bar{\tau}_i(m)-\tau_i(m)\}}] = [B_i^*(s)]^{L_i(\tau_i(m))} \tag{5.27a}$$

Hence the joint GF for the number of message arrivals during the service period for station i is given by

$$\{B_i^*[\sum_{j=1}^{N} (\lambda_j - \lambda_j z_j)]\}^{L_i(\tau_i(m))} \tag{5.27b}$$

The joint GF for the number of message arrivals during the reply interval between station i and station i+1 is given by

$$R_i^*[\sum_{j=1}^{N} (\lambda_j - \lambda_j z_j)] \tag{5.28}$$

Hence, we have

$$\begin{aligned} F_{i+1}(z_1,z_2,\ldots,z_N) = R_i^*[\sum_{j=1}^{N}(\lambda_j-\lambda_j z_j)]\cdot \\ \cdot F_i(z_1,z_2,\ldots,z_{i-1},B_i^*[\sum_{j=1}^{N}(\lambda_j-\lambda_j z_j)],\ z_{i+1},\ldots,z_N) \end{aligned} \tag{5.29}$$

Now, define the moments for $[L_1(t), L_2(t), \ldots, L_N(t)]$ at time $t=\tau_i(m)$ as in (3.27), (3.28a) and (3.28b). A set of N^2 equations for $\{f_i(j);\ i,j=$

$1,2,\ldots,N\}$ are given by

$$f_{i+1}(i) = r_i\lambda_i + \rho_i f_i(i) \tag{5.30a}$$

$$f_{i+1}(j) = r_i\lambda_j + \lambda_j b_i f_i(i) + f_i(j) \qquad j\neq i \tag{5.30b}$$

The solution to (5.30a) and (5.30b) is

$$E[L_i^*] = f_i(i) = \frac{\lambda_i \sum_{k=1}^{N} r_k}{1 - \sum_{k=1}^{N} \rho_k} \tag{5.31a}$$

$$f_i(j) = \lambda_j \left[\sum_{k=j}^{i-1} r_k + \frac{\left(\sum_{k=j}^{i-1} \rho_k\right)\left(\sum_{k=1}^{N} r_k\right)}{1 - \sum_{k=1}^{N} \rho_k} \right] \qquad j\neq i \tag{5.31b}$$

A set of N^3 equations for $\{f_i(j,k);i,j,k=1,2,\ldots,N\}$ are given by

$$\begin{aligned} f_{i+1}(j,k) = {} & \lambda_j\lambda_k(\delta_i^2+r_i^2)+r_i\lambda_k f_i(j)+r_i\lambda_j f_i(k)+f_i(i)\lambda_j\lambda_k(2r_ib_i+b_i^{(2)})+f_i(j,k) \\ & +b_i\lambda_k f_i(i,j)+b_i\lambda_j f_i(i,k)+b_i^2\lambda_j\lambda_k f_i(i,i) \qquad i\neq j,\ i\neq k \end{aligned} \tag{5.32a}$$

$$\begin{aligned} f_{i+1}(i,k) = {} & \lambda_i\lambda_k(\delta_i^2+r_i^2)+r_i\lambda_i f_i(k)+f_i(i)\lambda_i\lambda_k(2r_ib_i+b_i^{(2)})+\lambda_i b_i f_i(i,k) \\ & +\lambda_i\lambda_k b_i^2 f_i(i,i) \qquad i\neq k \end{aligned} \tag{5.32b}$$

$$f_{i+1}(i,i)=\lambda_i^2(\delta_i^2+r_i^2)+f_i(i)\lambda_i^2(2r_ib_i+b_i^{(2)})+(\lambda_ib_i)^2 f_i(i,i) \tag{5.32c}$$

In the case of identical stations, we have

$$E[L^*] = \frac{Nr}{1 - N\rho} \tag{5.33a}$$

$$f_i(i,i) = \frac{\delta^2\lambda^2 N}{(1-N\rho)(1+\rho)} + \frac{N^2\lambda^3 r b^{(2)}}{(1-N\rho)^2(1+\rho)} + \frac{N^2 r^2\lambda^2}{(1-N\rho)^2} \qquad (5.33b)$$

Let us now consider the service time, intervisit time, cycle time, and number of messages served in a cycle. First, the LST of the distribution function for the service time for station i is given by (5.27a), i.e.,

$$S_i^*(s) = F_i[B_i^*(s)] \qquad (5.34)$$

from which we have

$$E[S_i] = b_i E[L_i^*] = \frac{\rho_i \sum_{k=1}^{N} r_k}{1 - \sum_{k=1}^{N} \rho_k} \qquad (5.35a)$$

$$\mathrm{Var}[S_i] = b_i^2 \mathrm{Var}[L_i^*] + (b_i^{(2)} - b_i^2) E[L_i^*] \qquad (5.35b)$$

In the case of identical stations, by using (5.33b), we have

$$E[S] = \frac{Nr\rho}{1-N\rho}, \quad \mathrm{Var}[S] = \frac{\delta^2 N\rho^2}{(1+\rho)(1-N\rho)} + \frac{N\lambda r b^{(2)}(1+\rho-N\rho)}{(1+\rho)(1-N\rho)^2} \qquad (5.36)$$

which corresponds to (5.6) in the discrete-time system.

The number of messages served in a cycle is exactly the number of messages found at the polling instant:

$$T_i(m) = L_i(\tau_i(m)) \qquad (5.37a)$$

and so, by (5.31a), we have

$$E[T_i(m)] = E[L_i(\tau_i(m))] = E[L_i^*] = \frac{\rho_i \sum_{k=1}^{N} r_k}{1 - \sum_{k=1}^{N} \rho_k} \qquad (5.37b)$$

The LST of the distribution function for the cycle time for station i is given through a relationship similar to (5.7):

$$F_i(z) = C_i^*(\lambda_i - \lambda_i z) \tag{5.38}$$

or

$$C_i^*(s) = F_i(1 - s/\lambda_i) \tag{5.39a}$$

from which we have

$$E[C_i] = \frac{E[L_i^*]}{\lambda_i} = \frac{\sum_{k=1}^{N} r_k}{1 - \sum_{k=1}^{N} \rho_k} \quad , \quad E[\{C_i\}^2] = \frac{f_i(i,i)}{\lambda_i^2} \tag{5.39b}$$

In the case of identical stations, by use of (5.33b), we have

$$E[C] = \frac{Nr}{1-N\rho} \quad , \quad Var[C] = \frac{\delta^2 N}{(1-N\rho)(1+\rho)} + \frac{N^2 \lambda r b^{(2)}}{(1-N\rho)^2 (1+\rho)} \tag{5.40}$$

which corresponds to (5.9) in the discrete-time system, and to (4.24) in the exhaustive service system. From (5.39b), the stability condition is given by (4.25).

As for the intervisit time, we only know its mean:

$$E[I_i] = E[C_i] - E[S_i] = \frac{(1-\rho_i) \sum_{k=1}^{N} r_k}{1 - \sum_{k=1}^{N} \rho_k} = (1-\rho_i)E[C_i] \tag{5.41}$$

Compare (4.14a) and (5.35a); (4.20b) and (5.41); (4.23a) and (5.39b); (4.17a) and (5.37b). We see that they are respectively identical.

We next derive the GF, $Q_i(z)$, for the number of messages left at station i after the message service completion at station i. By the same regenerative arguments as in Section 4.3, $Q_i(z)$ is given by (4.28). Note that the denominator is given by (5.37b). To evaluate the sum in the numerator of (4.28), note that

$$L_i(\tau_i^{(n)}(m)) = T_i(m) - n + \text{number of arrivals during n service times} \tag{5.42a}$$

Thus we have

$$E\left[\sum_{n=1}^{T_i(m)} z^{L_i(\tau_i^{(n)}(m))}\right] = E\left[\sum_{n=1}^{T_i(m)} z^{T_i(m)-n}\{B_i^*(\lambda_i-\lambda_i z)\}^n\right]$$

$$= \frac{B_i^*(\lambda_i-\lambda_i z)}{z-B_i^*(\lambda_i-\lambda_i z)} E\left[z^{T_i(m)} - \{B_i^*(\lambda_i-\lambda_i z)\}^{T_i(m)}\right]$$

$$= \frac{B_i^*(\lambda_i-\lambda_i z)}{z-B_i^*(\lambda_i-\lambda_i z)} \{F_i(z) - F_i[B_i^*(\lambda_i-\lambda_i z)]\} \tag{5.42b}$$

Using (5.37b) and (5.42b) in (4.28), we get

$$Q_i(z) = \frac{1-\sum_{k=1}^{N}\rho_k}{\lambda_i \sum_{k=1}^{N} r_k} \cdot \frac{B_i^*(\lambda_i - \lambda_i z)}{z - B_i^*(\lambda_i - \lambda_i z)} \cdot \{F_i(z)-F_i[B_i^*(\lambda_i - \lambda_i z)]\} \tag{5.43}$$

From (5.43) we obtain

$$E[L_i] = \rho_i + \frac{(1-\sum_{k=1}^{N}\rho_k)(1+\rho_i)\, f_i\,(i,i)}{2\lambda_i \sum_{k=1}^{N} r_k} \tag{5.44a}$$

In the case of identical stations, by using (5.33b) we have

$$E[L] = \rho + \frac{\delta^2\lambda}{2r} + \frac{Nr\lambda(1+\rho)}{2(1-N\rho)} + \frac{N\lambda^2 b^{(2)}}{2(1-N\rho)} \tag{5.44b}$$

We have derived (5.43), (5.44a) and (5.44b) for the number of messages at message departure instants. However, for the same reason mentioned after (4.30b), we may think that they hold at any instant.

By an argument similar to Section 4.4, we may find $W_i^*(s)$, the LST of the distribution function for the message waiting time at station i. Using (5.43), we have

$$\bar{W}_i^*(s) = \frac{Q_i(1-s/\lambda_i)}{B_i^*(s)} = \frac{1-\sum_{k=1}^{N}\rho_k}{\sum_{k=1}^{N} r_k} \cdot \frac{F_i[B_i^*(s)]-F_i(1-s/\lambda_i)}{s-\lambda_i+\lambda_i B_i^*(s)} \tag{5.45}$$

From (5.45), we get

$$E[W_i] = \frac{(1-\sum_{k=1}^{N}\rho_k)(1+\rho_i)f_i(i,i)}{2\lambda_i^2 \sum_{k=1}^{N} r_k} \tag{5.46a}$$

From (5.44a) and (5.46a) we may again confirm Little's relation (4.34). In the case of identical stations, by using (5.33b) we have

$$E[W] = \frac{\delta^2}{2r} + \frac{Nr(1+\rho)}{2(1-N\rho)} + \frac{N\lambda b^{(2)}}{2(1-N\rho)} \tag{5.46b}$$

Again, note that (5.46b) has the same form as (5.23) when Poisson arrival is assumed ($\gamma^2=\lambda$) and "a half slot" is made shrink to zero in (5.23).

We may express $W_i^*(s)$ in terms of $C_i^*(s)$. Namely, by use of (5.39a) and (5.39b) in (5.45), we get

$$W_i^*(s) = \frac{1}{E[C_i]} \cdot \frac{C_i^*[\lambda_i - \lambda_i B_i^*(s)] - C_i^*(s)}{s - \lambda_i + \lambda_i B_i^*(s)} \qquad (5.47a)$$

from which it follows (also from (5.39b) and (5.46a)) that

$$E[W_i] = \frac{(1 + \rho_i) E[\{C_i\}^2]}{2E[C_i]} \qquad (5.47b)$$

We now present an algorithm to compute $E[\{C_i\}^2]$ by solving N^2 linear equations using an idea similar to that in Section 4.5. (Our treatment is based on [Ferg84]. A similar analysis is in [Cars77].) For the gated-service system, we define the station time $\tilde{x}_k$ at the k th visit as the service time for the current station plus the reply interval to the next station. See Figure 6. Thus, if the k th visit is to station i, we have

$$\tilde{x}_k = \tau_{i+1}(m) - \tau_i(m) \qquad (5.48)$$

The cycle time before the k th visit, C_k, may be expressed as

$$C_k = \sum_{j=1}^{N} \tilde{x}_{k-j} = \sum_{m=0}^{N-1} \tilde{x}_{k-N+m} \qquad (5.49)$$

The probability density function $P_k(x_{k-(N-1)}, \ldots, x_{k-1}, x_k)$ and its Laplace transform $X_k^*(s_1, s_2, \ldots, s_N)$ for the N successive station times $\underline{x}_k = [\tilde{x}_{k-(N-1)}, \ldots, \tilde{x}_{k-1}, x_k]$ are related by (4.43). By an argument similar to one leading to (4.44), we have

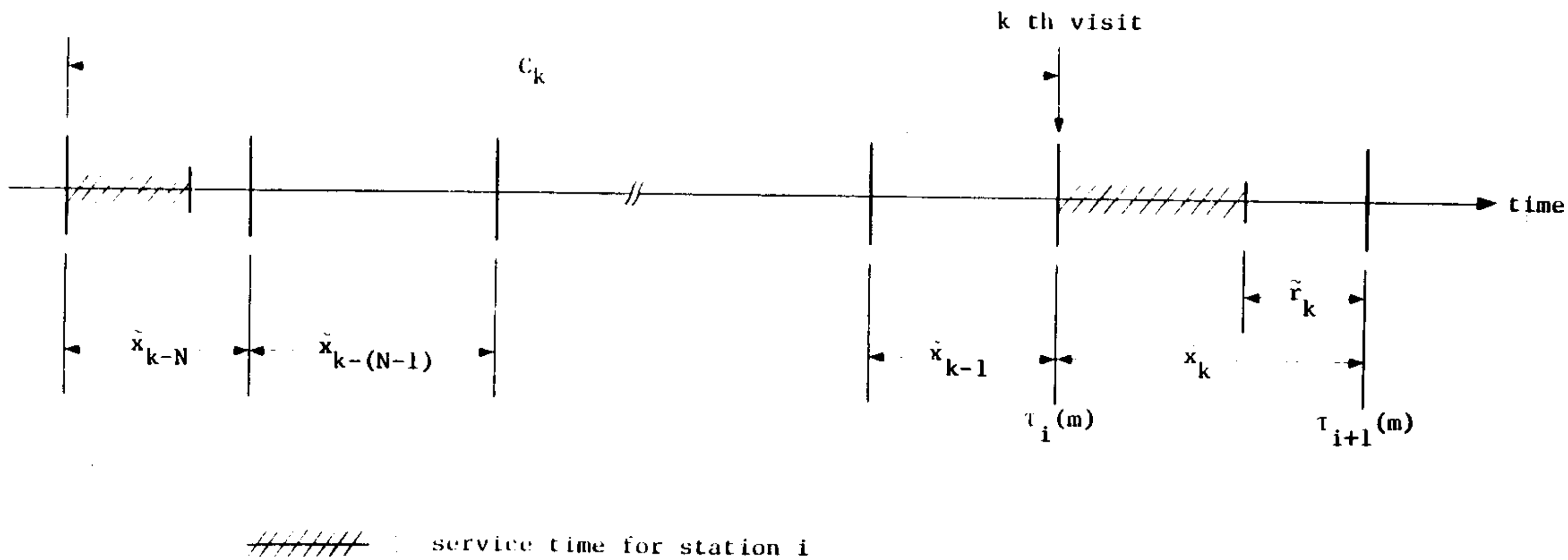

Figure 6. Cycle time.

$$P_k(x_{k-(N-1)},\ldots,x_{k-1},x_k) = \int_0^{\infty} P_{k-1}(x_{k-N},x_{k-N+1},\ldots,x_{k-1})\,dx_{k-N}$$

$$\cdot\int_0^{x_k} dR_k(t) \sum_{m=0}^{\infty} b_k^{*(m)}(x_k-t)\, \frac{1}{m!}\, (\lambda_k \sum_{j=1}^{N} x_{k-j})^m \exp[-(\lambda_k \sum_{j=1}^{N} x_{k-j})]$$

(5.50)

where $b_k^{*(m)}(t)$ denotes the m fold convolution of $b_k(t)$, the probability density function for the service time at the k th visited station. The Laplace transform of (5.50) is now given by

$$X_k^*(s_1, s_2, \ldots, s_N) = R_k^*(s_N)$$

$$\cdot X_{k-1}^*(\lambda_k[1-B_k^*(s_N)],\ s_1+\lambda_k[1-B_k^*(s_N)],\ \ldots,\ s_{N-1}+\lambda_k[1-B_k^*(s_N)])$$

(5.51)

From (5.51) we may derive the mean of $\tilde{x}_k$:

$$\overline{x}_k \triangleq E[\tilde{x}_k] = r_k + \rho_k E[C_k] \tag{5.52}$$

and the covariance of $\tilde{x}_k$ and $\tilde{x}_{k-j}$:

$$E[(\tilde{x}_k-\overline{x}_k)^2] = \sigma_k^2 + \lambda_k b_k^{(2)} E[C_k] + \rho_k^2\, Var[C_k] \tag{5.53a}$$

$$E[(\tilde{x}_k-\overline{x}_k)(\tilde{x}_{k-j}-\overline{x}_{k-j})] = \rho_k \sum_{m=1}^{N} E[(\tilde{x}_{k-m}-\overline{x}_{k-m})(\tilde{x}_{k-j}-\overline{x}_{k-j})] \tag{5.53b}$$

where (from (5.49))

$$Var[C_k] = \sum_{m=1}^{N} \sum_{n=1}^{N} E[(\tilde{x}_{k-n} - \overline{x}_{k-n})(\tilde{x}_{k-m} - \overline{x}_{k-m})] \tag{5.53c}$$

If we define r_{ij} as in (4.50), equations (5.53a-c) are converted to

$$r_{ii} = \sigma_i^2 + \lambda_i b_i^{(2)} E[C_i] + \rho_i^2\, Var[C_i] \tag{5.54a}$$

$$r_{ij} = \rho_i \left(\sum_{m=i}^{N} r_{jm} + \sum_{m=i}^{j-1} r_{jm} + \sum_{m=j}^{i-1} r_{mj} \right) \qquad j<i \tag{5.54b}$$

$$r_{ij} = \rho_i \left(\sum_{m=i}^{j-1} r_{jm} + \sum_{m=j}^{N} r_{mj} + \sum_{m=1}^{i-1} r_{mj} \right) \qquad j>i \qquad (5.54c)$$

where

$$\mathrm{Var}[C_i] = \sum_{n=1}^{i-1} \left(\sum_{m=i}^{N} r_{nm} + \sum_{m=1}^{n-1} r_{nm} + \sum_{m=n}^{i-1} r_{mn} \right)$$

$$+ \sum_{n=i}^{N} \left(\sum_{m=i}^{n-1} r_{nm} + \sum_{m=n}^{N} r_{mn} + \sum_{m=1}^{i-1} r_{mn} \right) \qquad (5.54d)$$

As before, it is possible to write $\mathrm{Var}[C_i]$ in a simpler form

$$\mathrm{Var}[C_i] = \frac{1}{\rho_i} \sum_{\substack{j=1 \\ (j\neq i)}}^{N} r_{ij} + \sum_{j=1}^{N} r_{ji} \qquad (5.55a)$$

Substituting (5.55a) back into (5.54a), we have

$$r_{ii} = \delta_i^{\,2} + \lambda_i b_i^{(2)} E[C_i] + \rho_i \sum_{\substack{j=1 \\ (j\neq i)}}^{N} r_{ij} + \rho_i^{\,2} \sum_{j=1}^{N} r_{ji} \qquad (5.55b)$$

Now we see that equations (5.54b), (5.54c) and (5.55b) constitute a system of N^2 linear equations for $\{r_{ij};\ i,j=1,2,\ldots,N\}$. Using the solution in (5.55a), we can compute $\mathrm{Var}[C_i]$ which is used in (5.47b) to calculate $E[W_i]$.

As an example, consider the case N=2. By solving the above equations, we have

$$\mathrm{Var}[C_1] = \frac{\begin{array}{c}(\delta_1^{\,2}+\lambda_1 b_1^{(2)} E[C_1])(1+2\rho_2-2\rho_2^{\,3}-\rho_1\rho_2-2\rho_1\rho_2^{\,2}) \\ +\ (\delta_2^{\,2}+\lambda_2 b_2^{(2)} E[C_2])(1+\rho_2\rho_2)\end{array}}{(1-\rho_1\rho_2)(1-\rho_1-\rho_2)(1+\rho_1+\rho_2+2\rho_1\rho_2)} \qquad (5.56a)$$

As before this can be used to obtain the mean waiting time in the system of zero reply intervals. In this case, from (5.47b) and (5.56a), we have

$$E[W_1] = \frac{(1+\rho_1)[\lambda_1 b_1^{(2)}(1+2\rho_2-2\rho_2^3-\rho_1\rho_2-2\rho_1\rho_2^2)+\lambda_2 b_2^{(2)}(1+\rho_1\rho_2)]}{2(1-\rho_1\rho_2)(1-\rho_1-\rho_2)(1+\rho_1+\rho_2+2\rho_1\rho_2)} \tag{5.56b}$$

We can again confirm the conservation law in (4.37a) and (4.37b).

In the case of identical stations, we have

$$r_{ij} = \frac{\rho[\delta^2(1-N\rho)+N\lambda rb^{(2)}]}{(1-N\rho)^2(1+\rho)} \qquad i\neq j \tag{5.57a}$$

$$r_{ii} = \frac{(1+\rho-N\rho)[\delta^2(1-N\rho)+N\lambda rb^{(2)}]}{(1-N\rho)^2(1+\rho)} \tag{5.57b}$$

which leads to (5.40).

Lastly, let us play with (5.46b), the mean message waiting time formula in the case of identical stations. From (4.33b) and (5.46b), we have

$$E[W]_{gated} - E[W]_{exhaustive} = \frac{Nr\rho}{1-N\rho} = \rho E[C] \tag{5.58}$$

As before, the mean message waiting time at station 1, $E[W_1]$, when all traffic is generated at station 1 is given by

$$E[W_1] = \frac{\delta^2}{2r} + \frac{Nr(1+N\rho)}{2(1-N\rho)} + \frac{N\lambda b^{(2)}}{2(1-N\lambda b)} \tag{5.59}$$

From (5.46b) and (5.59) we have the same relationship as (5.25b). Thus, we also have a relationship (5.26) in the continuous-time system. The mean message waiting time in a Poisson bulk arrival system (defined in Section 4.4) is given by

$$E[W] = \frac{\delta^2}{2r} + \frac{Nr(1+\lambda gb)}{2(1-N\lambda gb)} + \frac{N\lambda gb^{(2)}}{2(1-N\lambda gb)} + \frac{(g^{(2)}-g)b}{2g(1-N\lambda gb)} \tag{5.60}$$

6 Limited Service Systems

6.1 Symmetric, Continuous-Time System

6.1.1 Problem Statement and Formulation

We consider a system of N identical stations served by a single server. Each station, with infinite capacity, has an independent Poisson message arrival stream with a rate λ. Let $B^*(s)$, b and $b^{(2)}$ be the LST of the distribution function, mean and second moment, respectively, for the message length (measureed in service time): $b = -B^{*(1)}(0)$ and $b^{(2)} = B^{*(2)}(0)$. The server administers service to one message at a time, if any, from station i, and with a finite reply interval it goes to inspect station i+1. Let $R^*(s)$, r and δ^2 be the LST of the distribution function, mean and variance, respectively, for the reply interval: $r = -R^{*(1)}(0)$ and $\delta^2 = R^{*(2)}(0) - r^2$. We note that all these given parameters are the same for all stations, i.e., we have a symmetric system.

Let

$L_i(t) \triangleq$ number of messages at station i at time t

and define the joint GF for $[L_1(t), L_2(t), \ldots, L_N(t)]$ at time $t = \tau_i(m)$, i.e., at the time when station i is polled as in (3.21a):

$$F_i(z_1, z_2, \ldots, z_N) \triangleq E\left[\prod_{j=1}^{N} z_j^{L_j(\tau_i(m))}\right] \tag{6.1}$$

Similarly, define the GF for $L_i(t)$ at the time when a service at station i is finished:

$$Q(z) = E[z^{L_i(\text{service completion time})}] \tag{6.2}$$

(Here subscritp i is dropped from Q(z) because it is the same for all i.) Note that this "service completion time" is different from $\tau_i(m)$; since

$\{\bar{\tau}_i(m)\}$ refer not only to the instants when the service is actually given and finished (service completion times) but also to the instants when there are no messages found at the polling and the server goes to check the next station. Hereafter, we choose station 1 as a representative. Now, the relationship between $F_1(z, 1, \ldots, 1)$ and $Q(z)$ is provided by

$$Q(z) = \frac{F_1(z, 1, \ldots, 1) - F_1(0, 1, \ldots, 1)}{z[1-F_1(0, 1, \ldots, 1)]} B^*(\lambda-\lambda z) \qquad (6.3)$$

where $F_1(0, 1, \ldots, 1)$ is the (marginal) probability that there are no messages at station 1 when station 1 is polled. From (6.3), we have

$$Q^{(1)}(1) = \frac{\left[\frac{dF_1(z, 1, \ldots, 1)}{dz}\right]_{z=1}}{1-F_1(0, 1, \ldots, 1)} - 1 + \lambda b \qquad (6.4)$$

Our aim is to find the mean message waiting time $E[W]$ which is evaluated from $W^*(s)$, the LST of the distribution function for the message waiting time via $E[W] = -W^{*(1)}(0)$. Since the messages left at station 1 when a message at station 1 has completed service are those which arrive while that message is in the system, it follows that

$$W^*(\lambda - \lambda z)B^*(\lambda - \lambda z) = Q(z) \qquad (6.5)$$

Hence we have

$$\lambda E[W] + \lambda b = Q^{(1)}(1) \qquad (6.6)$$

which is the mean number of messages in each station at the time of message departure from that station. Thus, by use of (6.4) and (6.6) we may compute $E[W]$ once we have evaluated $F_1(0, 1, \ldots, 1)$ and $[dF_1(z, 1, \ldots, 1)/dz]_{z=1}$. This we do in the next section.

6.1.2 Mean Message Waiting Time

Let us derive the relation between $F_i(z_1, z_2, \ldots, z_N)$ and $F_{i+1}(z_1, z_2, \ldots, z_N)$. To express $F_{i+1}(z_1, z_2, \ldots, z_N)$ in terms of $F_i(z_1, z_2, \ldots, z_N)$, note that $[L_1(t), L_2(t), \ldots, L_N(t)]$ at $t = \tau_{i+1}(m)$ is $[L_1(t), L_2(t), \ldots, L_N(t)]$ at $t = \tau_i(m)$ plus the number of arrivals during a message service time (if $L_i(\tau_i(m)) > 0$; in this case also minus one from $L_i(t)$) and a reply interval. Thus we get

$$F_{i+1}(z_1, z_2, \ldots, z_N) = R^*[\sum_{j=1}^{N} (\lambda-\lambda z_j)]$$

$$\cdot \{B^*[\sum_{j=1}^{N}(\lambda-\lambda z_j)]\frac{1}{z_i}[F_i(z_1, z_2, \ldots, z_N) - F_i(z_1, \ldots, z_{i-1}, 0, z_{i+1}, \ldots, z_N)]$$

$$+ F_i(z_1, \ldots, z_{i-1}, 0, z_{i+1}, \ldots, z_N)\} \qquad i = 1, 2, \ldots, N \qquad (6.7)$$

This is the governing equation of our system.

Now, setting $z_k = z$ and $z_1 = \ldots = z_{k-1} = z_{k+1} = \ldots = z_N = 1$ in (6.7) yields

$$F_{i+1}(1, \ldots, 1, z, 1, \ldots, 1) = R^*(\lambda-\lambda z)$$

$$\cdot\{B^*(\lambda-\lambda z)\ \frac{1}{z}[F_i(1, \ldots, 1, z, 1, \ldots, 1) - F_i(1, \ldots, 1, 0, 1, \ldots, 1)]$$

$$+ F_i(1, \ldots, 1, 0, 1, \ldots 1)\} \qquad k = i \qquad (6.8)$$

and

$$F_{i+1}(1, \ldots, 1, z, 1, \ldots, 1) = R^*(\lambda-\lambda z)$$

$$\cdot\{B^*(\lambda-\lambda z)[F_i(1, \ldots, 1, z, 1, \ldots, 1) - F_i(1, \ldots, 1, z, 1, \ldots, 1, 0, 1, \ldots 1)]$$

$$+ F_i(1, \ldots, 1, z, 1, \ldots, 1, 0, 1, \ldots, 1)\} \qquad k \neq i \qquad (6.9)$$

where z and 0 appear in the k th and i th positions, respectively, of F_{i+1}

and F_i. Differentiating each of equations (6.8) and (6.9) with respect to z, letting $z = 1$ and adding them up, we have

$$F_1(0,\ 1,\ \ldots,\ 1) = \frac{1-N\lambda(r+b)}{1-N\lambda b} \tag{6.10}$$

where we have used symmetry condition.

Next, we obtain the equations which express $F_k(z,\ 1,\ \ldots,\ 1)$ in terms of $F_1(0,\ 1,\ \ldots,\ 1)$ and $F_j(z,\ 1,\ \ldots,\ 1,\ 0,\ 1,\ \ldots,\ 1)$ (where 0 is the j th argument) ($j = 2,\ 3,\ \ldots,\ N$) for each k. They are derived by repeatedly using (6.7). Introducing notational convention

$$R \equiv R^*(\lambda-\lambda z)\ ;\ B \equiv B^*(\lambda-\lambda z) \tag{6.11}$$

we have

$$F_1(z,1,\ldots,1) = \frac{(RB)^{N-1}R(z-B)F_1(0,1,\ldots,1)+\sum_{j=2}^{N}(RB)^{N-j}R(1-B)zF_j(z,1,\ldots,1,0,1,\ldots,1)}{z-(RB)^N} \tag{6.12}$$

and

$$F_k(z,1,\ldots,1) = \frac{\left\{\begin{array}{l}(RB)^{k-2}R(z-B)F_1(0,1,\ldots 1)+\sum_{j=2}^{k-1}(RB)^{k-j-1}R(1-B)zF_j(z,1,\ldots,1,0,1,\ldots 1)\\ \quad +\sum_{j=k}^{N}(RB)^{N-j+k-1}R(1-B)F_j(z,1,\ldots,1,0,1,\ldots 1)\end{array}\right\}}{z-(RB)^N}$$

$$k = 2,\ 3,\ \ldots,\ N \tag{6.13}$$

Thirdly, by setting $z_1=z_2=\ldots=z_N=z$ in (6.7) and noting symmetry, we obtain

$$F_1(z,\ldots,z) = \frac{R^*[N(\lambda-\lambda z)]\{z-B^*[N(\lambda-\lambda z)]\}F_1(0,z,\ldots,z)}{z-R^*[N(\lambda-\lambda z)]B^*[N(\lambda-\lambda z)]} \tag{6.14}$$

To find $[dF_1(z, 1, \ldots, 1)/dz]_{z=1}$, we use a relation

$$\left[\frac{dF_1(z,\ldots,z)}{dz}\right]_{z=1} = \sum_{k=1}^{N}\left[\frac{dF_k(z,1,\ldots,1)}{dz}\right]_{z=1} \tag{6.15}$$

which represents the sum of message numbers in all stations when station 1 is polled (due to symmetry). In evaluating each term of (6.15) from (6.12), (6.13) and (6.14), it turns out that each term of (6.15) can be expressed with a single unknown constant

$$\left[\frac{dF_1(0,z,\ldots,z)}{dz}\right]_{z=1} = \sum_{j=2}^{N}\left[\frac{dF_j(z,1,\ldots,1,0,1,\ldots,1)}{dz}\right]_{z=1} \tag{6.16}$$

which represents the sum of message numbers in stations 2 through N when station 1 is polled and found empty (due to symmetry). Thus we may solve (6.15) for $[dF_1(0,z,\ldots,z)/dz]_{z=1}$ to obtain

$$\left[\frac{dF_1(0,z,\ldots,z)}{dz}\right]_{z=1} = \frac{N(N-1)\lambda r}{2[1-N\lambda(r+b)]}\, F_1(0,1,\ldots,1) \tag{6.17}$$

The mean number of messages at station 1 when station 1 is polled is then given by

$$\left[\frac{dF_1(z,1,\ldots,1)}{dz}\right]_{z=1} = F_1(0,1,\ldots,1)$$

$$\cdot \frac{N\lambda^2(r^2+\delta^2)+2N\lambda r}{2[1-N\lambda(r+b)]} + \frac{N\lambda r[2N\lambda^2 rb-(N+1)\lambda r+N\lambda^2 b^{(2)}]}{2(1-N\lambda b)[1-N\lambda(r+b)]} \tag{6.18}$$

Finally, substituting (6.10) and (6.18) into (6.4) and (6.6), we get

$$E[W] = \frac{\delta^2}{2r} + \frac{N[\lambda b^{(2)}+r(1+\lambda b)+\lambda\delta^2]}{2[1-N\lambda(r+b)]} \tag{6.19}$$

Comparing (6.19) to (4.33b) and (5.46b), we have the inequalities in the symmetric, continuous-time systems:

$$E[W]|_{exhaustive} \le E[W]|_{gated} \le E[W]|_{limited} \qquad (6.20)$$

This disadvantage of the limited service system, however, is compensated for by the fairness among stations.

6.1.3 Decrementing Service System

A similar analysis can be carried out for the "decrementing" service system [Takag84] which is defined as follows: when there is at least one message found at the time of a polling for a station, the service for that station continues until the number of queued messages decreases to one less than that found at the polling instant. Although there seem no such service schemes applied in a real-world system at the present, this scheme complements the drawbacks in the exhaustive and limited service systems in the following sense. As we show below, this scheme gives the mean message waiting time smaller than that for the limited service (K = 1) system (though it is greater than that for the exhaustive service system). Yet our scheme is expected to prevent, to some extent, though not completely, a heavily loaded station from monopolizing the server. Thus, we claim that our shceme falls in between the exhaustive and limited service systems in terms of the mean message waiting time and fairness among stations.

The parameters given to our system are the same as described at the beginning of Section 6.1.1. We also define $F_i(z_1, z_2, \ldots, z_N)$ and $Q(z)$ as in (6.1) and (6.2).

We now note that the number of messages left behind at station 1 when message service at station 1 is finished consists of two independent components. One is the number of messages at the polling instant minus one whose GF is denoted by $Q_1(z)$. The other is the number of messages which arrive and are served during the contiguous service period for station 1 which is generated by a single message service time; we denote by $Q_2(z)$ the GF for the latter. (Here we have assumed the change in order of service for messages; this is allowable since we are only discussing the number of queued messages.) Since the two numbers are independent, we have

$$Q(z) = Q_1(z)Q_2(z) \tag{6.21}$$

By definition, $Q_1(z)$ is given by

$$Q_1(z) = \frac{F_1(z,1,\ldots,1) - F_1(0,1,\ldots,1)}{z[1-F_1(0,1,\ldots,1)]} \tag{6.22a}$$

from which we have

$$Q_1^{(1)}(1) = \frac{\left[\frac{dF_1(z,1,\ldots,1)}{dz}\right]_{z=1}}{1-F_1(0,1,\ldots 1)} - 1 \tag{6.22b}$$

Next we see that $Q_2(z)$ is the same as the z-transform for the distribution of the number of customers left behind by a departing customer in a single M/G/1 queueing system; see for example [Klei75,Sec.5.6]. Thus, we have

$$Q_2(z) = \frac{(1-\lambda b)(z-1)B^*(\lambda-\lambda z)}{z-B^*(\lambda-\lambda z)} \tag{6.23a}$$

and

$$Q_2^{(1)}(1) = \frac{\lambda^2 b^{(2)}}{2(1-\lambda b)} + \lambda b \tag{6.23b}$$

The message waiting time can be found by (6.5) and

$$\lambda E[W] + \lambda b = Q^{(1)}(1) = Q_1^{(1)}(1) + Q_2^{(1)}(1) \tag{6.24}$$

Note that the service time for each station has the same distribution as that for a busy period in a single M/G/1 queue. Namely, it is given by (4.3a) and (4.3b) where the subscript i is dropped.

Our analysis to find $F_1(0,1,\ldots,1)$ and $[dF_1(z,1,\ldots,1)/dz]_{z=1}$ goes quite parallel to Section 6.1.2. First, $F_{i+1}(z_1,z_2,\ldots z_N)$ is expressed as

$$F_{i+1}(z_1,z_2,\ldots,z_N) = R^*[\sum_{j=1}^{N}(\lambda-\lambda z_j)].$$

$$.\{\Theta^*[\sum_{\substack{j=1\\(j\neq i)}}^{N}(\lambda-\lambda z_j)]\frac{1}{z_i}[F_i(z_1,z_2,\ldots,z_N)-F_i(z_1,\ldots,z_{i-1},0,z_{i+1},\ldots,z_N)]$$

$$+F_i(z_1,\ldots,z_{i-1},0,z_{i+1},\ldots,z_N)\}\quad i=1,2,\ldots,N \tag{6.25}$$

from which we have

$$F_1(0,1,\ldots,1) = \frac{1-N\lambda r-(N-1)\lambda\theta}{1-(N-1)\lambda\theta} \tag{6.26}$$

Relations (6.15) and (6.16) hold also in the decrementing service system. They lead to

$$\left[\frac{dF_1(0,z,\ldots,z)}{dz}\right]_{z=1} = \frac{N(N-1)\lambda r[\lambda^2\theta^{(2)}-(\lambda\theta)^2+1]}{2[1-N\lambda r-(N-1)\lambda\theta](1+\lambda\theta)}F_1(0,1,\ldots,1) \tag{6.27}$$

The mean number of messages at station 1 when it is polled is then given by

$$\left[\frac{dF_1(z,1,\ldots,1)}{dz}\right]_{z=1} = \frac{[1-(N-1)\lambda\theta]N\lambda^2(\delta^2+r^2)+2N\lambda r(1-N\lambda r)}{2[1-N\lambda r-(N-1)\lambda\theta]^2}$$

$$+\frac{N(N-1)\lambda r[\lambda^2\theta^{(2)}+2\lambda r\lambda\theta+\lambda r(\lambda\theta)^2+\lambda r-2\lambda\theta(1+\lambda\theta)]}{2[1-N\lambda r-(N-1)\lambda\theta]^2(1+\lambda\theta)}F_1(0,1,\ldots,1) \tag{6.28}$$

Substituting (6.26) and (6.28) into (6.22b), we get

$$Q_1^{(1)}(1) = \frac{[1-(N-1)\lambda\theta]\lambda(\delta^2+r^2)}{2[1-N\lambda r-(N-1)\lambda\theta]r} + \frac{(N-1)\lambda[\lambda\theta^{(2)}+r(1+\lambda\theta)^2]}{2[1-N\lambda r-(N-1)\lambda\theta](1+\lambda\theta)} \tag{6.29}$$

Thus, from (6.23b), (6.24) and (6.29), we have the mean message waiting time as

$$E[W] = \frac{[1-(N-1)\lambda\theta]\delta^2+Nr^2}{2[1-N\lambda r-(N-1)\lambda\theta]r} + \frac{(N-1)\lambda\theta^{(2)}}{2[1-N\lambda r-(N-1)\lambda\theta](1+\lambda\theta)} + \frac{\lambda b^{(2)}}{2(1-\lambda b)} \tag{6.30}$$

Comparing (6.30) with (4.33b) and (6.19), we get

$$E[W]\Big|_{exhaustive} \leq E[W]\Big|_{decrementing} \leq E[W]\Big|_{limited} \tag{6.31}$$

for the symmetric, continuous-time systems. In Table I, we show a numerical example for the mean message waiting times E[W] in various continuous time systems. In this example, E[W] (decrementing) $\leq$ E[W](gated) for small values of the total load λN, while E[W] (gated) $\leq$ E[W] (decrementing) for large values of λN.

A note for the mean cycle time. Since messages are found by polling with probability $1 - F_1(0,1,\ldots,1)$, and the mean duration of the service time for a station is given by θ, we have

$$E[C] = Nr + N\theta[1-F_1(0,1,\ldots,1)] \tag{6.32a}$$

Using (6.26), we get

$$E[C] = \frac{Nr}{1-N\lambda b} \tag{6.32b}$$

which is again identical to those for other service systems.

Table I. **Comparison of the Mean Message Waiting Times in the Exhaustive, Gated, Limited (K=1) and Decrementing Service Systems.**
(Parameters: N=10, r=0.1, δ^2=0.01, b=1, and $b^{(2)}$=1)

λN	exhaustive	gated	limited	decrementing
.00	.5500	.5500	.5500	.5500
.05	.6000	.6053	.6085	.6031
.10	.6556	.6667	.6742	.6628
.15	.7176	.7353	.7485	.7302
.20	.7857	.8125	.8333	.8070
.25	.8667	.9000	.9310	.8953
.30	.9571	1.000	1.045	.9980
.35	1.062	1.115	1.179	1.119
.40	1.183	1.250	1.339	1.263
.45	1.327	1.409	1.535	1.438
.50	1.500	1.600	1.778	1.655
.55	1.711	1.833	2.089	1.931
.60	1.975	2.125	2.500	2.294
.65	2.314	2.500	3.070	2.793
.70	2.767	3.000	3.913	3.523
.75	3.400	3.700	5.286	4.690
.80	4.350	4.750	7.917	6.858
.85	5.933	6.500	15.00	12.27
.90	9.100	10.00	100.0	50.00
.95	18.60	20.50	-	-

6.2 General, Continuous-Time System

We now proceed to consider a limited service, continuous-time service system such that at most K_i messages are served continuously at station i out of those found at the polling (i=1,2,...,N). All messages which arrive once the service has started are reserved for the next round. We employ the same assumptions and parameters for the message arrival process and reply intervals as Chapter 4 and Section 5.2. Thus, the special case of $K_i = \infty$ for all i reduces to the gated service, continuous-time system considered in Section 5.2.

For the limited service, continuous-time system, we have an explicit solution to the mean message waiting time only in a special case of statistically identical stations and $K_i=1$ for all i. This special case has been considered in Section 6.1. In a general case considered in this section, however, we can only find explicitly the mean number of messages served in a cycle, and the mean cycle, intervisit and service times. They are all shown to be independent of $\{K_j;\ 1 \le j \le N\}$ and even equal to those for the exhaustive service system, respectively.

We may refer to [Hash81b,Sec.5.3] for the analysis for mean waiting time in the case of $N=1$ (a single queue where at most $K>1$ messages are served at a time followed by a forced idle period). It gives the mean waiting time in terms of the K-1 roots (within a unit circle) of the equation

$$z^K - R^*(\lambda - \lambda z)[B^*(\lambda - \lambda z)]^K = 0 \qquad (6.33)$$

6.2.1 Number of Messages at Polling Instants

Due to the limited service discipline stated above, if there are $k < K_i$ messages are found at the polling of station i, then k messages are served; if K_i or more messages are found, then K_i messages are served. This scheme manifests itself in the following euqations for the joint GF for the number of messages found at the polling instants:

$$F_{i+1}(z_1,z_2,\ldots,z_N) = R_i^*[\sum_{j=1}^{N} (\lambda_j-\lambda_j z_j)] .$$

$$\cdot \sum_{k=0}^{K_i-1} (B_i^*[\sum_{j=1}^{N}(\lambda_j-\lambda_j z_j)])^k \left(\frac{1}{k!}\frac{\partial^k F_i(z_1,z_2,\ldots,z_N)}{\partial z_i^k}\right)_{z_i=0}$$

$$+ \left(\frac{1}{z_i} B^*[\sum_{j=1}^{N}(\lambda_j-\lambda_j z_j)]\right)^{K_i} \cdot F_i(z_i,z_2,\ldots,z_N)$$

$$- \sum_{k=0}^{K_i-1} z_i^k \left(\frac{1}{k!}\frac{\partial^k F_i(z_1,z_2,\ldots,z_N)}{\partial z_i^k}\right)_{z_i=0} . \qquad (6.34)$$

Using Taylor expansions

$$\sum_{k=0}^{K_i-1} z_i^k \left(\frac{1}{k!}\frac{\partial^k F_i(z_1,\ldots,z_N)}{\partial z_i^k}\right)_{z_i=0} = F_i(z_1,\ldots,z_N)$$

$$- \sum_{k=K_i}^{\infty} z_i^k \left(\frac{1}{k!}\frac{\partial^k F_i(z_1,\ldots,z_N)}{\partial z_i^k}\right)_{z_i=0} \qquad (6.35a)$$

and

$$\sum_{k=0}^{K_i-1} (B_i^*[\sum_{j=1}^{N}(\lambda_j-\lambda_j z_j)])^k \left(\frac{1}{k!}\frac{\partial^k F_i(z_1,\ldots,z_N)}{\partial z_i^k}\right)_{z_i=0}$$

$$= F_i(z_i,\ldots,z_{i-1}, B_i^*[\sum_{j=1}^{N}(\lambda_j-\lambda_j z_j)], z_{i+1},\ldots,z_N)$$

$$- \sum_{k=K_i}^{\infty} (B_i^*[\sum_{j=1}^{N}(\lambda_i-\lambda_j z_j)])^k \left(\frac{1}{k!}\frac{\partial^k F_i(z_1,\ldots,z_N)}{\partial z_i^k}\right)_{z_i=0} \qquad (6.35b)$$

we may rewrite (6.34) as

$$F_{i+1}(z_1,z_2,\ldots,z_N) = R_i^*[\sum_{j=1}^{N}(\lambda_j-\lambda_j z_j)] .$$

$$\cdot F_i(z_1,\ldots,z_{i-1},B_i^*[\sum_{j=1}^{N}(\lambda_j-\lambda_j z_j)], z_{i+1},\ldots, z_N)$$

$$- \sum_{k=K_i}^{\infty} \left(B_i^*[\sum_{j=1}^{N} (\lambda_j - \lambda_j z_j)] \right)^k \left(\frac{1}{k!} \frac{\partial^k F_i(z_1,z_2,\ldots,z_N)}{\partial z_i^k} \right)_{z_i=0}$$

$$+ \left(\frac{1}{z_i} B_i^*[\sum_{j=1}^{N} (\lambda_j - \lambda_j z_j)] \right)^{K_i} \sum_{k=K_i}^{\infty} z_i^k \left(\frac{1}{k!} \frac{\partial^k F_i(z_1,z_2,\ldots,z_N)}{\partial z_i^k} \right)_{z_i=0} \Bigg\} \tag{6.36}$$

We now define

$$\tilde{F}_i(z_1,z_2,\ldots,z_N) \triangleq \sum_{k=K_i}^{\infty} (z_i)^{k-K_i} \left(\frac{1}{k!} \frac{\partial^k F_i(z_1,z_2,\ldots,z_N)}{\partial z_i^k} \right)_{z_i=0}$$

$$= E \left[z_i^{-K_i} \prod_{j=1}^{N} z_j^{L_j(\tau_i(m))} \Big| L_i(\tau_i(m)) \geq K_i \right] \tag{6.37}$$

Then (6.36) can be written as

$$F_{i+1}(z_1,z_2,\ldots,z_N) = B_i^*[\sum_{j=1}^{N} (\lambda_j - \lambda_j z_j)] .$$

$$\cdot \; F_i(z_1,\ldots,z_{i-1},B^*[\sum_{j=1}^{N} (\lambda_j - \lambda_j z_j)],\; z_{i+1},\; \ldots,\; z_N)$$

$$+ \left(B_i^*[\sum_{j=1}^{N} (\lambda_j - \lambda_j z_j)] \right)^{K_i} \cdot \Big\{ \tilde{F}_i(z_1,z_2,\ldots,z_N)$$

$$- \tilde{F}_i(z_1,\ldots,z_{i-1},B_i^*[\sum_{j=1}^{N} (\lambda_j - \lambda_j z_j)],\; z_{i+1},\ldots,z_N \Big\} \tag{6.38}$$

It can be seen that (6.38) reduces to (5.29) when $K_i = \infty$ (in this case, $\tilde{F}_i(z_1,\ldots,z_N) \equiv 0$).

Let us use $f_i(j)$ and $f_i(j,k)$ defined by (3.27), and $\tilde{f}_i$ and $\tilde{f}_{ij}$ defined by

$$\tilde{f}_i \triangleq \frac{\partial \tilde{F}_i(z_1,z_2,\ldots,z_N)}{\partial z_i} \Bigg|_{z_1=z_2\ldots=z_N=1} .$$

$$\tilde{f}_{ij} \triangleq \frac{\partial^2 \tilde{F}_i(z_1,z_2,\ldots,z_N)}{\partial z_i \partial z_j} \Bigg|_{z_1=z_2\ldots=z_N=1} \tag{6.39}$$

Taking the first derivatives of (6.38), we get

$$f_{i+1}(i) = \tilde{f}_i + r_i\lambda_i + \rho_i\,[f_i(i) - \tilde{f}_i] \tag{6.40a}$$

$$f_{i+1}(j) = r_i\lambda_j + \lambda_j b_i\,[f_i(i) - \tilde{f}_i] + f_i(j) \qquad j\neq i \tag{6.40b}$$

We may solve (6.40a) and (6.40b) for $\{f_i(j) - \tilde{f}_i\}$ to obtain

$$E[L_i^*] = f_i(i) = \tilde{f}_i + \frac{\lambda_i \sum_{k=1}^{N} r_k}{1-\sum_{k=1}^{N}\rho_k} \tag{6.41a}$$

$$f_i(j) = \tilde{f}_i + \lambda_j\left[\sum_{k=j}^{i-1} r_k + \frac{\left(\sum_{k=j}^{i-1}\rho_k\right)\left(\sum_{k=1}^{N} r_k\right)}{1-\sum_{k=1}^{N}\rho_k}\right] \qquad j\neq i \tag{6.41b}$$

A set of equations for the second derivatives are given by

$$\begin{aligned} f_{i+1}(j,k) = {} & \lambda_j\lambda_k(\delta_i^{\,2}+r_i^{\,2}) + r_i\lambda_k f_i(j) + r_i\lambda_j f_i(k) \\ & + \lambda_j\lambda_k[f_i(i)-\tilde{f}_i](2r_ib_i+b_i^{(2)}) + f_i(j,k) \\ & + b_i\lambda_k[f_i(i,j)-\tilde{f}_{ij}] + b_i\lambda_j[f_i(i,k)-\tilde{f}_{ik}] \\ & + b_i^{\,2}\lambda_j\lambda_k[f_i(i,i)-\tilde{f}_{ii}-2K_i\tilde{f}_i] \qquad i\neq j,\ i\neq k \end{aligned} \tag{6.42a}$$

$$\begin{aligned} f_{i+1}(i,k) = {} & \lambda_i\lambda_k(\delta_i^{\,2}+r_i^{\,2})+r_i\lambda_i f_i(k)+\lambda_i\lambda_k[f_i(i)-\tilde{f}_i](2r_ib_i+b_i^{(2)}) \\ & + b_i\lambda_i[f_i(i,k)-\tilde{f}_{ik}]+b_i^{\,2}\lambda_i\lambda_k[f_i(i,i)-\tilde{f}_{ii}-2K_i\tilde{f}_i] \\ & + \tilde{f}_{ik} + K_i\lambda_k b_i\tilde{f}_i \qquad k\neq i \end{aligned} \tag{6.42b}$$

$$\begin{aligned} f_{i+1}(i,i) = {} & \lambda_i^{\,2}(\delta_i^{\,2}+r_i^{\,2})+\lambda_i^{\,2}[f_i(i)-\tilde{f}_i](2r_ib_i+b_i^{(2)}) \\ & +(\lambda_ib_i)^2[f_i(i,i)-\tilde{f}_{ii}-2K_i\tilde{f}_i]+\tilde{f}_{ii}+2K_i\lambda_ib_i\tilde{f}_i \end{aligned} \tag{6.42c}$$

Although these equations have forms similar to (5.32a-c), we have so far been unable to find any solution even in the case of identical stations.

6.2.2 Number of Messages Served in a Cycle

We define $F_i(z)$ by (3.21b) and $\tilde{F}_i(z)$ by

$$\tilde{F}_i(z) \triangleq \tilde{F}_i(1,\ldots,1,z,1,\ldots,1) \tag{6.43}$$

where z is the i th argument in $\tilde{F}_i(1,\ldots,1,z,1,\ldots,1)$. Since the number of messages served in the m th cycle at station i, $T_i(m)$, is equal to $L_i(\tau_i(m))$ if $L_i(\tau_i(m)) \le K_i-1$, and equal to K_i if $L_i(\tau_i(m) \ge K_i$, it follows that the GF for $T_i(m)$ is given by

$$E[z^{T_i(m)}] = \sum_{k=0}^{K_i-1} z^k \frac{1}{k!} \left[\frac{d^k F_i(z)}{dz^k}\right]_{z=0}$$

$$+ z^{K_i} \left[1-\sum_{k=0}^{K_i-1} \frac{1}{k!} \frac{d^k F_i(z)}{dz^k}\right]_{z=0} \tag{6.44a}$$

Using $\tilde{F}_i(z)$, this can be written as

$$E[z^{T_i(m)}] = F_i(z) + z^{K_i} [\tilde{F}_i(1) - \tilde{F}_i(z)] \tag{6.44b}$$

The mean number of messages served for station i in a cycle is thus given by

$$E[T_i(m)] = f_i(i) - \tilde{f}_i = \frac{\lambda_i \sum_{k=1}^{N} r_k}{1-\sum_{k=1}^{N} \rho_k} \tag{6.45}$$

We note that this is independent of K_j; $1 \le j \le N$; and even equal to that for the exhaustive service system; see (4.17a) and (5.37b).

6.2.3 Mean Intervisit and Cycle Times

In order to find the mean intervisit time $E[\tau_i(m+1)-\bar{\tau}_i(m)]$, let us use a relationship

$$L_i(\tau_i(m+1))=L_i(\bar{\tau}_i(m))+\text{number of arrivals in }[\bar{\tau}_i(m),\ \tau_i(m+1)] \tag{6.46a}$$

where we must note that the two quantities on the right-hand side are dependent. However, we can still argue about their means; thus, from (6.46a) we have

$$E[L_i^*] = E[L_i(\bar{\tau}_i(m))] + \lambda_i\ E[I_i] \tag{6.46b}$$

The GF for $L_i(\bar{\tau}_i(m))$, i.e., the number of messages at station i when the service at station i has finished is given by the second factor of (6.38) with $z_i = z$ and $z_j = 1$ $(j\neq i)$:

$$E[z^{L_i(\bar{\tau}_i(m))}] = F_i[B_i^*(\lambda_i-\lambda_i z)]$$

$$+ [B_i^*(\lambda_i-\lambda_i z)]^{K_i}\ \{\tilde{F}_i(z)-\tilde{F}_i[B_i^*(\lambda_i-\lambda_i z)]\} \tag{6.47a}$$

from which we have

$$E[L_i(\bar{\tau}_i(m))] = \rho_i E[L_i^*] + (1-\rho_i)\ \tilde{f}_i \tag{6.47b}$$

Substituting (6.47b) into (6.46b) and using (6.41a), we have the mean intervisit time

$$E[I_i] = \frac{(1-\rho_i)\sum_{k=1}^{N} r_k}{1-\sum_{k=1}^{N}\rho_k} \tag{6.48}$$

Again, this is independent of $\{K_j;\ 1\leq j\leq N\}$, and is identical to (4.20b) (exhaustive case) and (5.41) (gated case).

Since the mean service time for station i in a cycle is given by

$b_i E[T_i(m)]$, we have the mean cycle time

$$E[C_i] = E[I_i] + b_i E[T_i(m)] = \frac{\sum_{k=1}^{N} r_k}{1 - \sum_{k=1}^{N} \rho_k} \tag{6.49}$$

which is once again equal to that for the exhaustive service system in (4.23a).

6.2.4. Queue Length and Waiting Time

Here we derive formal expressions for the mean queue length and mean waiting time. The GF for the number of messages left at station i after the completion of each message service, $Q_i(z)$, can be expressed in the same way as in Section 5.2. From the second line of (5.42b), we have

$$Q_i(z) = \frac{1}{E[T_i(m)]} \cdot \frac{B_i^*(\lambda_i - \lambda_i z)}{z - B_i^*(\lambda_i - \lambda_i z)} \cdot E[z^{T_i(m)} - \{B_i^*(\lambda_i - \lambda_i z)\}^{T_i(m)}] \tag{6.50}$$

where the GF of $T_i(m)$ is now given by (6.44b). As before, we may think that (6.50) holds not only at message departure instants but also all other times. From (6.50), the mean queue length at station i is given by

$$E[L_i] = Q_i^{(1)}(1) = \rho_i + \frac{(1 - \sum_{k=1}^{N} \rho_k)(1 + \rho_i)[f_i(i,i) - \tilde{f}_{ii} - 2K_i \tilde{f}_i]}{2\lambda_i \sum_{k=1}^{N} r_k} \tag{6.51a}$$

Applying Little's result in the form of (4.34), we have the mean message waiting time

$$E[W_i] = \frac{(1 - \sum_{k=1}^{N} \rho_k)(1 - \rho_i)[f_i(i,i) - \tilde{f}_{ii} - 2K_i \tilde{f}_i]}{2\lambda_i^2 \sum_{k=1}^{N} r_k} \tag{6.51b}$$

6.3 Symmetric, Discrete-Time System

This section presents the discrete-time version of a model in Section 6.1.

Let us define our system specifically. We consider a system of N identical stations (indexed as i = 1, 2, ..., N) served by a single server. Each station has an infinite capacity buffer. The reference time is slotted with slot size equal to unity. The number of messages arriving at each station is assumed to be independent and identically distributed with $A(z)$, λ, and γ^2 the GF, mean and variance, respectively; $\lambda = A^{(1)}(1)$ and $\gamma^2 = A^{(2)}(1) + \lambda - \lambda^2$. We assume that message service times are integral multiples of the slot size. Let $B(z)$, b, and $b^{(2)}$ be the GF, mean and second moment, respectively, for the message service time; $b = B^{(1)}(1)$ and $b^{(2)} = B^{(2)}(1) + b$. The server inspects the stations in cyclic order of indices, and administers service to one message, if any, for a station. After serving at most one message for station i, it goes to inspect station i + 1 mod N with a generally distributed reply interval (also an integral multiple of the slot size). Let $R(z)$, r, and δ^2 be the GF, mean and variance, respectively, for the time needed by the server to switch from one station to another; $r = R^{(1)}(1)$ and $\delta^2 = R^{(2)}(1) + r - r^2$.

We define $F_i(z_1, z_2, \ldots, z_N)$ and $Q(z)$ as in (6.1) and (6.2).

It is clear that the number of messages left behind at station 1 when message service at station 1 is finished consists of two independent components. One is the number of messages at the polling instant minus one assuming that more than one messages are found at the polling. The other is the number of arrivals during the message service time whose GF is given by the compound distribution $B[A(z)]$. Since the two numbers are independent, we have

$$Q(z) = \frac{F_1(z,1,\ldots,1) - F_1(0,1,\ldots,1)}{z[1 - F_1(0,1,\ldots,1)]} B[A(z)] \tag{6.52}$$

from which we have (6.4).

Let $W(z)$ be the GF for the waiting time W of an arbitrarily chosen message (called a tagged message); $E[W] = W^{(1)}(1)$. We assume that the waiting time is counted from the slot following the arrival instant to the slot preceding the start of service. Note that the messages left behind at station 1 when the service of a tagged message at station 1 has been completed are those which arrive while the tagged message is in the system (waiting time plus service time), and those which arrived in the same slot as the tagged message but were placed behind the tagged message. The GF for the first number is given by

$$Q_1(z) = W[A(z)]B[A(z)] \tag{6.53a}$$

while that for the second number (which is the backward recurrence time in the renewal process where the interevent time distribution is given by $A(z)$) is given by (see (3.8a))

$$Q_2(z) = \frac{1 - A(z)}{a(1-z)} \tag{6.53b}$$

It is important to note that these two numbers are not independent since, e.g., a large group of arrivals in a slot implies both a long waiting time for an arbitrary message in the group and a large number of messages placed behind it. However we can still argue about the mean numbers, i.e.,

$$Q^{(1)}(1) = Q_1^{(1)}(1) + Q_2^{(1)}(1) \tag{6.54a}$$

where

$$Q_1^{(1)}(1) = \lambda(E[W] + b) \tag{6.54b}$$

and

$$Q_2^{(1)}(1) = \frac{\gamma^2 + \lambda^2 - \lambda}{2\lambda} \tag{6.54c}$$

From (6.4) and (6.54a-c), we have

$$E[W] = \frac{\left[\frac{dF_1(z,1,\ldots,1)}{dz}\right]_{z=1}}{\lambda[1 - F_1(0,1,\ldots,1)]} - \frac{1}{\lambda} - \frac{\gamma^2 + \lambda^2 - \lambda}{2\lambda^2} \tag{6.55}$$

We proceed to find $F_1(0, 1, \ldots, 1)$ and $[dF_1(z, 1, \ldots, 1)/dz]_{z=1}$ in the same way as in Section 6.1.2. First, we have

$$F_{i+1}(z_1, z_2, \ldots, z_N) = R[\prod_{j=1}^{N} A(z_j)].$$

$$\cdot \{B[\prod_{j=1}^{N} A(z_j)]\frac{1}{z_i}[F_i(z_1,z_2,\ldots,z_N)-F_i(z_1,\ldots,z_{i-1},0,z_{i+1},\ldots,z_N)]$$

$$+ F_i(z_1,\ldots,z_{i-1},0,z_{i+1},\ldots,z_N)\} \qquad i=1,2,\ldots,N \tag{6.56}$$

from which we again have (6.10). By using convention

$$R \equiv R[A(z)] \; ; \quad B \equiv B[A(z)] \tag{6.57}$$

we have exactly the same equations as (6.12), (6.13) and (6.14). We also have the relationships in (6.15) and (6.16). Thus we have

$$\left[\frac{dF_1(0,z,\ldots,z)}{dz}\right]_{z=1} = \frac{N(N-1)r\gamma^2 F_1(0,1,\ldots,1)}{2[1-N\lambda(r+b)]} \tag{6.58}$$

and then

$$\left[\frac{dF_1(z,1,\ldots,1)}{dz}\right]_{z=1} = \frac{NF_1(0,1,\ldots,1)}{2[1-N\lambda(r+b)]^2}\{\lambda r[1-N\lambda(r+b)] +$$

$$+ (1-N\lambda b)\lambda^2\delta^2+r[1-(N-1)\lambda b]\gamma^2- \lambda r\lambda b + N\lambda^3(br^2+rb^{(2)})\} \qquad (6.59)$$

Substituting (6.10) and (6.59) into (6.55), we get the mean message waiting time

$$E[W] = \frac{\delta^2}{2r} + \frac{N[\lambda b^{(2)} +r(1+\lambda b)+\lambda\delta^2]}{2[1-N\lambda(r+b)]} + \frac{(b+Nr)(\gamma^2-\lambda)}{2\lambda[1-N\lambda(r+b)]} - \frac{1}{2} \qquad (6.60)$$

which corresponds to (3.63b) in the exhaustive service ssytem, and to (5.23) in the gated service system. Thus, if $\gamma^2 \geq \lambda$ (e.g. $\gamma^2 = \lambda$ for Poisson arrival), we have inequalities

$$E[W]\big|_{exhaustive} \leq E[W]\big|_{gated} \leq [W]\big|_{limited} \qquad (6.61)$$

As before, the mean message waiting time at station 1, $E[W_1]$, when all traffic is generated at station 1 is easily obtained from (6.60) (see Section 3.6.1):

$$E[W_1] = \frac{\delta^2}{2r} + \frac{N[\lambda b^{(2)}+r(1+N\lambda b)+N\lambda\delta^2]}{2[1-N\lambda(Nr+b)]} + \frac{(b+Nr)(\gamma^2-\lambda)}{2\lambda[1-N\lambda(Nr+b)]} - \frac{1}{2} \qquad (6.62)$$

which corresponds to (3.64b) in the exhaustive service system, and to (5.25a) in the gated service system. If we assume that the distribution for the number of message arrival in Poisson ($\gamma^2 = \lambda$), we have

$$E[W]\big|_{limited} \leq E[W_1]\big|_{limited};\ E[W_1]\big|_{gated} \leq E[W_1]\big|_{limited} \qquad (6.63)$$

in addition to (6.61). It seems that $E[W_1]\big|_{gated}$ can be greater or smaller than $E[W]\big|_{limited}$.

In order to have the mean packet waiting time for a system of single-packet messages only, we only have to let $b = b^{(2)} = 1$ and replace λ and γ^2 by [illegible] and [illegible]^2, respectively, in (6.60). This conversion leads to

$$E[U] = \frac{\delta^2}{2r} + \frac{(1+Nr)\sigma^2 + N\mu^2\delta^2}{2\mu[1-N\mu(1+r)]} - 1 \qquad (6.64)$$

which corresponds to (3.61b) in the exhaustive service system, and to (5.21b) in the gated service system.

The correspondence between (6.19) and (6.60) should be clear; we get (6.19) if Poisson arrival is assumed ($\gamma^2 = \lambda$) and a half slot shrinks to zero in (6.60).

We finally note that the mean cycle time in the discrete-time system is also given by (6.32b).

7 Systems with Zero Reply Intervals

In this chapter we consider continuous-time systems of N nonidentical stations with zero reply intervals. The mean message waiting times in the case N=2 in the exhaustive and gated service systems have already been obtained in Sections 4.4 and 5.2, respectively, by taking the limit in the solutions for systems with general reply intervals. Note also that the mean waiting time in the case of identical stations should be identical to that in a single M/G/1 queueing system. Our presentation here follows the approach in [Coop69], [Coop70] and [Coop81,Sec.5.13] with simplification made possible by borrowing ideas from Chapters 4 and 5.

In polling systems with zero reply intervals, the server becomes idle when the whole system empties. As soon as a message arrives at some station, the server goes there to begin service immediately. The cycle times (in the sense in Chapters 4 and 5) during the system idle period are zero infinitely many times; thus the mean cycle time is zero, and so our previous approach (based on averaging over the cycle time) becomes unapplicable. This is why we need separate analysis for systems with zero reply intervals.

7.1 Exhaustive Service System

For system parameters, we use the same notations as given in the beginning of Chapter 4. Following [Coop69], let us introduce a set of <u>switch points</u>. Suppose that the system is idle and a message arrives at some station, say, station i, at time τ_0. The server immediately commences service at station i, and continues to serve messages at station i until the first time τ_1 at which station i becomes empty. If the system is not empty at τ_1, the server advances to station i+1. The server immediately starts work at station i+1 until the instant τ_2 at which station i+1 becomes empty (where $\tau_2=\tau_1$ if the server finds station i+1 empty), and continues on in this manner until for the first time, τ_n say, the server finishes serving a station and there are no messages waiting anywhere in the system. The process terminates at τ_n and is reinitiated by the next arrival. We call the points τ_1, .., τ_n switch points.

Note that τ_0 is not a switch point, whereas τ_n is a switch point. Successive switch points may occur simultaneously in time but are nevertheless considered distinct. We associate with each switch point a station at which the server has just completed its visit.

7.1.1 Number of Messages at Switch Points

Define P_i $(q_1, q_2, \ldots, q_{i-1}, 0, q_{i+1}, \ldots, q_N)$ as the joint probability that at an arbitrary switch point, the server has just completed a visit to station i, and q_j messages are waiting in station j (j=1, 2, ..., i-1, i+1, ..., N). This state (i: $q_1, \ldots, q_{i-1}, 0, q_{i+1}, \ldots, q_N$) can occur through the ~~following exhaustive and mutually~~ exclusive events:

(i) The server leaves station i-1 and finds $k_i \geq 1$ messages waiting for service in station i where it spends a length of time equal to k_i busy periods.

(ii) The server leaves station i-1 and finds $k_i=0$ messages waiting for service in station i, but at least one message waiting for service somewhere else in the system, so that the server then pass through station i in zero time.

(iii) The server leaves some station and finds no messages waiting anywhere in the system. With probability λ_i/λ_0, the next arrival (which reinitiates the process) occurs at station i, where the server then spends a single busy period. Here we have used

$$\lambda_0 = \sum_{i=1}^{N} \lambda_i \tag{7.1a}$$

These considerations lead to the following embedded (at the switch points) Markov chain probability state equations:

$$P_i(q_1, \ldots, q_{i-1}, 0, q_{i+1}, \ldots, q_N)$$

$$= \sum_{k_1=0}^{q_1} \cdots \sum_{k_{i-2}=0}^{q_{i-2}} \sum_{k_i=1}^{\infty} \sum_{k_{i+1}=0}^{q_{i+1}} \cdots \sum_{k_N=0}^{q_N} P_{i-1}(k_1,\ldots,k_{i-2},0,k_i,k_{i+1},\ldots,k_N)$$

$$\cdot \int_0^\infty \prod_{\substack{j=1\\(j\neq i-1,i)}}^{N} \frac{(\lambda_j t)^{q_j-k_j}}{(q_j-k_j)!} e^{-\lambda_j t} \cdot \frac{(\lambda_{i-1}t)^{q_{i-1}}}{q_{i-1}\,!} e^{-\lambda_{i-1}t}\ \theta_i^{*(k_i)}(t)dt$$

$$+ P_{i-1}(q_1,\ldots, q_{i-2}, 0, 0, q_{i+1},\ldots, q_N)(1-\delta(\sum_{\substack{j=1\\(j\neq i-1,i)}}^{N} q_j))\ \delta(q_{i-1})$$

$$+ \frac{\lambda_i}{\lambda_0} P(0) \int_0^\infty \prod_{\substack{j=1\\(j\neq i)}}^{N} \frac{(\lambda_j t)^{q_j}}{q_j\,!} e^{-\lambda_j t}\ \theta_i(t)dt \tag{7.2}$$

where

$$P(0) = \sum_{i=1}^{N} P_i(0, \ldots, 0) \tag{7.1b}$$

and

$$\delta(x) \triangleq \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{if } x \neq 0 \end{cases} \tag{7.3}$$

We have also used $\theta_i(t)$ as the pdf for the busy period at station i and $\theta_i^{*(k)}(t)$ as its k fold convolution. The normalization condition is given by

$$\sum_{i=1}^{N} \sum_{q_1=0}^{\infty} \cdots \sum_{q_{i-1}=0}^{\infty} \sum_{q_{i+1}=0}^{\infty} \cdots \sum_{q_N=0}^{\infty} P_i(q_1, \ldots, q_{i-1}, 0, q_{i+1}, \ldots, q_N) = 1 \tag{7.4a}$$

We next define the GF

$$H_i(z_1, \ldots, z_{i-1}, 0, z_{i+1}, \ldots, z_N)$$

$$\triangleq \sum_{q_1=0}^{\infty} \cdots \sum_{q_{i-1}=0}^{\infty} \sum_{q_{i+1}=0}^{\infty} \cdots \sum_{q_N=0}^{\infty} P_i(q_1, \ldots, q_{i-1}, 0, q_{i+1}, \ldots, q_N) \cdot \prod_{\substack{j=1 \\ (j \neq i)}}^{N} (z_j)^{q_j} \tag{7.5}$$

Note then that $H_i(1, \ldots, 1, 0, 1, \ldots, 1)$ is the probability that an arbitrary switch point is associated with station i, and the normalization condition is given by

$$\sum_{i=1}^{N} H_i(1, \ldots, 1, 0, 1, \ldots, 1) = 1 \tag{7.4b}$$

Let us substitute (7.2) into (7.5) to have

$$H_i(z_1, \ldots, z_{i-1}, 0, z_{i+1}, \ldots, z_N)$$

$$= \sum_{k_1=0}^{\infty} \cdots \sum_{k_{i-2}=0}^{\infty} \sum_{k_i=1}^{\infty} \sum_{k_{i+1}=0}^{\infty} \cdots \sum_{k_N=0}^{\infty} P_{i-1}(k_1, \ldots, k_{i-2}, 0, k_i, k_{i+1}, \ldots, k_N)$$

$$\cdot \int_0^{\infty} \prod_{\substack{j=1 \\ (j \neq i)}}^{N} e^{-\lambda_j(1-z_j)t} \, \theta_i^{*(k_i)}(t)dt \prod_{\substack{j=1 \\ (j \neq i, i-1)}}^{N} (z_j)^{k_j}$$

$$+ \sum_{q_1=0}^{\infty} \cdots \sum_{q_{i-2}=0}^{\infty} \sum_{q_{i+1}=0}^{\infty} \cdots \sum_{q_N=0}^{\infty} P_{i-1}(q_1, \ldots, q_{i-2}, 0, 0, q_{i+1}, \ldots, q_N)$$

$$\cdot (1-\delta(\sum_{\substack{j=1\\(j\neq i-1,i)}}^{N} q_j)) \prod_{\substack{j=1\\(j\neq i-1,i)}}^{N} (z_j)^{q_j} + \frac{\lambda_i}{\lambda_0} P(0) \int_0^\infty \prod_{\substack{j=1\\(j\neq i)}}^{N} e^{-\lambda_j(1-z_j)t} \theta_i(t)dt \qquad (7.6a)$$

The first term on the r.h.s. of (7.6a) is

$$H_{i-1}(z_1, \ldots, z_{i-2}, 0, \Theta_i^*[\sum_{\substack{j=1\\(j\neq i)}}^{N} (\lambda_j - \lambda_j z_j)], z_{i+1}, \ldots z_N)$$

$$- \sum_{k_1=0}^{\infty} \cdots \sum_{k_{i-2}=0}^{\infty} \sum_{k_{i+1}=0}^{\infty} \cdots \sum_{k_N=0}^{\infty} P_{i-1}(k_1, \ldots, k_{i-2}, 0, 0, k_{i+1}, \ldots, k_N)$$

$$\cdot \prod_{\substack{j=1\\(j\neq i,i-1)}}^{N} (z_j)^{k_j} \qquad (7.6b)$$

where $\Theta_i^*(s)$ is the Laplace transform of $\theta_i(t)$, and satisfies (4.3a). The second term on the r.h.s. of (7.6a) and the second term of (7.6b) cancel when any one of $q_1, \ldots, q_{i-2}, q_{i+1}, \ldots, q_N$ is positive. When all of them are zero, the former is zero and the latter remains to give $-P_i(0, \ldots, 0)$. Thus we obtain

$$H_i(z_1, \ldots, z_{i-1}, 0, z_{i+1}, \ldots, z_N)$$

$$= H_{i-1}(z_1, \ldots, z_{i-2}, 0, \Theta_i^*[\sum_{\substack{j=1\\(j\neq i)}}^{N} (\lambda_j - \lambda_j z_j)], z_{i+1}, \ldots, z_N)$$

$$- P_{i-1}(0, \ldots, 0) + \frac{\lambda_i}{\lambda_0} P(0) \Theta_i^* [\sum_{\substack{j=1\\(j\neq i)}}^{N} (\lambda_j - \lambda_j z_j)] \qquad (7.7)$$

We proceed to derive from (7.7) the equations for the moments defined by

$$h_i(j) = \frac{\partial H_i(z_1, \ldots, z_{i-1}, 0, z_{i+1}, \ldots, z_N)}{\partial z_j} \Big|_{z_1 = \cdots = z_N = 1}$$

$$h_i(j,k) = \frac{\lambda_0(1-\rho_0)}{P(0)} \left. \frac{\partial^2 H_i(z_1, \ldots, z_{i-1}, 0, z_{i+1}, \ldots, z_N)}{\partial z_j \partial z_k} \right|_{z_1 = \ldots = z_N = 1}$$

$$i, j, k = 1, 2, \ldots, N \quad (7.8)$$

where

$$\rho_0 = \sum_{i=1}^{N} \rho_i \tag{7.1c}$$

By differentiating (7.7) with respect to $z_j (j \neq i)$ and then setting $z_1 = \ldots = z_N = 1$, we have a set of equations for $\{h_i(j);\ i,j = 1, 2, \ldots, N\}$:

$$h_i(i-1) = h_{i-1}(i)\, \lambda_{i-1}\, \theta_i + \frac{\lambda_i}{\lambda_0} P(0)\, \lambda_{i-1}\, \theta_i \tag{7.9a}$$

$$h_i(j) = h_{i-1}(j) + h_{i-1}(i)\, \lambda_j \theta_i + \frac{\lambda_i}{\lambda_0} P(0)\, \lambda_j \theta_i \qquad j \neq i,\ i-1 \tag{7.9b}$$

where θ_i and $\theta_i^{(2)}$ (which appears later) are defined in (4.3b). To solve (7.9a) and (7.9b), add up their both sides over $i=j+1, j+2, \ldots, j-1$ to have

$$\sum_{i=j+1}^{j-1} h_i(j) = \sum_{i=j+2}^{j-1} h_{i-1}(j) + \lambda_j \sum_{i=j+1}^{j-1} h_{i-1}(i)\theta_i + \frac{\lambda_j}{\lambda_0} P(0) \sum_{i=j+1}^{j-1} \lambda_i \theta_i$$

or

$$h_{j-1}(j) = \lambda_j \sum_{i=j+1}^{j-1} h_{i-1}(i)\, \theta_i + \frac{\lambda_j}{\lambda_0} P(0) \sum_{i=j+1}^{j-1} \lambda_i \theta_i \tag{7.10a}$$

Adding $\lambda_j h_{j-1}(j)\theta_j$ to both sides of (7.10a), and then dividing them by λ_j, we have after rearrangement

$$\frac{1+\lambda_j\theta_j}{\lambda_j} h_{j-1}(j) - \frac{P(0)}{\lambda_0} \sum_{i=j+1}^{j-1} \lambda_i \theta_i = \sum_{i=j}^{j-1} h_{i-1}(i)\, \theta_i \tag{7.10b}$$

The sum on the r.h.s. of (7.10b) is independent of j. Hence we have N-1 equations for $\{h_{j-1}(j);\ j=1, 2, \ldots, N\}$. They and (7.10a) are satisfied by

$$h_{i-1}(i) = \frac{\lambda_i}{\lambda_0} P(0) \frac{\rho_0 - \rho_i}{1-\rho_0} \tag{7.11}$$

By differentiating (7.7) with respect to z_j and z_k and then setting $z_1 = \cdots = z_N = 1$ we have a set of equations for $\{h_i(j,k);\ i, j, k = 1, 2, \ldots, N\}$

$$h_i(j,k) = h_{i-1}(j,k) + h_{i-1}(i,j)\lambda_k\theta_i + h_{i-1}(i,k)\lambda_j\theta_i$$
$$+ h_{i-1}(i,i)\ \lambda_j\lambda_k\theta_i^{\ 2} + \lambda_j\lambda_k\lambda_i(1-\rho_i)\theta_i^{(2)}$$
$$j \neq i,\ i-1; \quad k \neq i,\ i-1 \tag{7.12a}$$

$$h_i(i-1,k) = h_{i-1}(i,k)\ \lambda_{i-1}\theta_i + h_{i-1}(i,i)\ \lambda_{i-1}\lambda_k\theta_i^{\ 2}$$
$$+ \lambda_{i-1}\lambda_k\lambda_i(1-\rho_i)\theta_i^{(2)} \qquad k \neq i,\ i-1 \tag{7.12b}$$

$$h_i(i-1,i-1) = h_{i-1}(i,i)(\lambda_{i-1}\theta_i)^2 + \lambda_{i-1}^{\ 2}\lambda_i(1-\rho_i)\theta_i^{(2)} \tag{7.12c}$$

where (7.11) has been used. We assume that (7.12a-c) are generally solved numerically. In the case $N=2$, only (7.12c) for $i=1$ and 2 may be solved to give

$$h_2(1,1) = \frac{\lambda_1^{\ 2}\lambda_2[\lambda_1\lambda_2(1-\rho_1)\theta_1^{(2)}\theta_2^{\ 2} + (1-\rho_2)\theta_2^{(2)}]}{1-(\lambda_1\lambda_2\theta_1\theta_2)^2}$$

$$= \lambda_1^{\ 2} \frac{\lambda_1\rho_2^{\ 2}b_1^{(2)} + \lambda_2(1-\rho_1)^2 b_2^{(2)}}{(1-\rho_1-\rho_2)(1-\rho_1-\rho_2+2\rho_1\rho_2)} \tag{7.13}$$

7.1.2 Message Waiting Time

Let us define as a <u>super cycle</u> the elapsed time between the arrival instant of a message at any station when the system is empty, and the first instant at which the system again becomes empty. Then messages that arrive at station i can be classified into the following two exclusive and exhaustive types:

(1) arrivals at station i that either initiate a super cycle or occur during the first busy period at station i generated by an arrival at station i which initiated a super cycle; or

(2) all other arrivals at station i (i.e., all arrivals that occur after (and including) the second busy period at station i in a super cycle).

Consider now waiting times of messages that arrive at station i. Those messages of type (1) are served during a busy period originated by a single message. Therefore, the LST $W_i^*(s|\text{ type } 1)$ of the waiting time distribution function for messages at station i of type (1) is given by the Pollaczek-Khinchin formula

$$W_i^*(s|\text{ type } 1) = \frac{s(1-\rho_i)}{s-\lambda_i+\lambda_i B_i^*(s)} \tag{7.14}$$

Messages of type (2) are served during a busy period originated by those messages waiting in station i when the server leaves station i-1. This means that at the end of intervisit time for station i, the server finds a number of messages whose GF is given by

$$I_i^*(\lambda_i-\lambda_i z) = \frac{H_{i-1}(1, \ldots, 1, 0, z, 1, \ldots, 1)}{H_{i-1}(1, \ldots, 1, 0, 1, 1, \ldots, 1)} \tag{7.15}$$

where 0 and z are the i-1 st and i th arguments, respectively, of H_{i-1}. Using (7.15) in (4.35a), the LST $W_i^*(s|\text{ type } 2)$ of the waiting time distribution function for messages of type (2) is given by

$W_i^*(s|\text{ type 2}) =$

$$= \frac{\lambda_i(1-\rho_i)[H_{i-1}(1, \ldots, 1, 0, 1, \ldots, 1)-H_{i-1}(1, \ldots, 1, 0, 1-s/\lambda_i, 1, \ldots, 1)]}{h_{i-1}(i)[s-\lambda_i+\lambda_i B_i^*(s)]}$$

(7.16)

We next find the probability P_i(type 1) that an arbitrary arrival at station i is of type (1). The mean number of messages that arrive at station i during an interval of length t is $\lambda_i t$. The probability that an arbitrary arrival at station i finds the system empty is $1-\rho_0$, so that $\lambda_i t(1-\rho_0)$ is the mean number of arrivals at station i that initiate a super cycle during any elapsed time t. The mean number of messages served in a busy period generated by each such arrival is $1/(1-\rho_i)$ (see (4.5b)), and hence the mean number of messages of type (1) served at station i during time t is $\lambda_i t(1-\rho_0)/(1-\rho_i)$. Thus we have

$$P_i(\text{type 1}) = \frac{\lambda_i t(1-\rho_0)/(1-\rho_i)}{\lambda_i t} = \frac{1-\rho_0}{1-\rho_i} \qquad (7.17a)$$

and the probability P_i(type 2) that an arbitrary arrival at station i is of type (2)

$$P_i(\text{type 2}) = 1-P_i(\text{type 1}) = \frac{\rho_0-\rho_i}{1-\rho_i} \qquad (7.17b)$$

The LST $W_i^*(s)$ of the waiting time distribution function is the weighted sum of the LSTs for both types:

$$W_i^*(s) = P_i(\text{type 1})\, W_i^*(s|\text{ type 1}) + P_2(\text{type 2})\, W_i^*(s|\text{ type 2})$$

$$= \frac{s(1-\rho_0)}{s-\lambda_i+\lambda_i B_i^*(s)} +$$

$$+\frac{\lambda_i(\rho_0-\rho_i)[H_{i-1}(1,\ldots,1,0,1,\ldots,1)-H_{i-1}(1,\ldots,1,0,1-s/\lambda_i,1,\ldots,1)]}{h_{i-1}(i)[s-\lambda_i+\lambda_i B_i^*(s)]}$$

(7.18)

From (7.18), we have the mean waiting time

$$E[W_i] = \frac{\lambda_i b_i^{(2)}}{2(1-\rho_i)} + \frac{h_{i-1}(i,i)}{2\lambda_i^2(1-\rho_i)} \tag{7.19}$$

where $h_{i-1}(i,i)$ is to be determined by solving (7.12a-c). In the case $N=2$, by using (7.13) in (7.19) we recover (4.36b).

7.2 Gated Service System

In a gated service system with zero reply intervals, only those messages that find the system empty upon arrival are classified into type (1), and their waiting time is zero. We thus immeidately have

$$P_i(\text{type } 1) = 1-\rho_0 \quad ; \quad W_i^*(s \mid \text{type } 1) = 1 \tag{7.20}$$

To find $W_i^*(s \mid \text{type } 2)$, we begin as before by considering the number of messages at switch points.

Define $P_i(q_i, q_2, \ldots, q_N)$ as the joint probability that at an arbitrary switch point, the server has just left station i, and q_j messages are waiting in station j (j=1, 2, ..., N). Denote by $b_i(t)$ as the pdf for the service time at station i and $b_i^{*(k)}(t)$ as its k fold convolution. Then, for a gated service system, we have

$P_i(q_1, \ldots, q_{i-1}, q_i, q_{i+1}, \ldots, q_N)$

$$= \sum_{k_1=0}^{q_1} \cdots \sum_{k_{i-1}=0}^{q_{i-1}} \sum_{k_i=1}^{\infty} \sum_{k_{i+1}=0}^{q_{i+1}} \cdots \sum_{k_N=0}^{q_N} P_{i-1}(k_1, \ldots, k_{i-1}, k_i, k_{i+1}, \ldots, k_N)$$

$$\cdot \int_0^\infty \prod_{\substack{j=1 \\ (j\neq i)}}^{N} \frac{(\lambda_j t)^{q_j - k_j}}{(q_j - k_j)!} e^{-\lambda_j t} \frac{(\lambda_i t)^{q_i t}}{q_i!} e^{-\lambda_i t} b_i^{*(k_i)}(t)dt$$

$$+ P_{i-1}(q_1, \ldots, q_{i-1}, 0, q_{i+1}, \ldots, q_N)(1-\delta(\sum_{\substack{j=1 \\ (j\neq i)}}^{N} q_j))\ \delta(q_i)$$

$$+ \frac{\lambda_i}{\lambda_0} P(0) \int_0^\infty \prod_{j=1}^{N} \frac{(\lambda_j t)^{q_j}}{q_j!} e^{-\lambda_j t} b_i(t)dt \tag{7.21}$$

We define the GF

$$H_i(z_i, \ldots, z_N) = \sum_{q_1=0}^{\infty} \cdots \sum_{q_N=0}^{\infty} P_i(q_1, \ldots, q_N) \prod_{j=1}^{N} (z_j)^{q_j} \tag{7.22}$$

and obtain the equation

$$H_i(z_1, \ldots, z_N) = H_{i-1}(z_1, \ldots, z_{i-1}, B_i^*[\sum_{j=1}^{N} (\lambda_j - \lambda_j z_j)], z_{i+1}, \ldots, z_N)$$

$$- P_i(0, \ldots, 0) + \frac{\lambda_i}{\lambda_0} P(0)\ B_i^*[\sum_{j=1}^{N} (\lambda_j - \lambda_j z_j)] \tag{7.23}$$

The moments are defined by

$$h_i(j) = \frac{\partial H_i(z_1, \ldots, z_N)}{\partial z_j} \Bigg|_{z_1 = \ldots = z_N = 1}$$

$$h_i(j,k) = \frac{\lambda_0(1-\rho_0)}{P(0)} \frac{\partial^2 H_i(z_1, \ldots, z_N)}{\partial z_j \partial z_k} \Bigg|_{z_1 = \ldots = z_N = 1}$$

$$i,j,k = 1, 2, \ldots, N \tag{7.24}$$

From the first derivatives of (7.23), we have equations for $\{h_i(j);\ i,j = 1, 2, \ldots, N\}$

$$h_i(i) = h_{i-1}(i)\lambda_i b_i + \frac{\lambda_i}{\lambda_0} P(0)\ \lambda_i b_i \tag{7.25a}$$

$$h_i(j) = h_{i-1}(j) + h_{i-1}(i)\lambda_j b_i + \frac{\lambda_i}{\lambda_0} P(0)\ \lambda_j b_i \qquad j \neq i \tag{7.25b}$$

from which we get

$$h_{i-1}(i) = \frac{\lambda_i}{\lambda_0} P(0)\ \frac{\rho_0}{1-\rho_0} \tag{7.26}$$

From the second derivatives of (7.23), using (7.26) we have equations for $\{h_i(j,k);\ i,j,k = 1, 2, \ldots, N\}$

$$h_i(j,k) = h_{i-1}(j,k) + h_{i-1}(i,j)\ \lambda_k b_i + h_{i-1}(i,k)\ \lambda_j b_i + h_{i-1}(i,i)\ \lambda_j \lambda_k b_i^2 + \lambda_j \lambda_k \lambda_i b_i^{(2)} \qquad j \neq i,\ k \neq i \tag{7.27a}$$

$$h_i(i,k) = h_{i-1}(i,k)\ \lambda_i b_i + h_{i-1}(i,i)\ \lambda_i \lambda_k b_i^2 + \lambda_i^2 \lambda_k b_i^{(2)} \qquad k \neq i \tag{7.27b}$$

$$h_i(i,i) = h_{i-1}(i,i)(\lambda_i b_i)^2 + \lambda_i^3 b_i^{(2)} \tag{7.27c}$$

Although (7.27a-c) must be solved numerically, we have an explicit solution in the case $N=2$.

$$h_2(1,1) = \frac{\lambda_1^2[\lambda_1(1+2\rho_2-\rho_1\rho_2-2\rho_1\rho_2^2-2\rho_2^3)b_1^{(2)}+\lambda_2(1+\rho_1\rho_2)b_2^{(2)}]}{(1-\rho_1\rho_2)(1-\rho_1-\rho_2)(1+\rho_1+\rho_2+2\rho_1\rho_2)} \tag{7.28}$$

We now turn our attention to $W_i^*(s \mid \text{type } 2)$. Since the number of messages found by the server at station i is the number of arrivals during the cycle time, we have

$$C_i^*(\lambda_i-\lambda_i z) = \frac{H_{i-1}(1, \ldots, 1, z, 1, \ldots, 1)}{H_{i-1}(1, \ldots, 1, 1, 1, \ldots, 1)} \tag{7.29}$$

where z appears in the i th position of H_{i-1}. Using (7.29) in (5.47a), we get

$$W_i^*(s|\text{ type 2}) = \frac{\lambda_i}{h_{i-1}(i)} \cdot \frac{H_{i-1}(1,\ldots, 1, B_i^*(s), 1,\ldots, 1) - H_{i-1}(1,\ldots, 1, 1-s/\lambda_i, 1,\ldots, 1)}{s-\lambda_i+\lambda_i B_i^*(s)} \tag{7.30}$$

From (7.20) and (7.30), by unconditioning we get

$$W_i^*(s) = 1 - \rho_0 + \frac{\lambda_i\rho_0[H_{i-1}(1,\ldots, 1, B_i^*(s), 1,\ldots, 1) - H_{i-1}(1,\ldots, 1, 1-s/\lambda_i, 1,\ldots, 1)]}{h_{i-1}(i)[s-\lambda_i+\lambda_i B_i^*(s)]} \tag{7.31}$$

This form is given in [Coop81,Prob.5.31]. The mean message waiting time is given by

$$E[W_i] = \frac{(1+\rho_i)\rho_0}{2h_{i-1}(i)\lambda_i} \cdot \frac{P(0)}{\rho_0(1-\rho_0)} h_{i-1}(i,i) \tag{7.32a}$$

Using (7.26), we get

$$E[W_i] = \frac{(1+\rho_i)h_{i-1}(i,i)}{2\lambda_i^2} \tag{7.32b}$$

We note that, in the case M=2, (7.28) and (7.32b) recover (5.56b).

8 Applications

In this chapter, we look at some examples (in computer communication networks) to which the above analysis results for the polling system are applied. Other work referring to or related to the polling system analysis includes (i) polling system with ARQ (automatic repeat request) retransmission policy [Kueh81] (the case where service may be unsuccessful), (ii) SCAN disk access policy [Coff82] and [Swar82] (an example where stations are polled in non-cyclic order), (iii) polling with priorities [Manf83], (iv) polling system with multiple servers [Morr84], (v) finite capacity polling system [Tran84], (vi) continuous polling system [Coff84], (vii) single and multiple Newhall loops [Cars77] and [Cars78], (viii) Prioritized-Round-Robin (PRR) and Fair-Round-Robin (FRR) channel access protocols for cable or radio networks [Gold83], and (ix) Fasnet [Heym83]. Expressnet and Fasnet are also modeled in [Toba83b]..

8.1 Polling in Wide-Area Networks

According to books such as [Mart67,Chap.20] and [Schw77,Chap.12], there are two types of polling discipline that have been most commonly adopted in practice. In the roll-call polling system, each station is in turn interrogated by the central processor. On the arrival of a polling message the station transmits all messages waiting to the central processor. Upon completion of message transmission, polling of the next station is initiated. In multi-drop network configuration, the roll-call polling is used. Another type, hub polling, is easier to implement in the loop configuration such that all the stations are serially connected to the central processor. The central processor initiates polling by interrogating the station at the end of the loop. This station transmits its waiting messages if any to which it appends a polling message for the next upstream station. The latter station similarly adds its messages followed by a polling message. In this way, at the completion of the polling cycle, with all stations interrogated, the central processor regains control.

In [Schw77,Chap.12], the reply intervals in both types of polling are assessed as follows. Let $\bar{R}$ be the sum of reply intervals (assumed to be constant) for all stations. In other words, $\bar{R}$ is the time needed for the polling message to circulate through all stations. In the hub polling, a delay is needed at each station to read, recognize, and add the polling message. If time C is needed at each station, and Y is needed for the round-trip propagation delay for the loop, we have

$$\bar{R}_{hub} = NC + Y \tag{8.1}$$

In the roll-call polling, the central processor polls each station one by one. Let P be the time needed to transmit a polling message. Then we have

$$\bar{R}_{roll\text{-}call} = NP + NC + Y' \tag{8.2}$$

where Y' is the total propagation delay incurred, which depends on the network configuration. It is generally greater than Y, the loop propagation delay.

From (8.1) and (8.2), it is the presence of the additional polling message delay NP plus the greater propagation delay that accounts for the increased time delay of roll-call polling system.

8.2 Polling in Packet Radio System

[Toba76] considers the performance of (roll-call) polling technique in the data flow from a population of packet radio terminals to the central station. After the completion of data message transmissions from a terminal, the central station sends a short polling message to the next terminal in sequence. The polling message transmission time is assumed to be a constant T_p seconds, and the one-way (electromagnetic) wave propagation delay is given by τ (also a constant) seconds. Let the time slot size be τ. Then the reply interval is given by

$$r = (2\tau + T_p)/\tau = 2 + T_p/\tau, \qquad \delta^2 = 0 \tag{8.3}$$

(Acknowledgment transmission time is neglected.) Data messages, of constant length T_m seconds (so, 1 message = (T_m/τ) packets), are assumed to have Poisson arrival with λ messages per second (so $\lambda\tau$ messages per slot). Thus, the packet arrival parameters are given by

$$\mu = \lambda T_m, \qquad \sigma^2 = \lambda T_m^2/\tau = \mu T_m/\tau \tag{8.4}$$

We measure the propagation delay by T_m:

$$a \triangleq \tau/T_m \tag{8.5}$$

and define the total throughput (time fraction spent for message transmission) by

$$S = \mu N \tag{8.6}$$

[Toba76] uses (3.61a) to evaluate the average data message delay (including the transmission time) in seconds

$$D_{polling} = T_m + \tau\left[\frac{\delta^2}{2r} + \frac{\sigma^2 N}{2(1-N\mu)} + \frac{1-\mu}{2} + \frac{Nr(1-\mu)}{2(1-N\mu)}\right] \tag{8.7a}$$

Substituting (8.3) through (8.6) into (8.7a), it gets

$$D_{polling}/T_m = 1 + \frac{S}{2(1-S)} + \frac{a}{2}\left(1 - \frac{S}{N}\right)\left(1 + \frac{Nr}{1-S}\right) \tag{8.7b}$$

However, as we have shown in Section 3.6.1, we must use (3.63b) to compute the mean waiting time for an arbitrary message. Using $\lambda\tau$ in palce of λ, $\gamma^2 = \lambda$ (Poisson arrival), $b = T_m/\tau$, $b^{(2)} = (T_m/\tau)^2$ in (3.63b), we have

$$D_{polling} = T_m + \tau\left\{ \frac{N[\lambda T_m^{\ 2}/\tau + r(1-\lambda T_m)]}{2(1 - N\lambda T_m)} - \frac{1}{2} \right\} \tag{8.8a}$$

In terms of a and S defined above, this is written as

$$D_{polling}/T_m = 1 + \frac{S}{2(1-S)} + \frac{a}{2}\left(1 - \frac{S}{N}\right)\left(\frac{Nr}{1-S}\right) - \frac{a}{2} \tag{8.8b}$$

where r is given by (8.3).

If we adopt a continuous-time model, we can use (4.33b) to have

$$D_{polling}/T_m = 1 + \frac{S}{2(1-S)} + \frac{a}{2}\left(1 - \frac{S}{N}\right)\left(\frac{Nr}{1-S}\right) \tag{8.8c}$$

8.3 MSAP (Mini-Slotted Alternating Priorities)

[Klei80] proposes MSAP scheme for channel multi-access in a population of packet radio terminals. This scheme takes advantage of the broadcasting nature of the messages. Specifically, each terminal can know its turn to transmit after the channel idle time of 1 slot allotted to the preceding terminal. Thus we have the reply interval

$$r = 1, \quad \delta^2 = 0 \tag{8.9}$$

Hence, in [Klei80] the average delay for MSAP is given by (using(8.9) in (8.7b))

$$D_{MSAP}/T_m = 1 + \frac{S}{2(1-S)} + \frac{a}{2}\left(1 - \frac{S}{N}\right)\left(1 + \frac{N}{1-S}\right) \qquad (8.10a)$$

which is smaller than (8.7b). However, for the same reason as before, this must be replaced by

$$D_{MSAP}/T_m = 1 + \frac{S}{2(1-S)} + \frac{a}{2}\left(1 - \frac{S}{N}\right)\left(\frac{N}{1-S}\right) - \frac{a}{2} \qquad (8.10b)$$

which is smaller than (8.8b).

MSAP can also be modeled in the framework of the one-message buffer system treated in Chapter 2. Using the constant message length T_m and propagation delay τ, we only have to replace

$$\tau_j = \bar{R} + jb - N\tau + jT_m \qquad (8.11)$$

in expressions in Section 2.4. The capacity (maximum throughput) of MSAP in the one-message buffer model in given by

$$C(N,a) = \frac{NT_m}{N\tau + NT_m} = \frac{1}{1 + a} \qquad (8.12)$$

8.4 UBS-RR and Expressnet

UBS-RR (Unidirectional Broadcast System - Round Robin) achieves broadcast communication by folding the unidirectional transmission cable to provide inbound and outbound channels for each station [Toba83a]. See Figure 7a. Let τ be the propagation delay between two extreme stations 1 and N, and T_m be the constant message transmission time. After the ongoing transmission period, each station is allotted a slot of length $2\tau/N$ to send the reservation burst sequentially (in round-robin order). After hearing its own burst in the inbound channel, it can start transmission. Thus, for the one-message buffer model of Chapter 2, we have substitution

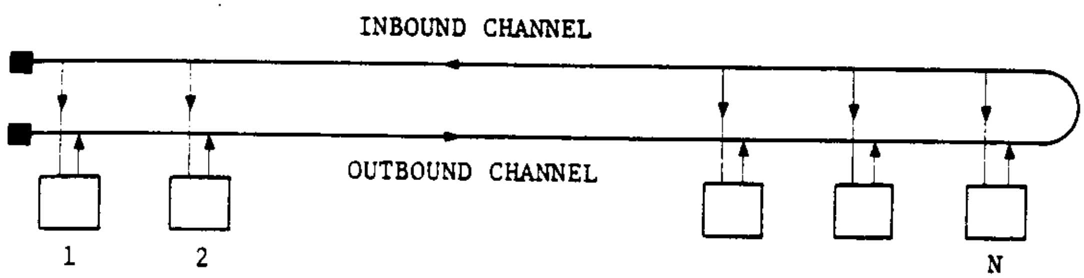

Figure 7 a. Topology of UBS-RR.

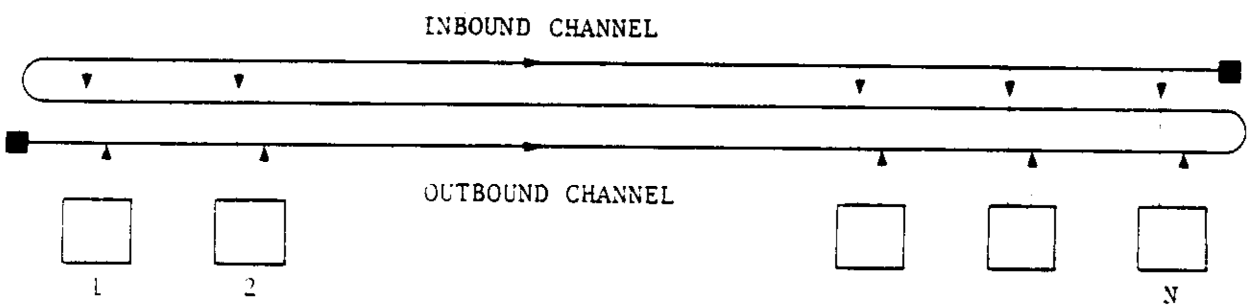

Figure 7 b. Topology of Expressnet.

$$\tau_j = 2\tau + j(T_m + 2\tau) \tag{8.13}$$

The capacity of UBS-RR is thus given by

$$C(N,a) = \frac{NT_m}{2\tau + N(T_m+2\tau)} = \frac{1}{1 + 2a + 2a/N} \tag{8.14}$$

Expressnet is an improved version of UBS-RR, now configured as in Figure 7b [Toba83a]. The data message, of constant length T_m, has a preamble, of constant length t_d, for synchronization and collision detection. Let τ be the end-to-end propagation delay on the inbound or outbound channel. The propagation delay along the connecting cable is denoted by τ_c. Assume that each station can have at most one outstanding message. From nonempty stations, data messages are transmitted into the inbound channel interleaved with preambles (this succession is called a train.). After the propagation delay of length $\tau + \tau_c$ (during this period data messages are received through the outbound channel) plus the final t_d (to ensure the end of train), the transmission of the next train becomes possible. Thus, application of our one-message buffer model in Chapter 2 can be attained by substitution

$$\tau_j = \bar{R} + jb + \tau + \tau_c + 2t_d + j(T_m + t_d) \tag{8.15}$$

The capacity of Expressnet is given by

$$C(N,a) = \frac{NT_m}{\tau + \tau_c + 2t_d + N(T_m + t_d)} \tag{8.16a}$$

Neglecting t_d in comparison to T_m and τ, and taking $\tau_c = \tau$, (8.16a) is rewritten in terms of $a = \tau/T_m$ as

$$C(N,a) = \frac{1}{1 + 2a/N} \tag{8.16b}$$

Comparing (8.12), (8.14) and (8.16b), we see that the capacity of Expressnet is greater than those of MSAP and UBS-RR.

8.5 Token Ring and Token Bus

In the performance comparison of channel access schemes in local-area networks, [Bux81] applies (4.33b) to the token ring scheme where a control token for transmission right is circulated in a loop of stations. In Bux's notation

S = number of stations λ = total message arrival rate

T_p = message transmission time $\rho = \lambda E[T_p]$ channel utilization

τ = ring round-trip delay

Note that τ is given by

$$\tau(\text{sec}) = P(\text{sec/km}) \times \ell(\text{km}) + N \times L(\text{bits}) \times C^{-1}(\text{sec/bit}) \qquad (8.17)$$

where P = signal propagation delay (e.g., 5×10^{-6} sec/km), ℓ = length of the ring, L_R = bit latency at each station, and C (bits/sec) = ring speed. Thus, by substitution

$$N \to S,\ r \to \frac{\tau}{S},\ \delta^2 \to 0,\ \lambda \to \frac{\lambda}{S},\ b^{(2)} \to E[T_p{}^2],\ \rho \to \frac{\lambda}{S}E[T_p] = \frac{\rho}{S} \qquad (8.18)$$

in (4.33b), we get

$$E[W] = \frac{\tau(1-\rho/S)}{2(1-\rho)} + \frac{\rho E[T_p{}^2]}{2(1-\rho)E[T_p]} \qquad (8.19)$$

Adding (8.19) by the mean message length $E[T_p]$ and the mean propagation delay to the destination (assumed to be uniformly distributed over the loop) $\tau/2$, we have the mean message transfer time t_f as

$$t_f = \frac{E[T_p{}^2]}{2(1-\rho)E[T_p]} + E[T_p] + \frac{\tau(1-\rho/S)}{2(1-\rho)} + \frac{\tau}{2} \qquad (8.20a)$$

Note that (8.20a) assumes the exhaustive service; for other types of services we have (from (5.46b) and (6.19))

$$t_f\big|_{gated} = \frac{\rho E[T_p^{\,2}]}{2(1-\rho)E[T_p]} + E[T_p] + \frac{\tau(1+\rho/S)}{2(1-\rho)} + \frac{\tau}{2} \qquad (8.20b)$$

$$t_f\big|_{limited} = \frac{\rho E[T_p^{\,2}]}{2(1-\rho-\lambda\tau)E[T_p]} + E[T_p] + \frac{\tau(1+\rho/S)}{(1-\rho-\lambda\tau)} + \frac{\tau}{2} \qquad (8.20c)$$

The results in (8.20a) and (8.20b) are also given in [DeMo84] with different notation.

In the token bus system analyzed in [DeMo84], stations connected via bidirectional medium consistute a logical ring such that station i + 1 is not necessarily next to station i. Due to broadcast transmission, τ for the token bus is given by

$$\tau(\text{sec}) = N \times P(\text{sec/km}) \times \ell(\text{km}) + N \times L_B(\text{bits}) \times C^{-1}(\text{sec/bit}) \qquad (8.21)$$

The mean message transfer times have expressions similar to (8.20a-c) with the last term $\tau/2$ replaced by $P\ell/2$.

we have (from (5.46b) and (6.19))

$$t_f\Big|_{gated} = \frac{\rho E[T_p^2]}{2(1-\rho)E[T_p]} + E[T_p] + \frac{\tau(1+\rho/S)}{2(1-\rho)} + \frac{\tau}{2} \qquad (8.20b)$$

$$t_f\Big|_{limited} = \frac{\rho E[T_p^2]}{2(1-\rho-\lambda\tau)E[T_p]} + E[T_p] + \frac{\tau(1+\rho/S)}{(1-\rho-\lambda\tau)} + \frac{\tau}{2} \qquad (8.20c)$$

The results in (8.20a) and (8.20b) are also given in [DeMo84] with different notation.

In the token bus system analyzed in [DeMo84], stations connected via bidirectional medium consistute a logical ring such that station i + 1 is not necessarily next to station i. Due to broadcast transmission, τ for the token bus is given by

$$\tau(\text{sec}) = N \times P(\text{sec/km}) \times \ell(\text{km}) + N \times L_B(\text{bits}) \times C^{-1}(\text{sec/bit}) \qquad (8.21)$$

The mean message transfer times have expressions similar to (8.20a-c) with the last term $\tau/2$ replaced by $P\ell/2$.

9 Future Research Topics

Let us make clear what has been solved and what has not in the presentation of Chapters 2 through 7. (By "solved" we mean that a set of linear simultaneous equations are explicitly derived in terms of given system parameters to find the mean message or packet waitng time which we think is the single most important performance measure [Klei76, p.161].)

For the one-message buffer system (Chapter 2), the case of constant message service time and constant reply intervals is explicitly solved by (2.29b), (2.31b) and (2.44a). Even the distribution of the waiting time is explicitly given in (2.42). However, when the message service time and/or reply interval are generally distributed (even if the distribution functions are identical for all stations), we have to solve $O(2^N)$ equations. We note that a model in [Coff84] (continuous polling system) has the same problem.

For the infinite buffer systems with exhaustive and gated service (Chapters 3 through 5), we have derived $O(N^2)$ equations in continuous-time systems ((4.51a-d) and (5.54a-d)), and $O(N^3)$ equations in discrete-time systems ((3.31a-d) and (5.5a-d)). Thus we think that these are basically solved. To find a set of $O(N\ )$ equations for discrete-time systems will be a good exercise. We have been unable to obtain the distribution for the intervisit time in gated service systems although the mean intervisit times are given in (5.11) and (5.41).

For the infinite buffer system with limited service (Chapter 6), the only case for which we can ever calculate the mean waiting time is the case of identical stations with "limited-to-one" service. We do not have an ex-

9 Future Research Topics

Let us make clear what has been solved and what has not in the presentation of Chapters 2 through 7. (By "solved" we mean that a set of linear simultaneous equations are explicitly derived in terms of given system parameters to find the mean message or packet waitng time which we think is the single most important performance measure [Klei76, p.161].)

For the one-message buffer system (Chapter 2), the case of constant message service time and constant reply intervals is explicitly solved by (2.29b), (2.31b) and (2.44a). Even the distribution of the waiting time is explicitly given in (2.42). However, when the message service time and/or reply interval are generally distributed (even if the distribution functions are identical for all stations), we have to solve $O(2^N)$ equations. We note that a model in [Coff84] (continuous polling system) has the same problem.

For the infinite buffer systems with exhaustive and gated service (Chapters 3 through 5), we have derived $O(N^2)$ equations in continuous-time systems ((4.51a-d) and (5.54a-d)), and $O(N^3)$ equations in discrete-time systems ((3.31a-d) and (5.5a-d)). Thus we think that these are basically solved. To find a set of O (N[illegible]) equations for discrete-time systems will be a good exercise. We have been unable to obtain the distribution for the intervisit time in gated service systems although the mean intervisit times are given in (5.11) and (5.41).

For the infinite buffer system with limited service (Chapter 6), the only case for which we can ever calculate the mean waiting time is the case of identical stations with "limited-to-one" service. We do not have an explicit solution to any asymmetric system, even the case of N = 2 stations considered in [Eise79], [Iisa81a] and [Boxm84]. We feel the need of some breakthrough if we try to solve this case exactly.

For the systems with zero reply intervals (Chapter 7), we have shown O (N^3) equations in (7.12a-c) and (7.27a-c) for exhaustive and gated services, respectively. [Coop70] shows it possible to have O (N^2) equations for the exhaustive service system. No attempt has been made to consider the limited service system with zero intervals exactly for N>2.

In this monograph, we have dealt with only exact approach. However, there are many practical problems which need approximate treatment. Let us indicate some of them in the area of computer communication networks. First, any communication buffer is finite; so a work such as [Tran84] seems useful. As another example, to integrate the data and real-time voice communication on the token ring, a priority mechanism has been purposed [Bux83b]. To consider the performance of prioritized token ring, we may need analysis of polling system with priority such as [Manf83]. Also, the token ring draft standards (e.g., [IEEE83]) recommend to set the maximum message length that each station can transmit at one time. This problem is formulated as a limited service polling system where at most K packets can be served in each cycle. When the maximum transmission time from each station is limited, a data file must be blocked for transfer (a block is the unit of transmission). Although we can formulate this case as a "limited-to-one" service polling system (viewing a block as a fixed-size message) of Section 6.1, a user is interested not in the mean block transfer time but in the mean file transfer time. Note that a file transfer is interleaved with transmissions from other stations.

We have no analysis for such systems so far.

The mean waiting time has been our major performance criterion. The mean cycle time has shown to be identical for all service systems considered in this report. We finally propose to take the fairness among stations as another performance measure. Although the limited service system is said to be fair (as compaired to the exhaustive and gated service systems) in the sense that it prevents heavily loaded stations from monopolyzing the service, no attempts seem to have appeared to quantify this claim.

The analysis of polling systems is an intriguing application of probabilistic thinking with elegant mathematical manipulation. We hope that this writing has presented an organized view of the state-of-the-art and stimulated further research.

Appendix

A: Derivation of (3.29a), (3.29b), (3.30a) and (3.30b)

In the case of exhaustive, discrete-time system, we derive from (3.26) a set of equations for $\{f_i(j)\}$ and $\{f_i(j,k)\}$ $(i,j,k = 1,2,\ldots, N)$ defined by (3.27), and solve them. The equations (3.26) is reproduced here with index change $j \to n$:

$$F_{i+1}(z_1,\ldots,z_N) = R_i[\prod_{n=1}^{N} P_n(z_n)] . F_i(z_1,\ldots,z_{i-1}, \beta_i[\prod_{\substack{n=1\\(n\neq i)}}^{N} P_n(z_n)], z_{i+1},\ldots,z_N) \qquad (A.1)$$

Throughout the appendix, $\underline{z}=\underline{1}$ stands for $z_1=z_2=\ldots=z_N=1$.

We first derive the equations (3.29a) and (3.29b) for $\{f_i(j)\}$. Take the derivative of (A.1) with respect to z_j and then let $\underline{z}=\underline{1}$:

$$\left[\frac{\partial}{\partial z_j} R_i[\prod_{n=1}^{N} P_n(z_n)]\right]_{\underline{z}=\underline{1}} = R_i^{(1)}(1)P_j^{(1)}(1) = r_i \lambda_j \qquad 1\leq j\leq N$$

$$\frac{\partial}{\partial z_i} F_i(z_1,\ldots,z_{i-1}, \beta_i[\prod_{\substack{n=1\\(n\neq i)}}^{N} P_n(z_n)], z_{i+1}, \ldots, z_N) = 0$$

$$\left[\frac{\partial}{\partial z_j} F_i(z_1,\ldots,z_{i-1}, \beta_i[\prod_{\substack{n=1\\(n\neq i)}}^{N} P_n(z_n)], z_{i+1},\ldots,z_N)\right]_{\underline{z}=\underline{1}}$$

$$= \left[\frac{\partial F_i}{\partial z_j}\right]_{\underline{z}=\underline{1}} + \left[\frac{\partial F_i}{\partial z_i}\right]_{\underline{z}=\underline{1}} \beta_i^{(1)}(1)P_j^{(1)}(1)$$

$$= f_i(j) + f_i(i) \frac{\lambda_j}{1-\lambda_i} \qquad j\neq i$$

where F_i stands for $F_i(z_1,\ldots,z_N)$ and we have used (3.23b) for $\beta_i^{(1)}(1)$. Thus we have (3.29a) and (3.29b).

We now solve (3.29a) and (3.29b) for $\{f_i(j)\}$. From (3.29b),

$$f_{k+1}(i) - f_k(i) = \mu_i\left[r_k + \frac{f_k(k)}{1-\mu_k}\right] \qquad k\neq i$$

Take the sum of this equation over k=j,j+1,···,i-1 to get

$$f_i(i) - f_j(i) = \mu_i\left[\sum_{k=j}^{i-1} r_k + \sum_{k=j}^{i-1}\frac{f_k(k)}{1-\mu_k}\right] \tag{A.2}$$

Put j=i+1 in (3.29a) and (A.2) to obtain

$$f_i(i) = \mu_i\left[\sum_{k=1}^{N} r_k + \sum_{\substack{k=1\\(k\neq i)}}^{N}\frac{f_k(k)}{1-\mu_k}\right]$$

which solves to give (3.30a).

To find $f_i(j)$, $j\neq i$, use (3.30a) and (A.2) to get

$$f_i(j) = f_j(j) - \mu_j\left[\sum_{k=i}^{j-1} r_k + \sum_{k=i}^{j-1}\frac{f_k(k)}{1-\mu_k}\right]$$

$$= \mu_j\left[\frac{(1-\mu_j)\sum_{k=1}^{N} r_k}{1-\sum_{k=1}^{N}\mu_k} - \sum_{k=i}^{j-1} r_k - \frac{\left(\sum_{k=i}^{j-1}\mu_k\right)\left(\sum_{k=1}^{N} r_k\right)}{1-\sum_{k=1}^{N}\mu_k}\right]$$

$$= \mu_j\left[\frac{\left(1-\sum_{k=i}^{j}\mu_k\right)\left(\sum_{k=1}^{N} r_k\right)}{1-\sum_{k=1}^{N}\mu_k} - \sum_{k=i}^{j-1} r_k\right]$$

However,

$$1 - \sum_{k=i}^{j} \mu_k = 1 - \sum_{k=1}^{N} \mu_k + \sum_{k=j+1}^{i-1} \mu_k$$

Hence, we get (3.30b).

B: Derivation of (3.31a-d) and (3.32b)

Let us derive a set of equations for $\{f_i(j,k)\}$. Differentiating (A.1) with respect to z_j and z_k, we have

$$f_{i+1}(j,k) = \frac{\partial}{\partial z_j}\left\{\left[\frac{\partial}{\partial z_k} R_i\left(\prod_{n=1}^{N} P_n(z_n)\right)\right] \cdot F_i\left(z_1,\cdots,z_{i-1},\Theta_i\left(\prod_{\substack{n=1\\(n\neq i)}}^{N} P_n(z_n)\right),z_{i+1},\cdots,z_N\right)\right.$$

$$\left.\left. + R_i\left(\prod_{n=1}^{N} P_n(z_n)\right) \cdot \left[\frac{\partial}{\partial z_k} F_i\left(z_1,\ldots,z_{i-1},\Theta_i\left(\prod_{\substack{n=1\\(n\neq i)}}^{N} P_n(z_n)\right),z_{i+1},\ldots,z_N\right)\right]\right\}\right|_{\underline{z}=\underline{1}}$$

$$= \left[\frac{\partial^2}{\partial z_j \partial z_k} R_i\left(\prod_{n=1}^{N} P_n(z_n)\right)\right]_{\underline{z}=\underline{1}} + \left[\frac{\partial}{\partial z_k} R_i\left(\prod_{n=1}^{N} P_n(z_n)\right)\right]_{\underline{z}=\underline{1}}$$

$$\cdot \left[\frac{\partial}{\partial z_j} F_i\left(z_1,\cdots,z_{i-1},\Theta_i\left(\prod_{\substack{n=1\\(n\neq i)}}^{N} P_n(z_n)\right),z_{i+1},\cdots,z_N\right)\right]_{\underline{z}=\underline{1}}$$

$$+ \left[\frac{\partial}{\partial z_j} R_i\left(\prod_{n=1}^{N} P_n(z_n)\right)\right]_{\underline{z}=\underline{1}}$$

$$\cdot\left[\frac{\partial}{\partial z_k} F_i(z_1,\cdots,z_{i-1},\Theta_i(\prod_{\substack{n=1\\(n\neq i)}}^{N} P_n(z_n)),z_{i+1},\cdots,z_N)\right]_{\underline{z}=\underline{1}}$$

$$+\left[\frac{\partial^2}{\partial z_j \partial z_k} F_i(z_1,\cdots,z_{i-1},\Theta_i(\prod_{\substack{n=1\\(n\neq i)}}^{N} P_n(z_n)),z_{i+1},\cdots,z_N)\right]_{\underline{z}=\underline{1}} \qquad (A.3)$$

Now, evaluating the first term, we have

$$\left[\frac{\partial^2}{\partial z_j \partial z_k} R_i(\prod_{n=1}^{N} P_n(z_n))\right]_{\underline{z}=\underline{1}} = \begin{cases} \mu_j\mu_k(\delta_i^{\,2}+r_i^{\,2}) & j\neq k \\ \mu_j^{\,2}(\delta_i^{\,2}+r_i^{\,2}) + r_i(\sigma_j^{\,2}-\mu_j) & j=k \end{cases}$$

The second and third terms of (A.3) have been evaluated in Appendix A. The last term is evaluated as

$$\left[\frac{\partial^2}{\partial z_j \partial z_k} F_i(z_1,\cdots,z_{i-1},\Theta_i(\prod_{\substack{n=1\\(n\neq i)}}^{N} P_n(z_n)),z_{i+1},\cdots,z_N)\right]_{\underline{z}=\underline{1}}$$

$$= \begin{cases} f_i(i)\mu_j\mu_k[\Theta_i^{(1)}(1)+\Theta_i^{(2)}(1)]+\Theta_i^{(1)}(1)[f_i(i,j)\mu_k+f_i(i,k)\mu_j] \\ \quad +f_i(j,k)+f_i(i,i)[\Theta_i^{(1)}(1)]^2\mu_j\mu_k & i\neq j, i\neq k, j\neq k \\ f_i(i)[\Theta_i^{(2)}(1)\mu_j^{\,2}+\Theta_i^{(1)}(1)P_j^{(2)}(1)]+2\mu_j\Theta_i^{(1)}(1)f_i(i,j) \\ \quad +f_i(j,j)+f_i(i,i)[\Theta_i^{(1)}(1)]^2\mu_j^{\,2} & i\neq j=k \\ 0 & i=j \text{ or } i=k \end{cases}$$

Using these evaluations in (A.3), we obtain (3.31a-d).

Now, assuming the case of identical stations, we find $f^{(2)} \triangleq f_i(i,i)$. By dropping the subscripts from the parameters, the equations (3.31a-d) become

$$f_{i+1}(j,k) = a + b[f_i(j) + f_i(k)] + cf^{(1)} + d[f_i(i,j) + f_i(i,k)] + f_i(j,k) + d^2 f^{(2)} \qquad i \neq j,\ i \neq k,\ j \neq k \qquad \text{(A.4a)}$$

$$f_{i+1}(j,j) = a + r(\sigma^2 - \mu) + 2bf_i(j) + \left(\frac{\sigma^2 - \mu}{1-\mu} + c\right) f^{(1)} + f_i(j,j) + 2df_i(i,j) + d^2 f^{(2)} \qquad i \neq j \qquad \text{(A.4b)}$$

$$f_{i+1}(i,k) = a + b[f_i(k) + df^{(1)}] \qquad i \neq k \qquad \text{(A.4c)}$$

$$f_{i+1}(i,i) = a + r(\sigma^2 - \mu) \qquad \text{(A.4d)}$$

where

$$a \triangleq \mu^2(\sigma^2 + r^2), \quad b \triangleq r\mu, \quad c \triangleq \mu^2\left[\frac{2r}{1-\mu} + \frac{1}{(1-\mu)^2} + \frac{\sigma^2}{(1-\mu)^3}\right]$$

$$d \triangleq \frac{\mu}{1-\mu}, \quad f^{(1)} \triangleq f_i(i) = \frac{Nr\mu(1-\mu)}{1 - N\mu} \qquad \text{(A.5)}$$

First, consider the case $j \neq k$. Sum (A.4a) and (A.4c) over all i to have

$$\sum_{i=1}^{N} f_{i+1}(j,k) = Na + b\left\{\sum_{\substack{i=1\\(i \neq j)}}^{N} f_i(j) + \sum_{\substack{i=1\\(i \neq k)}}^{N} f_i(k)\right\} + [(N-2)c + 2bd]f^{(1)}$$

$$+ d \sum_{\substack{i=1 \\ (i \neq j,k)}}^{N} [f_i(i,j)+f_i(i,k)] + (N-2)d^2 f^{(2)} + \sum_{\substack{i=1 \\ (i \neq j,k)}}^{N} f_i(j,k)$$

Noting $\Sigma_{i=1}^{N} f_{i+1}(j,k)=\Sigma_{i=1}^{N} f_i(j,k)$, we have

$$f_j(j,k)+f_k(j,k) = Na+b\{\sum_{i=1}^{N} [f_i(j)+f_i(k)] - 2f^{(1)}\}$$

$$+ d\{\sum_{i=1}^{N} [f_i(i,j)+f_i(i,k)] - 2f^{(2)} - f_j(j,k) - f_k(k,j)\}$$

$$+ [(N-2)c+2bd]f^{(1)} + (N-2)d^2 f^{(2)} \qquad j \neq k \qquad \text{(A.6a)}$$

Next, consider the case j=k. Sum (A.4b) and (A.4d) over all i to have

$$f_j(j,j) = N[a+r(\sigma^2-\mu)]+2b\{\sum_{i=1}^{N} f_i(j)-f^{(1)}\}+(N-1)(\frac{\sigma^2-\mu}{1-\mu}+c)f^{(1)}$$

$$+ (N-1)d^2 f^{(2)} + 2d\{\sum_{i=1}^{N} f_i(i,j)-f^{(2)}\} \qquad \text{(A.6b)}$$

Note that (3.30b) for identical stations takes form

$$f_i(j) = \begin{cases} \mu\left[(i-j)r + \dfrac{(i-j-1)Nru}{1-N\mu}\right] & i \geq j+1 \\ \mu\left[(N-j+i)r + \dfrac{(N-j+i-1)Nru}{1-\mu}\right] & i \leq j-1 \end{cases}$$

It follows that

$$e \triangleq \sum_{i=1}^{N} f_i(j) - f^{(1)} = \frac{N(N-1)r\mu(1-2\mu)}{2(1-N\mu)} \qquad \text{(A.7a)}$$

Also let

$$g \triangleq \sum_{i=1}^{N} f_i(i,j) = \sum_{i=1}^{N} f_j(j,i) \qquad \text{(A.7b)}$$

By symmetry among stations, e and g are independent of j. Thus, (A.6a) and (A.6b) can be rewritten, respectively, as

$$f_j(j,k)+f_k(j,k) = Na + 2be + d[2g-f_j(j,k)-f_k(k,j)]$$
$$+ [(N-2)c+2bd]f^{(1)} + [(N-2)d^2-2d]f^{(2)} \qquad j \neq k \qquad \text{(A.8a)}$$

$$f^{(2)}=N[a+r(\sigma^2-\mu)]+2be+(N-1)\left(\frac{\sigma^2-\mu}{1-\mu}+c\right)f^{(1)}+[(N-1)d^2-2d]f^{(2)}+2dg \qquad \text{(A.8b)}$$

Finally, sum (A.8a) and (A.8b) over all k to get

$$2[1-(N-1)d]g = N^2a + Nr(\sigma^2-\mu) + 2Nbe$$
$$+ (N-1)\left[\frac{\sigma^2-\mu}{1-\mu} + (N-1)c+2bd\right]f^{(1)} + [1-(N-1)d]^2 f^{(2)} \qquad \text{(A.9)}$$

Solve the equations (A.8b) and (A.9) for two unknowns g and $f^{(2)}$. Substituting (A.5) and (A.7a) in the expression for $f^{(2)}$ gives

$$f^{(2)} = \frac{\delta^2\mu^2 N(1-\mu)}{1 - N\mu} + \frac{\sigma^2 rN[1-(N+1)\mu+(2N-1)\mu^2]}{(1 - N\mu)^2}$$

$$- \frac{Nr\mu(1-\mu)}{1 - N\mu} + \frac{N^2 r^2 \mu^2 (1-\mu)^2}{(1 - N\mu)^2} \qquad \text{(A.10)}$$

This leads to (3.32b).

C: Derivation of (4.49a) and (4.49b)

We derive (4.49a) and (4.49b) from (4.46). If

$$t_m \triangleq s_m + \lambda_k[1-\theta_k^*(s_N)] \qquad m=1,2,\cdots,N \qquad \text{(A.11)}$$

then the first derivative of the right-hand side of (4.46) with respect to s_N is given by

$$\frac{\partial}{\partial s_N}[R_{k-1}^*(t_N)X_{k-1}^*(0,t_1,t_2,\cdots,t_{N-1})]$$

$$= \frac{dt_N}{ds_N}\frac{dR_{k-1}^*(t_N)}{dt_N} X_{k-1}^*(0,t_1,t_2,\ldots,t_{N-1})$$

$$+ R_{k-1}^*(t_N)\sum_{m=1}^{N-1}\frac{dt_m}{ds_N}\frac{\partial X_{k-1}^*(0,t_1,t_2,\ldots,t_{N-1})}{\partial t_m} \qquad \text{(A.12)}$$

Differentiating (A.12) with respect to s_N once more yields

$$\frac{\partial^2 (RX)}{\partial s_N^{\,2}} = \left[\frac{d^2 t_N}{ds_N^{\,2}} \frac{dR}{dt_N} + \left(\frac{dt_N}{ds_N} \right)^2 \frac{d^2 R}{dt_N^{\,2}} \right] X$$

$$+ 2 \frac{dt_N}{ds_N} \frac{dR}{dt_N} \sum_{m=1}^{N-1} \frac{dt_m}{ds_N} \frac{\partial X}{\partial t_m} + R \sum_{m=1}^{N-1} \frac{d^2 t_m}{ds_N^{\,2}} \frac{\partial X}{\partial t_m}$$

$$+ R \sum_{n=1}^{N-1} \sum_{m=1}^{N-1} \frac{dt_n}{ds_N} \frac{dt_m}{ds_N} \frac{\partial^2 X}{\partial t_n \partial t_m} \qquad \text{(A.13)}$$

where, for simplicity, we have set

$$R \equiv R_{k-1}^{\,*}(t_N) \qquad X \equiv X_{k-1}^{\,*}(0, t_1, t_2, \ldots, t_{N-1})$$

If $\underline{s} = \underline{0}$ stands for $s_1 = \ldots = s_N = 0$, then

$$\left(\frac{dR}{dt_N} \right)_{\underline{s}=\underline{0}} = -r_{k-1} \;, \quad \left(\frac{d^2 R}{dt_N^{\,2}} \right)_{\underline{s}=\underline{0}} = \sigma_{k-1}^{\,2} + r_{k-1}^{\,2}$$

$$\left(\frac{dt_N}{ds_N} \right)_{\underline{s}=\underline{0}} = 1 + \lambda_k \theta_k \;, \quad \left(\frac{d^2 t_N}{ds_N^{\,2}} \right)_{\underline{s}=\underline{0}} = -\lambda_k \theta_k^{(2)}$$

$$\left(\frac{dt_m}{ds_N} \right)_{\underline{s}=\underline{0}} = \lambda_k \theta_k \;, \quad \left(\frac{d^2 t_m}{ds_N^{\,2}} \right)_{\underline{s}=\underline{0}} = -\lambda_k \theta_k^{(2)} \qquad m=1,2,\ldots,N-1$$

$$\left(\frac{\partial X}{\partial t_m} \right)_{\underline{s}=\underline{0}} = -E[\tilde{x}_{k-N+m}] \;, \quad \left(\frac{\partial^2 X}{\partial t_n \partial t_m} \right)_{\underline{s}=\underline{0}} = E[\tilde{x}_{k-N+n}\, \tilde{x}_{k-N+m}]$$

$$m,n = 1,2,\ldots,N-1 \qquad \text{(A.14)}$$

where θ_k and $\theta_k^{(2)}$ are given in (4.3b). Using (A.14) to evaluate (A.13) at $\underline{s}=\underline{0}$, we have

$$E[\tilde{x}_k^{\,2}] = \lambda_k \theta_k^{(2)} r_{k-1} + (1+\lambda_k \theta_k)^2 (\sigma_{k-1}^{\,2} + r_{k-1}^{\,2}) + 2(1+\lambda_k \theta_k) r_{k-1} \sum_{m=1}^{N-1} \lambda_k \theta_k E[\tilde{x}_{k-N+m}]$$

$$+ \sum_{m=1}^{N-1} \lambda_k \theta_k^{(2)} E[\tilde{x}_{k-N+m}] + \sum_{n=1}^{N-1} \sum_{m=1}^{N-1} (\lambda_k \theta_k)^2 E[\tilde{x}_{k-N+n}\, \tilde{x}_{k-N+m}] \qquad \text{(A.15a)}$$

From (4.47a), we have

$$(E[\tilde{x}_k])^2 = (1+\lambda_k\theta_k)^2 r_{k-1}{}^2 + 2(1+\lambda_k\theta_k)\, r_{k-1}\, \lambda_k\theta_k \sum_{m=1}^{N-1} E[\tilde{x}_{k-m}]$$

$$+ (\lambda_k\theta_k)^2 \left(\sum_{n=1}^{N-1} E[\tilde{x}_{k-n}]\right)\left(\sum_{m=1}^{N-1} E[\tilde{x}_{k-m}]\right) \qquad \text{(A.15b)}$$

Thus, we get

$$E[(\tilde{x}_k-\bar{x}_k)^2] = (1+\lambda_k\theta_k)^2 \delta_{k-1}{}^2 + \lambda_k\theta_k{}^{(2)} \left\{ r_{k-1} + \sum_{m=1}^{N-1} E[\tilde{x}_{k-m}] \right\}$$

$$+ (\lambda_k\theta_k)^2 \sum_{n=1}^{N-1} \sum_{m=1}^{N-1} E[(\tilde{x}_{k-n} - \bar{x}_{k-n})(\tilde{x}_{k-m} - \bar{x}_{k-m})]$$

(A.16)

If we use (4.3b), (4.42) and (4.49c) in (A.16), we obtain (4.49a).

Differentiating (A.12) with respect to s_{N-j}, $j < N$, yields

$$\frac{\partial^2 (RX)}{\partial s_N s_{N-j}} = \frac{dt_N}{ds_N}\, \frac{dR}{dt_N}\, \frac{\partial X}{\partial t_{N-j}}\, \frac{dt_{N-j}}{ds_{N-j}} + R \sum_{m=1}^{N-1} \frac{dt_m}{ds_N}\, \frac{\partial^2 X}{\partial t_m \partial t_{N-j}}\, \frac{dt_{N-j}}{ds_{N-j}} \qquad \text{(A.17)}$$

Using (A.14) and $d(t_{N-j})/ds_{N-j} = 1$ to evaluate (A.17), we have

$$E[\tilde{x}_k \tilde{x}_{k-j}] = (1+\lambda_k\theta_k) r_{k-1} E[\tilde{x}_{k-j}] + \sum_{m=1}^{N-1} \lambda_k\theta_k\, E[\tilde{x}_{k-N+m}\, \tilde{x}_{k-j}] \qquad \text{(A.18a)}$$

From (4.47a), we have

$$E[\tilde{x}_k] E[\tilde{x}_{k-j}] = E[\tilde{x}_{k-j}] \left\{ (1+\lambda_k\theta_k) r_{k-1} + \lambda_k\theta_k \sum_{m=1}^{N-1} E[\tilde{x}_{k-m}] \right\} \qquad \text{(A.18b)}$$

Thus we get

$$E[(\tilde{x}_k-\bar{x}_k)(\tilde{x}_{k-j}-\bar{x}_{k-j})] = \lambda_k\theta_k \sum_{m=1}^{N-1} E[(\tilde{x}_{k-m}-\bar{x}_{k-m})(\tilde{x}_{k-j}-\bar{x}_{k-j})] \qquad \text{(A.19)}$$

which is (4.49b).

References

[Amin75] Y.Aminetzah, "An Exact Approach to the Polling System," Ph.D. Thesis, March 1975, McGill University.

[Avi65] B.Avi-Itzhak, W.L.Maxwell and L.W.Miller, "Queueing with Alternating Priorities," Operations Research, Vol.13, No.2, pp.306-318, March-April 1965.

[Bhar60] A.T.Bharucha-Reid, Elements of the Theory of Markov Processes and Their Applications, McGraw-Hill Book Company, 1960.

[Boxm84] O.J.Boxma, "Two Symmetric Queues with Alternating Service and Switching Times," Performance '84.

[Bux81] W.Bux, "Local-Area Subnetworks: A Performance Comparison," IEEE Transactions on Communications, Vol.COM-29, No.10, pp.1465-1473, October 1981.

[Bux83a] W.Bux and H.L.Truong, "Mean-Delay Approximation for Cyclic Service Queueing Systems," Performance Evaluation, Vol.3, No.3, pp.187-196, August 1983.

[Bux83b] W.Bux, F.H.Closs, K.Kuemmerle, H.J.Keller and H.R.Mueller, "Architecture and Design of a Reliable Token-Ring Network," IEEE Journal on Selected Areas in Communications, Vol.SAC-1, No.5, pp.756-765, November 1983.

[Cars77] R.T.Carsten, E.E.Newhall and M.J.M.Posner, "A Simplified Analysis of Scan Times in an Asymmetrical Newhall Loop with Exhaustive Service," IEEE Transactions on Communications, Vol.COM-25, No.9, pp.951-957, September 1977.

[Cars78] R.T.Carsten and M.J.M.Posner, "Simplified Statistical Models of Single and Multiple Newhall Loops," Conference Record, 1978 National Telecommunications Conference, pp.44.5.1-44.5.7, Birmingham, Alabama, December 3-6, 1978, IEEE 78CH1354-0.

[Chu72] W.W.Chu and A.G.Konheim, "On the Analysis and Modeling of a Class of Computer Communication Systems," IEEE Transactions on Communications, Vol.COM-20, No.3, pp.645-660, June 1972.

[Coff82] E.G.Coffman, Jr. and M.Hofri, "On the Expected Performance of Scanning Disks," SIAM Journal on Computing, Vol. 11, No.1, pp.60-70, February 1982.

[Coff84] E.G.Coffman, Jr. and E.N.Gilbert, "A Continuous Polling System with Constant Service Times," Preprint.

[Cohe83] J.W.Cohen and O.J.Boxma, Boundary Value Problems in Queueing System Analysis, Mathematics Studies 79, North-Holland, 1983.

[Conw67] R.W.Conway, W.L.Maxwell and L.W.Miller, Theory of Scheduling, Addison-Wesley, 1967.

[Coop69] R.B.Cooper and G.Murray, "Queues Served in Cyclic Order," The Bell System Technical Journal, Vol.48, No.3, pp.675-689, March 1969.

[Coop70] R.B.Cooper, "Queues Served in Cyclic Order: Waiting Times," The Bell System Technical Journal, Vol,49, No.3, pp.399-413, March 1970.

[Coop81] R.B.Cooper, Introduction to Queueing Theory, Second Edition, North-Holland, 1981.

[DeMo81] L.F.M.de Moraes, "Message Queueing Delays in Polling Schemes with Applications to Data Communication Networks," Department of System Science, School of Engineering and Applied Science, University of California, Los Angeles, UCLA-ENG-8106, May 1981.

[DeMo84] L.F.M.de Moraes and I.Rubin, "Analysis and Comparison of Message Queueing Delays in Token-Rings and Token-Buses Local Area Networks," Proceedings of IEEE International Conference on Communications, pp.130-135, Amsterdam, May 14-17, 1984, IEEE 84CH2028-9.

[Eise71] M.Eisenberg, "Two Queues with Changeover Times," Operations Research, Vol.19, No.2, pp.386-401, March-April 1971.

[Eise72] M.Eisenberg, "Queues with Periodic Service and Changeover Times," Operations Research, Vol.20, No.2, pp.440-451, March-April 1972.

[Eise79] M.Eisenberg, "Two Queues with Alternating Service," SIAM Journal on Applied Mathematics, Vol.36, No.2, pp.287-303, April 1979.

[Ferg84] M.J.Ferguson and Y.Aminetzah, "Exact Results for Non-Symmetric Token Ring Systems," Preprint.

[Gold83] Y.I.Gold and W.R.Franta, "An Efficient Collision-Free Protocol for Prioritized Access-Control of Cable or Radio Channels," Computer Networks, Vol.7, No.2, pp.83-98, 1983.

[Hash70] O.Hashida, "Gating Multiqueues Served in Cyclic Order," Systems-Computers-Controls, Vol.1, No.1, pp.1-8, 1970.

[Hash72] O.Hashida, "Analysis of Multiqueue," Review of the Electrical Communication Laboratories, Vol.20, Nos.3-4, pp.189-199, March-April 1972.

[Hash81a] O.Hashida and K.Kawashima, "Analysis of a Polling System with Single User at Each Terminal," Review of the Electrical Communication Laboratories, Vol.29, Nos.3-4, pp.245-253, March-April 1981.

[Hash81b] O.Hashida, Study of Multi-Queues in Communication Control, May 1981 (in Japanese).

[Haye81] J.F.Hayes, "Local Distribution in Computer Communications," IEEE Communications Magazine, Vol.19, No.2, pp.6-14, March 1981.

[Heym82] D.P.Heyman and M.J.Sobel, Stochastic Models in Operations Research, Volume I: Stochastic Processes and Operating Characteristics, McGraw-Hill Book Company, 1982.

[Heym83] D.P.Heyman, "Data-Transport Performance Analysis of Fasnet," The Bell System Technical Journal, Vol.62, No.8, pp.2547-2560, October 1983.

[Humb78] P.A.Humblet, "Source Coding for Communication Concentrators," Report ESL-R-798, Electronic Systems Laboratory, Massachusetts Institute of Technology, January 1978.

[IEEE83] IEEE Project 802, Local-Area Network Standards, Draft IEEE Standard 802.5, Token-Ring Access Method and Physical-Layer Specifications, Working Draft, June 10, 1983.

[Iisa80] S.Iisaku, N.Miki, N.Nagai and K.Hatori, "A Consideration of Multiqueues with Nonexhaustive and Cyclic Service," The Transactions of the Institute of Electronics and Communication Engineers of Japan, Vol.J63-B, No.11, pp.1102-1109, November 1980 (in Japanese).

[Iisa81a] S.Iisaku, N.Miki, N.Nagai and K.Hatori, "Two Queues with Alternating Service and Walking Time," The Transactions of the Institute of Electronics and Communication Engineers of Japan, Vol.J64-B, No.4, pp.342-343, April 1981 (in Japanese).

[Iisa81b] S.Iisaku, N.Miki, N.Nagai and K.Hatori, "Relationship of the Number of Waiting Customers to Waiting Time of Multiqueues with Nonexhaustive and Cyclic Service," The Transactions of the Institute of Electronics and Communication Engineers of Japan, Vol.J64-B, No.8, pp.891-892, August 1981 (in Japanese).

[Kaye72] A.R.Kaye, "Analysis of a Distributed Control Loop for Data Transmission," Proceedings of the Symposium on Computer-Communications Networks and Teletraffic, pp.47-58, Polytechnic Institute of Brooklyn, April 4-6, 1972.

[Klei75] L.Kleinrock, Queueing Systems, Volume 1: Theory, John Wiley and Sons, 1975.

[Klei76] L.Kleinrock, Queueing Systems, Volume 2: Computer Applications, John Wiley and Sons, 1976.

[Klei80] L.Kleinrock and M.O.Scholl, "Packet Switching in Radio Channels: New Conflict-Free Access Schemes," IEEE Transactions on Communications, Vol.COM-28, No.7, pp.1015-1029, July 1980.

[Koba77] H.Kobayashi and A.G.Konheim, "Queueing Models for Computer Communications System Analysis," IEEE Transactions on Communications, Vol. COM-25, No.1, pp.[illegible]-29, January 1977.

[Konh74] A.G.Konheim and B.Meister, "Waiting Lines and Times in a System with Polling," Journal of the Association for Computing Machinery, Vol. 21, No.3, pp.470-490, July 1974.

[Konh76] A.G.Konheim, "Chaining in a Loop System," IEEE Transactions on Communications, Vol.COM-24, No.2, pp.203-210, February 1976.

[Konh80] A.G.Konheim, "Mathematical Models for Computer Data Communication," Case Studies in Mathematical Modeling, W.E.Boyce (ed.), pp.256-334, Pitman Advanced Publishing Program, Boston, 1980.

[Kueh79] P.J.Kuehn, "Multiqueue Systems with Nonexhaustive Cyclic Service," The Bell System Technical Journal, Vol.58, No.3, pp.671-698, March 1979.

[Kueh81] P.J.Kuehn, "Performance of ARQ-Protocols for HDX Transmission in Hierarchical Polling Systems," Performance Evaluation, Vol.1, No.1, pp. 19-30, January 1981.

[Kuro81] K.Kurosawa and S.Tsujii, "Analysis on Asymmetric Polling Systems with Limiting Service," National Telecommunications Conference 1981, pp. G.4.2.1-G.4.2.5, New Orleans, Louisiana, November 29 - December 3, 1981, IEEE 81CH1679-0.

[Leib61] M.A.Leibowitz, "An Approximate Method for Treating a Class of Multiqueue Problems," IBM Journal of Research and Development, Vol.5, No.3, pp.204-209, July 1961.

[Leib62] M.A.Leibowitz, "A Note on Some Fundamental Parameters of Multiqueue Systems," IBM Journal of Research and Development, Vol.6, No.4, pp.470-471, October 1962.

[Litt61] J.D.C.Little, "A Proof for the Queueing Formula: $L = \lambda W$," Operations Research, Vol.9, No.3, pp.383-387, May-June 1961.

[Mack57a] C.Mack, T.Murphy and N.L.Webb, "The Efficiency of N Machines Uni-Directionally Patrolled by One Operative when Walking Time and Repair Times are Constants," Journal of the Royal Statistical Society, Ser.B. Vol.19, No.1, pp.166-172, 1957.

[Mack57b] C.Mack, "The Efficiency of N Machines Uni-Directionally Patrolled by One Operative when Walking Time is Constant and Repair Times are Variable," Journal of the Royal Statistical Society, Ser.B, Vol.19, No.1, pp.173-178, 1957.

[Manf83] D.R.Manfield, "Analysis of a Polling System with Priorities," IEEE Global Telecommunications Conference, pp.43.4.1-43.4.5, San Diego, California, November 28 - December 1, 1983, IEEE 83CH1956-2.

[Mart67] J.Martin, Design of Real-Time Computer Systems, Prentice-Hall, Inc., 1967.

[Morr84] R.J.T.Morris and Y.T.Wang, "Some Results for Multi-Queue Systems with Multiple Cyclic Servers," Performance of Computer-Communication Systems, H.Rudin and W.Bux (eds.), pp.245-258, Elsevier Science Publishers B.V. (North-Holland), 1984.

[Nomu78] M.Nomura and K.Tsukamoto, "Traffic Analysis on Polling Systems," The Transactions of the Institute of Electronics and Communication Engineers of Japan, Vol.J61-B, No.7, pp.600-607, July 1978 (in Japanese).

[Reis82] M.Reiser, "Performance Evaluation of Data Communications Systems," Proceedings of the IEEE, Vol.70, No.2, pp.171-196, February 1982.

[Rubi83] I.Rubin and L.F.M.de Moraes, "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.

[Saat61] T.L.Saaty, Elements of Queueing Theory with Applications, McGraw-Hill Book Company, 1961.

[Scho78] M.Scholl and D.Potier. "Finite and Infinite Source Models for Communication Systems under Polling," IRIA Rapport de Recherche, No.308, May 1978.

[Schw77] M.Schwartz, Computer-Communication Network Design and Analysis, Prentice-Hall, 1977.

[Stid72] S.Stidham, "Regenerative Processes in the Theory of Queues, with Applications to the Alternating-Priority Queue," Advances in Applied Probability, Vol.4, No.3, pp.542-577, December 1972.

[Swar80] G.B.Swartz, "Polling in a Loop System," Journal of the Association for Computing Machinery, Vol.27, No.1, pp.42-59, January 1980.

[Swar82] G.B.Swartz, "Analysis of a SCAN Policy in a Gated Loop System," Applied Probability - Computer Science: The Interface, Volume II, R.L.Disney and T.J.Ott (eds.), pp.241-252, Birkhauser, 1982.

[Syke70] J.S.Sykes, "Simplified Analysis of an Alternating-Priority Queueing Model with Setup Times," Operations Research, Vol.18, No.6, pp.1182-1192, November-December 1970.

[Takac68] L.Takacs, "Two Queues Attended by a Single Server," Operations Research, Vol.16, No.3, pp.639-650, May-June 1968.

[Takag84] H.Takagi, "Mean Message Waiting Time in a Symmetric Polling System," Performance '84, 1984.

[Toba76] F.A.Tobagi and L.Kleinrock, "Packet Switching in Radio Channels: Part III - Polling and Dynamic Split-Channel Reservation Multiple Channel Access," IEEE Transactions on Communications, Vol.COM-24, No.8, pp.832-845, August 1976.

[Toba83a] F.A.Tobagi, F.Borgonovo and L.Fratta "Expressnet: A High-Performance Integrated-Services Local Area Network," IEEE Journal on Selected Areas in Communications, Vol.SAC-1, No.5, pp.898-913, November 1983.

[Toba83b] F.A.Tobagi and M.Fine, "Performance of Unidirectional Broadcast Local Area Networks: Expressnet and Fasnet," IEEE Journal on Selected Areas in Communications, Vol.SAC-1, No.5, pp.913-926, November 1983.

[Tran84] P.Tran-Gia and T.Raith, "Multiqueue Systems with Finite Capacity and Nonexhaustive Cyclic Service," Preprint.