Queuing Analysis of Polling Models

HIDEAKI TAKAGI

IBM Research, Tokyo Research Laboratory, 5-19 Sanban-cho, Chiyoda-ku, Tokyo 102, Japan

A polling model is a system of multiple queues accessed by a single server in cyclic order. Polling models provide performance evaluation criteria for a variety of demand-based, multiple-access schemes in computer and communication systems. This paper presents an overview of the state of the art of polling model analysis, as well as an extensive list of references. In particular, single-buffer systems and infinite-buffer systems with exhaustive, gated, and limited service disciplines are treated. There is also some discussion of systems with a noncyclic order of service and systems with priority. Applications to computer networks are illustrated, and future research topics are suggested.

Categories and Subject Descriptors: C.2.5 [**Computer-Communication Networks**]: Local Networks; C.4 [**Computer Systems Organization**]: Performance of Systems—*modeling techniques*; D.4.8 [**Operating Systems**]: Performance—*queuing theory*

General Terms: Performance

Additional Key Words and Phrases: Cyclic service systems, modeling of token rings, polling models

INTRODUCTION

A *polling model* is a system of multiple queues accessed in cyclic order by a single server. In recent decades, polling models have been used to analyze the performance of a variety of systems. In the late 1950s, a polling model with a single buffer for each queue was used in an investigation of a problem in the British cotton industry involving a patrolling machine repairman [Mack 1957; Mack et al. 1957]. In the 1960s, polling models with two queues were used to analyze traffic signal control (see a survey by Stidham [1969]). There were also some early studies from the viewpoint of queuing theory that were apparently independent of traffic analysis (e.g., Avi-Itzhak et al. [1965]). In the 1970s, with the advent of computer communication networks, an extensive study was carried out on a polling scheme for data transfer from terminals on multidrop lines to a central computer. Since the early 1980s, the same model has been revived by Bux [1981] and others to study token passing schemes (e.g., token ring and token bus) in local-area networks (LANs). More recently, it has been used for resource arbitration and load sharing for multiprocessor computers [Wang and Morris 1985]. Numerous other applications exist in manufacturing systems. A polling

CONTENTS

model was used in a nontechnical article in *Scientific American* [Liebowitz 1968] as an example of an interesting and important queuing system.

The usual objective in analyzing polling models is to find the *message waiting time*, defined as the time from the arrival of a randomly chosen message to the beginning of its service. The mean waiting time plus the mean service time is the mean message response time, which is the *single most important performance measure* in most computer communication systems [Kleinrock 1976, p. 161]. We are also interested in the *polling cycle time*, which is the time between the server's visit of the same queue in successive cycles. The mean cycle time is easily found for infinite-buffer systems. Here, we focus on methods for determining these two performance measures.

In this paper, we present an overview of the state of the art of polling model analysis, along with a comprehensive list of references. We also point out some challenging problems that require further research. For a more detailed presentation, the reader is referred to Takagi [1986]. (Since the publication of that study, some significant progress has already been made.)

There are many varieties of polling models, but we can classify them broadly according to the following characteristics:

- continuous-time or discrete-time systems,
- single or infinite buffers for each queue,
- service disciplines (exhaustive, gated, limited, decrementing),
- systems with or without server switchover times,
- symmetric or asymmetric parameters,
- cyclic or noncyclic order of service,
- exact or approximate solutions available.

Polling models are referred to in many survey articles and book chapters on data communication systems [Bertsekas and Gallager 1987, sec. 3.5.2; Chu and Konheim 1972; Hayes 1981, 1984, chap. 7; Hayes and Sherman 1972; Kaye and Richardson 1973; Kobayashi and Konheim 1977; Konheim 1980; Lam 1983; Penny and Baghdadi 1979a, 1979b; Reiser 1982].

The paper is organized as follows: In Section 1 we deal with single buffer systems. In Section 2 we consider various models of infinite buffer systems. In Section 3 we briefly discuss polling models with a noncylic order of service and those involving priority structure and in Section 4 applications of polling models. Future research topics are suggested in Section 5, and a list of references is provided at the end of the paper.

Some notation common to all polling models is given here. The number of queues in the system is denoted by N. Queues are indexed by i ($i = 1, 2, \ldots, N$)

in order of service. In continuous-time models, we assume a Poisson process for the arrival of messages at rate λ_i for queue i. The mean and the second moment of the message service time at queue i are denoted by b_i, and $b_i^{(2)}$, respectively. The offered load is given by

$$\rho = \sum_{i=1}^{N} \rho_i, \qquad \text{where} \quad \rho_i \triangleq \lambda_i b_i. \tag{1}$$

The mean and variance of the time needed by the server to switch from queue i to queue $i \bmod N + 1$ are denoted by r_i and δ_i^2, respectively. The switchover times are assumed to be independent of the arrival and service processes. The mean and variance of the total switchover time are given by

$$R \triangleq \sum_{i=1}^{N} r_i; \qquad \Delta^2 \triangleq \sum_{i=1}^{N} \delta_i^2. \tag{2}$$

The mean cycle time, generally independent of i, is denoted by $E[C]$. The mean message waiting time is denoted by $E[W_i]$ for queue i. For symmetric systems (all queues are statistically identical), we let $\rho_i = \rho/N = \lambda b$ and omit subscript i from other variables.

1. SINGLE-BUFFER SYSTEMS

We first consider a continuous-time polling model where each queue can have at most one outstanding message; that is, where the messages that find the buffer full on arrival at a queue are lost. We assume a Poisson arrival process at each queue. Thus, our system may also be thought of as one in which it takes an exponentially distributed time to generate a new message at each queue after the service to the previous message has been completed. Our model can be compared to a *machine repair system* in which a repairman walks from machine to machine in cyclic order fixing broken machines; the message arrival and its service correspond to a machine stoppage and its repair, respectively. This view appears in the early literature [Bharucha-Reid 1960, sec. 9.4D; Mack 1957; Mack et al. 1957; Runnenberg 1957].

Assuming a symmetric system and using the above notation, it is shown [Mack et al. 1957; Scholl and Potier 1978] that

$$E[C] = R + bE[Q], \tag{3}$$

$$E[W] = (N-1)b - \frac{1}{\lambda} + \frac{NR}{E[Q]}, \tag{4}$$

where $E[Q]$ is the mean number of messages served in a polling cycle. If γ denotes the *throughput* of the system (the mean number of messages served per unit time) and $E[X]$ denotes the *mean message interdeparture time*, we have

$$\gamma = \frac{1}{E[X]} = \frac{N}{E[W] + b + 1/\lambda}. \tag{5}$$

Thus, determination of $E[Q]$ or $E[X]$ is sufficient to compute these performance measures.

In a symmetric system with constant service and switchover times, $E[Q]$ is given explicitly by [Mack et al. 1957]

$$E[Q] = \frac{N\sum_{n=0}^{N-1} \binom{N-1}{n} \prod_{j=0}^{n} [e^{\lambda(R+jb)} - 1]}{1 + \sum_{n=1}^{N} \binom{N}{n} \prod_{j=0}^{n-1} [e^{\lambda(R+jb)} - 1]}. \tag{6}$$

The distribution of the message waiting time is found in Kaye [1972] and Takagi [1986, sec. 2.4] (which applies a method in Hashida and Kawashima [1979] to the case of constant parameters).

When the message service time and switchover time are random variables (symmetry is still assumed), we do not have an explicit expression for $E[Q]$ for an arbitrary value of N. This case is analyzed in Mack [1957] (very difficult to follow), Takagi [1985a], and Takine et al. [1986, 1987b] (which derive the distribution of X). We are usually led to a set of $O(2^N)$ linear equations to obtain $E[Q]$ or $E[X]$. We note that, in the limit $R \to 0$, $E[W]$ for this model should approach that for the corresponding first-come, first-served (FCFS) M/G/1//N queue, or *machine repairman model* [Saaty 1961, sec. 14-6; Takács 1962, chap. 5]; thus,

$$\lim_{R\to 0} E[W] = (N-1)b - \frac{1}{\lambda} + \frac{1}{\lambda[1 + \sum_{n=1}^{N-1} \binom{N-1}{n} \prod_{j=1}^{n} \{[B^*(j\lambda)]^{-1} - 1\}]}, \tag{7}$$

where $B^*(s)$ is the Laplace–Stieltjes transform of the distribution function for the message service time. As a result of symmetry, the sample paths of the *unfinished work* in both systems are statistically identical. It follows from Little's theorem that $E[W]$ should be the same for both systems. Analyses in Bharucha-Reid [1960, sec. 9.4D], Hashida and Kawashima [1979], and Runnenberg [1957] fail to meet this criterion and are therefore incorrect.

The arrival rate, service time distribution, and switchover time distribution can differ from queue to queue. However, it is only recently that such asymmetric cases have been analyzed exactly (there is an early approximate treatment in Yuen et al. [1972]). First, Ibe and Cheng [1986] found $E[W_i]$ for $N = 2$ by extending a method in Takagi [1985a]. Then, using a different approach, Takine et al. [1988] obtained the distribution of the message waiting time for N asymmetric queues. According to them, for a particular station, say station i, the $E[W_i]$ can be evaluated by solving a set of $O(2^N)$ linear simultaneous equations. Thus, to find $E[W_i]$'s for all the queues, we have to solve N sets of $O(2^N)$ linear equations.

Taking the limit $N \to \infty$ with ρ and R fixed at finite values in the constant-parameter case, we obtain the *continuous polling model* of Coffman and Gilbert [1986]. In this model, the server travels around a closed tour on which the service requests arrive uniformly. We then have

$$E[C] = \frac{R}{1-\rho}, \tag{8}$$

$$E[W] = \frac{R + \rho b}{2(1-\rho)}. \tag{9}$$

Note that all stable systems (whether with infinite or finite buffers) with symmetric constant parameters are reducible to the continuous polling model in the limit $N \to \infty$, since each λ becomes infinitesimal for stability, and more than one message is seldom buffered at each queue. On the basis of this idea for infinite-buffer systems, the limiting forms (8) and (9) are also given in Aminetzah [1975, chap. 6], Ferguson [1986b], and Fuhrmann and Cooper [1985a]. The difficulty of analyzing a continuous polling model with random parameters is pointed out by Coffman and Gilbert [1986].

To return to the case of a finite N, a slight variation on the above single-buffer model is the *round-robin scheduling* of services considered in Wu and Chen [1975] (approximately) and Takagi [1987b] (exactly). Here each message consists of a geometrically distributed (with mean $1/\sigma$) number of packets whose service time is a constant b. The server serves exactly one packet (if any) for each visit to a queue. There are several applications of this model:

(1) In the token-ring network, a long file is transmitted in segments, and the user must determine the mean time for the whole file to be transmitted.
(2) As a model for error-prone transmission channels, a message either departs (successful transmission) with probability σ or does not depart with probability $1 - \sigma$.
(3) In a time-sharing system with N multiprogamming levels, a processor allocates service quanta to users in a round-robin fashion.

An asymmetric model of the round-robin scheduling system was studied by Takine et al. [1987c]. Another variation is the *buffer relaxation model*, proposed by Rego and Ni [1986] and analyzed by Takine et al. [1987a], where it is assumed that a new message can be queued once the service to the previous message in that queue has been started.

A system consisting of N symmetric single-buffer queues and one infinite-buffer queue is analyzed exactly by Takine et al. [1986]. Exhaustive service is assumed for the infinite-buffer queue. An approximate analysis is given by Murata and Takagi [1987b] for a system of N single-buffer queues and one finite-buffer queue. These models may be applied to token-passing networks with a special "big" station (e.g., a file server and a bridge to an interconnected network).

2. INFINITE-BUFFER SYSTEMS

We next consider polling models in which any number of messages can be stored without loss at each queue. In this category, four main disciplines have been analyzed: exhaustive, gated, limited, and decrementing service. They differ in the instants at which the server switches from one queue to another. (These four types were already identified by Pittel [1973] for the case with two queues. More complicated disciplines for this case are considered by Gersht and Marbukh [1975], Hofri [1986a, 1986b], Yadin [1970], and Srivastava and Kashyap [1982, sec. 2.3]. Dou and Chang [1987] treated a correlated input.) In the *exhaustive* service system, the server continues to serve each queue until it is emptied. Messages arriving at the queue in service are also served in the current service period. This type of system is also called the *alternating priority discipline* (in queuing literature) or *zero-switch policy* (in vehicle traffic literature) in the case in which $N = 2$. In the *gated* service system, the server serves only those messages that are queued at the polling instant. Those that arrive at a queue during the service to that queue are set aside to be served in the next round of polling. In the *limited* (*or nonexhaustive*) scheme, each queue is served until either the queue is emptied or a specified number of messages are served, whichever occurs first. This maximum number is usually assumed to be 1, that is, at most one message is served at each visit. The term *alternating service discipline* is also used for the case in which $N = 2$. The *decrementing* (or *semiexhaustive*) scheme is defined as follows: When at least one message is found as a result of polling a queue, the service continues until the number of queued messages decreases to one less than that found at the polling instant.

For a broad class of infinite-buffer systems, including the four disciplines mentioned above and a mixture thereof, it can be shown [Kuehn 1979] that the mean cycle time is given by (8). For the mean message waiting times, a linear relationship among $E[W_i]$'s ($i = 1, 2, \ldots, N$), called the *pseudoconservation law*, has been derived by several authors [Boxma 1986; Boxma and Groenendijk 1986, 1988; Ferguson and Aminetzah 1985; Pang 1985; Pang and Donaldson 1986; Watson 1984]. For a continuous-time system of queues with a mixture of the above four disciplines, the pseudoconservation law takes the form [Boxma and Groenendijk 1986]

$$\begin{aligned} \sum_{i\in E,G} \frac{\rho_i}{\rho} E[W_i] + \sum_{i\in L} \frac{\rho_i}{\rho}\left(1-\frac{\lambda_i R}{1-\rho}\right) E[W_i] + \sum_{i\in D} \frac{\rho_i}{\rho}\left[1-\frac{\lambda_i(1-\rho_i)R}{1-\rho}\right] E[W_i] \\ = \frac{\sum_{i=1}^N \lambda_i b_i^{(2)}}{2(1-\rho)} + \frac{\Delta^2}{2R} + \frac{R(\rho-\sum_{i=1}^N \rho_i^2)}{2\rho(1-\rho)} + \frac{R\sum_{i\in G,L}\rho_i^2}{\rho(1-\rho)} - \frac{R\sum_{i\in D}\rho_i\lambda_i^2 b_i^{(2)}}{2\rho(1-\rho)}, \end{aligned} \tag{10}$$

where E, G, L, and D stand for the index sets of queues with exhaustive, gated, limited, and decrementing service disciplines, respectively. In the case of zero switchover times, our polling models become work-conserving, nonpreemptive service queues, and so this relationship reduces to Kleinrock's *conservation law* (see Kleinrock [1976, sec. 3.4]):

$$\sum_{i=1}^{N} \frac{\rho_i}{\rho} E[W_i] = \frac{\sum_{i=1}^N \lambda_i b_i^{(2)}}{2(1-\rho)}. \tag{11}$$

Although the pseudoconservation law in the form of (10) does not give expressions for the individual $E[W_i]$'s, it has several merits, which are discussed by Fuhrmann [1987]: It provides a measure of overall system performance, a demonstration of the effects of various parameters on the mean waiting times, a basis for approximation schemes, a validity check for simulations, and so on.

From (10), the $E[W_i]$'s are readily found in two extreme cases. One is the symmetric system with the same discipline for all queues, and the other is the case of extremely unbalanced traffic ($\lambda_j = 0$ for all $j \neq i$) [de Moraes and Rubin 1983]. For symmetric systems [see (14), (20), (25), and (28)], we have

$$E[W]_{\text{exhaustive}} \le E[W]_{\text{gated}} \le E[W]_{\text{limited}}, \tag{12a}$$

$$E[W]_{\text{exhaustive}} \le E[W]_{\text{decrementing}} \le E[W]_{\text{limited}}, \tag{12b}$$

and

$$E[W]_{\text{gated}} > E[W]_{\text{decrementing}}$$

for low traffic, whereas the opposite holds for heavy traffic.

There is an interesting qualitative difference between exhaustive and gated services in asymmetric systems. In exhaustive service systems, messages at heavily loaded queues experience lower delays than those at lightly loaded queues (because the server is more likely to be at one of the heavily loaded queues when a message arrives). To a lesser extent, lightly loaded queues that are close to the heavily loaded queues in the direction in which the server is moving also enjoy low delays (owing to the preceding heavily loaded queues). The opposite trends hold in gated service systems. If an exhaustive service queue is embedded in a group of gated service queues, the former plays a role similar to a heavily loaded queue in a group of lightly loaded queues all with exhaustive service. These observations are based on numerical results [Ferguson and Aminetzah 1985; Takagi 1987d]. An advantage of limited service systems is that they prevent heavily loaded queues from monopolizing the service.

A few remarks should be made with regard to *stability conditions* and the mean polling cycle time. Focusing on queue i, we see that time points when queue i is polled and all queues are empty are *regeneration points* when the system's state space consists of the queue length at each queue. Let us denote by T_i the interval between two such successive regeneration points for queue i. We say that the system is stable if and only if $\Pr[T_i < \infty] = 1$ and $E[T_i] < \infty$. Note that the regeneration cycle T_i is different from the polling cycle C_i. If the system is stable in the above sense, the mean queue length at every queue is finite, which implies that the mean polling cycle time is also finite. The converse is not necessarily true. For limited service systems, some queues may grow infinitely long whereas the polling cycle time can still be finite. The necessary conditions for stability of the four disciplines are given by Boxma [1986] as

$$\begin{array}{lll} \text{exhaustive and gated:} & \rho < 1, & \\ \text{limited:} & \rho + \lambda_i R < 1 & \text{for all } i \\ \text{decrementing:} & \rho + \lambda_i(1 - \rho_i)R < 1 & \text{for all } i. \end{array} \qquad (13)$$

According to Szpankowski and Rego [1987], these are also sufficient conditions for stability. The switchover times do not play any role in the stability condition for exhaustive and gated service systems because, at near-saturation load, the proportion of time the server spends in switchover is negligible [Eisenberg 1972]. Note that the necessary and sufficient condition for the finite mean polling cycle is given by $\rho < 1$ for the exhaustive and gated disciplines, and by $\rho_i < 1$ for all i for the decrementing discipline, whereas it is always finite for the limited discipline. Other discussions on the stability conditions and (possibly correlated) cycle times are found in Agrawal et al. [1984], Chou [1978], Rego [1985], Rego and Hughes [1986], Servi [1985], and Zhdanov and Saksonov [1979].

Let us now focus on analysis of the individual service disciplines.

2.1 Exhaustive Service Systems

The exhaustive service system with zero switchover times was studied in early literature for the case in which $N = 2$ [Avi-Itzhak et al. 1965; Conway et al. 1967, sec. 9-2; Darroch et al. 1964; Dukhovnyi 1969; Hawkes 1963; Jaiswal 1968, sec. VII.5; Miller 1964; Neuts and Yadin 1968; Stidham 1969, 1972; Takács 1968]. More recently, the asymmetric case of an arbitrary N has been analyzed in Cooper and Murray [1969] and Cooper [1970] (see also Cooper [1981, sec. 5.13] and Takagi [1987d]), where it is shown that we have to solve a set of $O(N^2)$ linear equations to compute $E[W_i]$. For the second moment $E[W_i^2]$, we have to solve $O(N^3)$ equations. As a variation of the above model, it is possible to analyze the cases of last come, first served (LCFS) or priority disciplines within each queue [Takagi 1987d].

The exhaustive service system with nonzero switchover times for $N = 2$ is analyzed by Sykes [1970] and Eisenberg [1971]. (Their assumptions on the behavior of the server when both queues are empty are different; in Sykes [1970] the server keeps switching, whereas in Eisenberg [1971] it remains at the queue last served, which may be called a "wait-and-see" policy; thus Sykes's "keep-switching" policy corresponds to the special case $N = 2$ in our cyclic service model with general N.) Exact analyses of exhaustive service systems with an arbitrary N and nonzero switchover times are available in Aminetzah [1975], Brodetskii and Vinnitskii [1973], Eisenberg [1972], Ferguson [1986b], Ferguson and Aminetzah [1975], Hashida [1972], Humblet [1978, sec. 7D], and Sarkar and Zangwill [1987] (continuous-time, asymmetric queues); in Konheim and Meister [1974] (discrete-time, symmetric queues); and in de Moraes [1981], Rubin and de Moraes [1983], and Swartz [1980] (discrete-time, asymmetric queues). For the continuous-time symmetric case, we simply have [from (10)]

$$E[W] = \frac{\delta^2}{2r} + \frac{N\lambda b^{(2)} + r(N - \rho)}{2(1 - \rho)}. \qquad (14)$$

Note that the case in which $N = 1$ is an M/G/1 queuing model with a server that goes on vacations [Cooper 1970; Doshi 1986; Scholl and Kleinrock 1983; Skinner 1967]. Explicit approximation formulas for $E[W_i]$ are proposed by Bux and Truong [1983] and Everitt [1986] (they are constructed so as to satisfy the pseudoconservation law); see (24) below. As a variation, a system in which the first message in a queue gets exceptional service is considered in Bradlow and Byrd [1987] and Ferguson [1986a] by approximation.

To date, the most efficient algorithm in the open literature for computing $E[W_i]$ (in the continuous-time system) is one given by Ferguson and Aminetzah [1985] (also in the Ph.D. dissertations of Aminetzah [1975] and Humblet [1978]), which requires us to solve a set of $O(N^2)$ linear equations. It is given by

$$E[W_i] = \frac{E[(I_i)^2]}{2E[I_i]} + \frac{\lambda_i b_i^{(2)}}{2(1-\rho_i)}, \tag{15}$$

where

$$E[I_i] = \frac{(1-\rho_i)R}{1-\rho} \tag{16}$$

and

$$\mathrm{var}[I_i] = \delta_{i-1}^2 + \frac{1-\rho_i}{\rho_i} \sum_{\substack{j=1 \\ (j \neq i)}}^{N} r_{ij} \tag{17}$$

are, respectively, the mean and variance of the *intervisit time* I_i for queue i, defined as the time starting at the instant the server leaves queue i and ending when the server visits queue i in the next polling cycle. The set $\{r_{ij};\ i, j = 1, 2, \ldots, N\}$ is computed by solving a system of linear equations:

$$r_{ij} = \frac{\rho_i}{1-\rho_i}\left(\sum_{m=i+1}^{N} r_{jm} + \sum_{m=1}^{j-1} r_{jm} + \sum_{m=j}^{i-1} r_{mj}\right), \qquad j < i; \tag{18a}$$

$$r_{ij} = \frac{\rho_i}{1-\rho_i}\left(\sum_{m=i+1}^{j-1} r_{jm} + \sum_{m=j}^{N} r_{mj} + \sum_{m=1}^{i-1} r_{mj}\right), \qquad j > i; \tag{18b}$$

$$r_{ii} = \frac{\delta_{i-1}^2}{(1-\rho_i)^2} + \frac{\lambda_i b_i^{(2)} E[I_i]}{(1-\rho_i)^3} + \frac{\rho_i}{1-\rho_i} \sum_{\substack{j=1 \\ (j \neq i)}}^{N} r_{ij}. \tag{18c}$$

Here, r_{ij} is the covariance of *station times* (introduced in Aminetzah [1975] and Ferguson and Aminetzah [1985] as *terminal service times*) for queues i and j, where, for exhaustive service systems, the station time for queue i is defined as the time interval between the successive instants the server leaves queue $i-1$ and queue i. Recently, Sarkar and Zangwill [1987] derived an algorithm to compute $\mathrm{var}[I_i]$ by solving a set of only $O(N)$ linear equations. Although their algorithm requires $O(N^3)$ arithmetic operations to compute the coefficients for the set of equations, it is a significant improvement over the previous algorithm given by (15)–(18). In contrast, Humblet [1978] and Levy [1986] discuss iterative solutions to the linear equations. Algorithms to compute $\mathrm{var}[W_i]$ are provided in

Aminetzah [1975] and Ferguson [1986b]; since they are so complicated, Ferguson uses a symbolic formula manipulation language.

We note that the Laplace–Stieltjes transform of the distribution function for the waiting time for queue i with FCFS can be expressed as

$$W_i^*(s) = \frac{1 - I_i^*(s)}{sE[I_i]} \cdot \frac{s(1 - \rho_i)}{s - \lambda_i + \lambda_i B_i^*(s)}, \tag{19}$$

where the first factor stands for the residual life of the intervisit time I_i, and the second factor is the well-known Pollaczek–Khinchin transform formula for the M/G/1 queue. This product form is an example of the M/G/1 *decomposition property* for generalized vacation models [Doshi 1986; Fuhrmann and Cooper 1985b; Wolff 1988, sec. X.5].

2.2 Gated Service Systems

The gated service system with N asymmetric queues without switchover times is considered in Cooper and Murray [1969] and Cooper [1970]. There is an early approximate treatment of the asymmetric system with switchover times by Leibowitz [1961, 1962] and Brodetsky [1973]. Exact analyses are given in Aminetzah [1975], Carsten et al. [1977], Carsten and Posner [1978], Ferguson and Aminetzah [1985], Hashida [1970, 1972], Humblet [1978, sec. 7C], Pang [1985], Pang and Donaldson [1986], and Vinnitskii and Brodetskii [1972] for continuous-time systems and in de Moraes [1981] and Rubin and de Moraes [1983] for discrete-time systems. Some of these works are integrated in Takagi [1986, sec. 5.2 and chap. 7]. For the continuous-time symmetric case, we simply have

$$E[W] = \frac{\delta^2}{2r} + \frac{N\lambda b^{(2)} + r(N + \rho)}{2(1 - \rho)}. \tag{20}$$

For asymmetric systems, we have

$$W_i^*(s) = \frac{C_i^*\,[\lambda_i - \lambda_i B_i^*(s)] - C_i^*(s)}{E[C][s - \lambda_i + \lambda_i B_i^*(s)]}, \tag{21a}$$

$$E[W_i] = \frac{(1 + \rho_i)E[(C_i)^2]}{2E[C]}, \tag{21b}$$

where $C_i^*(s)$ is the Laplace–Stieltjes transform of the distribution function for the cycle time C_i for queue i, and the mean cycle time $E[C]$ (independent of i) is given in (8). According to Ferguson and Aminetzah [1985] (after correcting some errors), we have

$$\mathrm{var}[C_i] = \frac{1}{\rho_i} \sum_{\substack{j=1 \\ (j \neq i)}}^{N} r_{ij} + \sum_{j=1}^{N} r_{ji}, \tag{22}$$

where r_{ij} is again the covariance of station time for queues i and j, but the station time for queue i for gated service systems is defined as the time interval between the successive instants when the server visits queue i and queue $i + 1$. The set $\{r_{ij};\ i, j = 1, 2, \ldots, N\}$ is given as a solution to the following $O(N^2)$ set of equations:

$$r_{ij} = \rho_i \left(\sum_{m=i}^{N} r_{jm} + \sum_{m=1}^{j-1} r_{jm} + \sum_{m=j}^{i-1} r_{mj} \right), \qquad j < i; \tag{23a}$$

$$r_{ij} = \rho_i \left(\sum_{m=i}^{j-1} r_{jm} + \sum_{m=j}^{N} r_{mj} + \sum_{m=1}^{i-1} r_{mj} \right), \qquad j > i; \tag{23b}$$

$$r_{ii} = \delta_i^2 + \lambda_i b_i^{(2)} E[C] + \rho_i \sum_{\substack{j=1 \\ (j \neq i)}}^{N} r_{ij} + \rho_i^2 \sum_{j=1}^{N} r_{ji}. \tag{23c}$$

Again, Sarkar and Zangwill [1987] give a set of $O(N)$ equations to compute $\mathrm{var}[C_i]$. An approximation to $E[W_i]$ given in Everitt [1986] is

$$E[W_i] \approx (1 \pm \rho_i) \cdot \frac{R}{2(1-\rho)} \cdot \left[1 + \frac{\rho}{\sum_{j=1}^{N} \rho_j (1 \pm \rho_j)} \cdot \left[\frac{(1-\rho)\Delta^2}{R^2} + \frac{\sum_{j=1}^{N} \lambda_j b_j^{(2)}}{R} \right] \right], \tag{24}$$

where we choose + for the gated service and − for the exhaustive service of queue i.

A system with a mixture of exhaustive and gated service disciplines is analyzed in Ozawa [1987a, 1987b] (for $N = 2$) and Takagi [1987c] (for general N). A "fully gated" discipline is defined in Bertsekas and Gallager [1987, sec. 5.3.2] as one in which each queue is inspected before the switchover time to that queue (our "gated" is called "partially gated"). This is based on the assumption that the switchover time is used to schedule the next service period. However, the analysis of this system can be reduced to that of our gated service system. Recently, Rubin and Zsai [1987a, 1987b] considered a gated polling for a backbone control with a local multiple-access policy.

2.3 Limited and Decrementing Service Systems

Analysis of limited and decrementing service systems is more difficult than analysis of exhaustive and gated service systems. In a general asymmetric system with exhaustive or gated service, $E[W_i]$'s can be simply computed from the values of the mean and second moment of the message service times and switchover times. However, this does not seem to be the case for limited and decrementing service systems.

For the case of $N = 2$, limited service systems are studied in Cohen and Boxma [1981] (symmetric, without switchover times), in Eisenberg [1979] and Cohen and Boxma [1983, secs. III.2 and IV.1] (asymmetric, without switchover times), in Iisaku et al. [1981a] and Boxma [1984] (symmetric, with switchover times), and in Boxma and Groenendijk [1987] (asymmetric, with switchover times). Boxma [1986] describes how the analysis can be formulated as a *Riemann-Hilbert boundary-value problem.* The exact $E[W]$ for an arbitrary N in the symmetric system is given in Fuhrmann [1985], Nomura and Tsukamoto [1978], and Takagi [1985b]. It can now be readily obtained from (10) as

$$E[W] = \frac{\delta^2}{2r} + \frac{N\lambda b^{(2)} + r(N + \rho) + N\lambda\delta^2}{2(1 - \rho - N\lambda r)}. \tag{25}$$

There are a number of approximate solutions proposed for asymmetric systems [Arndt and Sulanke 1984; Boxma and Meister 1986, 1987; Groenendijk 1988;

Hashida and Ohara 1972; Kuehn 1979; Kurosawa and Tsujii 1981; Mase 1977; Srinivasan 1986; Wang 1986]. Diffusion approximations are used in Fischer [1977] and Kimura and Takahashi [1986]. The one in Boxma and Meister [1986] is given by

$$E[W_i] \approx \frac{1 - \rho + \rho_i}{1 - \rho - \lambda_i R} \cdot \frac{1 - \rho}{(1 - \rho)\rho + \sum_{j=1}^{N} \rho_j^2} \cdot \left[\frac{\rho}{2(1 - \rho)} \sum_{j=1}^{N} \lambda_j b_j^{(2)} + \frac{\rho \Delta^2}{2R} + \frac{R}{2(1 - \rho)} \sum_{j=1}^{N} \rho_j (1 + \rho_j)\right], \tag{26}$$

which is constructed so as to satisfy the pseudoconservation law (10). (Note that in both (24) and (26), only the first factor depends on i, a point often used for approximation formulas.) This implies that (26) is exact in the case of $N = 1$, in the extremely unbalanced traffic case ($\lambda_j = 0$ for all $j \neq i$), and in the symmetric case. The case $N = 1$ is an M/G/1 model in which the server takes a vacation after each service; its solution is given in Skinner [1967]. Approximation (26) is used by Murata and Takagi [1987a] in local-area-network modeling for the media access control (MAC) layer model, together with a closed queuing network for the transport layer model.

A symmetric system with *feedback* is solved in Takagi [1987a] (whose error is corrected by de Moraes [1987]). Here, after a message has been served, it leaves the system with probability σ or rejoins the queue with probability $1 - \sigma$. (This is an infinite-buffer version of round-robin scheduling.) Applications of this model are shown in Halfin [1975] and Kuehn [1981]. Groenendijk [1988] and Ozawa [1987b] independently found the mean waiting times for a system of two queues with exhaustive and limited discipline that manifest their dependence on the distributions of the switchover times. (Skinner [1967] and Srinivasan and Lee [1987] solved the same problem with a constant switchover time from the exhaustive service queue to the limited service queue.) An interesting performance measure called *access delay* has been proposed by Pack and Whitaker [1976]. It is defined as the time from the instant at which a message reaches the head of the queue until the instant its service is started. The access delay at a queue is basically a measure of how much the other queues use the server, whereas the message waiting time includes both the access delay and the congestion at the queue.

A system of N queues where at most K_i messages are served at queue i is treated approximately by several authors, including Fuhrmann [1985, 1987] (with an upper bound on $E[W_i]$), Fuhrmann and Wang [1987, 1988], Hashida [1981, chap. 5] (extending his exact result for $N = 1$), Hayes and Jalali-Nadushan [1986], Iisaku et al. [1980, 1981b], Konheim [1976] (the term *chaining* is used), Nair [1976] ($N = 2$), Noguchi et al. [1984], and Takagi [1986, chap. 6]. We may distinguish two separate cases on the basis of whether the K_i messages include those that arrive during the service period (*E-limited*) or not (*G-limited*). For a symmetric E-limited service system, Furhmann [1985] derives an upper bound

$$E[W]_{\text{E-limited}} + b \leq \frac{1 - \rho}{1 - \rho - \lambda R/K} (E[W]_{\text{exhaustive}} + b), \tag{27}$$

where $E[W]_{\text{exhaustive}}$ is given by (14), and then argues that this bound is so tight that it can be used as an approximation [Fuhrmann 1987; Fuhrmann and Wang 1987, 1988]. Note that the equality holds in (27) for the cases of $K = 1$ (limited service) and $K = \infty$ (exhaustive service). Furhmann [1987] also establishes a similar upper bound for a symmetric G-limited service system and other upper bounds on weighted sums of $E[W_i]$'s for asymmetric E-limited and G-limited service systems. Recently, Coffman et al. [1988] have considered a system of two queues in which the service at each queue continues until either it has been emptied or an exponentially distributed time has expired, whichever occurs first; since they assume exponentially distributed service times, the remainder of a truncated service time is also exponentially distributed, which makes the system a two-dimensional birth-and-death process.

The decrementing service discipline was first introduced in Pittel [1973] and was independently analyzed in Takagi [1984], which gives an exact $E[W]$ for a symmetric system of N queues:

$$E[W] = \frac{\delta^2}{2r} + \frac{N\lambda b^{(2)}(1 - \lambda r) + (r + \lambda\delta^2)(N - \rho)}{2[1 - \rho - \lambda r(N - \rho)]}. \tag{28}$$

Since then, Cohen [1988] has analyzed an asymmetric system of two queues without switchover times, and Boxma [1986] has derived the pseudoconservation law (10).

3. NONCYCLIC POLLING AND PRIORITY

We can consider two types of noncyclic order of polling, one *deterministic*, the other *probabilistic.*

An example of deterministic order is the *scan*, named after a seek policy in the moving-arm disk device. Here, the order of polling is given by $1 \to 2 \to \cdots \to N - 1 \to N \to N \to N - 1 \to \cdots \to 2 \to 1 \to 1 \to \cdots$. It is analyzed in Coffman and Hofri [1982, 1986], Swartz [1982], and Takagi and Murata [1986]. Coffman and Gilbert [1987] consider the continuous polling model for scan service. Another instance of deterministic order is when frequently visited queues appear more often in the polling order table than others, thus receiving preferential treatment [Alford and Muntz 1975; Baker 1986, chap. 2; Baker and Rubin 1987; Eisenberg 1972; Kruskal 1969; Mapp and Manfield 1986]. As a result of these studies, the problem of any deterministic polling order can be solved in exactly the same way as for ordinary cyclic-order systems.

An example of probabilistic order is *Bernoulli scheduling*, which is parameterized by a vector $(p_1, p_2, \ldots, p_N)$ [Servi 1986]. Here, if queue i is empty when its service to a message has been completed, then queue $i + 1$ is polled. Otherwise, queue i is served again with probability p_i, and queue $i + 1$ is polled with probability $1 - p_i$. A switchover time is required to change from queue i to queue $i + 1$. Thus, the cases in which $p_i = 1$ and $p_i = 0$ can be reduced to the exhaustive and limited service disciplines, respectively, at queue i. The Bernoulli scheduling system has been solved only approximately. Fukagawa et al. [1986a] approximately analyzed a model in which the server starts service only when more than a certain number of messages are found at the server's visit. Another probabilistic case is one in which, after the completion of service at any queue, the next polled queue is queue j with probability p_j, where $\sum_{j=1}^{N} p_j = 1$; it has been considered by Suda et al. [1980] and Levy [1984, chap. 4], who use the term *random polling.* The random polling system can be analyzed exactly.

There are several ways to give priority to some queues over others, namely, higher frequency of visits, exhaustive service (gated or limited service to others), and more buffers. They may be called *queue priority* schemes. For example, a system of one exhaustive service queue and $N - 1$ limited service queues is (approximately) analyzed in Manfield [1983, 1985] and Srinivasan and Lee [1987]. Other schemes in which messages are assigned external priorities can be called *message priority* schemes. There are various ways to handle prioritized messages. The case in which the priority discipline applies only within each queue is easy to analyze [Takagi 1987d] since we may use the results for an M/G/1 priority system with vacation. Another case in which the priority discipline is enforced on the whole system has applications to the priority mode operation of the token ring network [Bux et al. 1983] as analyzed in Baker [1986, chap. 3], Gianini and Manfield [1988], Nishida et al. [1983, 1986], Saydam and Sethi [1985], Shen et al. [1985], and Yamamoto et al. [1984]. A priority scheme in which only the message with the highest priority is served at a visit to a queue is considered in Fukagawa et al. [1986b, 1987], Karvelas and Leon-Garcia [1986], and Kimura and Takahashi [1987b].

4. APPLICATIONS

We have already mentioned early applications of a single-buffer polling model to a patrolling machine repairman problem and of a two-queue exhaustive-service model to a vehicle traffic problem. Other applications include production systems [Sarkar and Zangwill 1987; Schay 1962], passenger transportation on a circular route [Dukhovnyy 1979], and mail delivery systems [Nahmias and Rothkopf 1984]. More applications, however, are drawn from the field of computer communication networks.

Half-duplex is a mode of transmitting data between two parties on a shared communication line. Transmission is possible in either direction but not in both directions simultaneously. To analyze this system, Sykes [1969a, 1969b] used his result for two queues with alternating priority discipline.

Roll-call polling is a technique often used in a configuration in which geographically dispersed terminals are connected to a central processor by communication lines in tree topology [Hayes 1984, chap. 7; Martin 1967, chap. 20; Schwartz 1977, chap. 12, 1987, chap. 8]. The processor has a polling sequence table according to which it interrogates each terminal. The addressed terminal then transmits all waiting messages to the processor. When the transmission is completed, polling of the next terminal is initiated. *Hub polling*, on the other hand, is suitable for loop topology. The central processor initiates polling by interrogating the terminal at the end of the loop. This terminal transmits its waiting messages, to which it appends a polling message for the next upstream terminal. The latter terminal similarly adds its own messages followed by another polling message, and so on. At the completion of the polling cycle, the central processor regains control. For an example of analysis, see Schwartz [1977] and Hammond and O'Reilly [1986, chap. 7]. Clearly, the overhead of switchover is greater for roll-call polling than for hub polling. Polling is also proposed for packet radio systems [Tobagi and Kleinrock 1976].

A discrete-time polling model is used for analysis of the *Minislotted Alternating Priorities* (*MSAP*) multiple-access scheme [Kleinrock and Scholl 1980; Scholl 1976]. This scheme takes advantage of the broadcasting nature of packet radio systems to realize a distributed-access control. Each terminal gets the transmission rights for one minislot (with the duration set equal to the signal propagation delay) after the service period for the preceding terminal. Thus the minislot accounts for a constant switchover time in the polling model.

In local-area networks, two token passing schemes, *token ring* and *token bus*, have been analyzed by using polling models [Berry and Chandy 1983; Bux 1981; Colvin and Weaver 1986; Davies and Ghani 1983; de Moraes and Rubin 1984; Sethi and Saydam 1985]. For a general survey, see Bux [1984, 1987]; for more detailed analysis, see Hammond and O'Reilly [1986, chap. 8]. Assuming that N = number of stations, P = signal propagation delay (e.g., 5×10^{-6}s/km, or two-thirds of the speed of light), l = cable length (kilometers), C = line speed (bits per second), L_R = bit latency (bits) at each station for the token ring, and L_B = bit latency (bits) at each station for the token bus, the total switchover time R (seconds) is given by

$$R = \begin{cases} Pl + \dfrac{NL_R}{C}, & \text{token ring;} \\ NPl + \dfrac{NL_B}{C}, & \text{token bus.} \end{cases} \tag{29}$$

Thus, the performance of token ring is clearly better than that of token bus. In the token ring network, the service times used in the previous formulas differ according to the modes of operation (single token, multiple token, and single message) [Hammond and O'Reilly 1986]. Multiple-ring networks have been considered by Carsten and Posner [1978], Hashimoto and Ohmae [1986], and Kanada and Ikeda [1987].

Other demand-based channel access schemes in local-area networks have also been analyzed by using polling models. Examples include Newhall loops [Carsten et al. 1977; Carsten and Posner 1978], Unidirectional Broadcast System—Round Robin (UBS-RR) [Tobagi et al. 1983], Expressnet [Tobagi et al. 1983; Tobagi and Fine 1983], Fasnet [Heyman 1983; Tobagi and Fine 1983], round-robin access in cable networks [Gold and Franta 1983], PBX [Appleton and Peterson 1986], Hybrinet [Ibe and Chen 1985], helical window token ring [Kschischang and Molle 1987], Fiber Distributed Data Interface (FDDI) token ring [Sevcik and Johnson 1987], and other timed token schemes [Goyal and Dias 1988; Yue 1987]. A number of demand access schemes are surveyed from the point of view of performance in Tropper [1981], Fine and Tobagi [1984], and Ibe [1987].

5. FUTURE RESEARCH TOPICS

From this survey, it may seem that most basic polling models have already been solved. By "solved" we mean that the distribution of the message waiting times can be expressed in terms of given system parameters. For single-buffer systems, we already have the distributions of message waiting times [Takine et al. 1988]. For exhaustive service and gated service systems with infinite buffers, the mean waiting times can be obtained by solving $O(N)$ linear equations [Sarkar and Zangwill 1987]. For limited service systems, we have the mean waiting times only for N symmetric queues [Takagi 1985b] and for two asymmetric queues [Boxma and Groenendijk 1987]; it appears unlikely that exact solutions will be found for asymmetric systems of N (> 2) queues. Thus, in the future it remains for us to consider more elaborate models and practical models.

One such model is a system with finite buffers. Queue i can store up to L_i messages, where $1 < L_i < \infty$. There are some approximate treatments [Kimura and Takahashi 1987a; Raith and Tran-Gia 1986; Tran-Gia 1988; Tran-Gia and Raith 1986], but further study may be done by extending the techniques in, for example, Lee [1983, 1984] for single queue M/G/1/L systems with vacations. The case of $L_i = 2$ has been studied by Robillard [1974]. The study of systems with multiple servers [Kamal and Hamacher 1986; Morris and Wang 1984; Raith 1985] is still in its infancy; assumptions on the relative movements of multiple servers depend on the applications. We have not obtained exact results for the waiting times in message-priority polling models. Also of interest are systems, such as FDDI, in which the number of messages allowed to be served in a polling cycle depends on the length of the previous cycle or on the number of messages served in the previous cycle.

It is interesting to note that the analysis for the case with zero switchover times is usually more difficult than that for the case with nonzero switchover times. To compute $E[W_i]$'s in the asymmetric infinite-buffer exhaustive and/or gated service systems, we have a set of $O(N)$ equations for systems with nonzero switchover times and a set of $O(N^2)$ equations for systems with zero switchover times. There has not been an analysis of single-buffer systems with zero switchover times. A technical difficulty in (stable) systems with zero switchover times is that the mean cycle time shrinks to zero because the server rotates with an infinite speed while the system is empty.

Asymmetric limited and decrementing service systems involve mathematical difficulty. To discover whether the Riemann–Hilbert formulation can be extended to more than two dimensions is a challenging task. This is an example of a problem whose solution would contribute a great deal to the fields of mathematics and engineering.

ACKNOWLEDGMENTS

I am grateful to the following people who commented on the early versions and/or helped organize the scope of the present subject through intensive discussions: O. J. Boxma (Centre for Mathematics and Computer Science, The Netherlands), R. B. Cooper (Florida Atlantic University, U.S.A.), L. F. M. de Moraes (Instituto Militar de Engenharia-RJ, Brazil), M. J. Ferguson (INRS-Télécommunications, Canada), and Y. Takahashi (Kyoto University, Japan).

An early version of this paper was presented at the Third International Conference on Data Communication Systems and Their Performance, Rio de Janeiro, June 22–25, 1987.

REFERENCES

AGRAWAL, S. C., BUZEN, J. P., AND THAREJA, A. K. 1984. A unified approach to scan time analysis of token rings and polling networks. *Perform. Eval. Rev. 12*, 3 (Aug.) 176–185. Also, Tech. Rep. BGS Systems, Waltham, Mass., 1983.

ALFORD, M., AND MUNTZ, R. R. 1975. Queueing models for polled multiqueues. In *Proceedings of the 4th Texas Conference on Computing Systems*, pp. 5B-2.1–2.5.

AMINETZAH, Y. 1975. An exact approach to the polling system. Ph.D. dissertation, Dept. of Electrical Engineering, McGill Univ., Montreal, Quebec.

APPLETON, J. M., AND PETERSON, M. M. 1986. Traffic analysis of a token ring PBX. *IEEE Trans. Commun. COM-34*, 5 (May), 417–422.

ARNDT, K., AND SULANKE, H. 1984. A queueing system with relative and cyclic priorities. *Elektron. Inf. Kybern. 20*, 7/9 (Sept.), 423–425.

AVI-ITZHAK, B., MAXWELL, W. L., AND MILLER, L. W. 1965. Queueing with alternating priorities. *Oper. Res. 13*, 2 (Mar.–Apr.), 306–318.

BAKER, J. E. 1986. Multiple-access communications with prioritized service. Ph.D. dissertation, Dept. of Electrical Engineering, Univ. of California, Los Angeles, Calif.

BAKER, J. E., AND RUBIN, I. 1987. Polling with a general-service order table. *IEEE Trans. Commun. COM-35*, 3 (Mar.), 283–288. Also in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '86)* (Houston, Tex., Dec. 1–4, 1986) pp. 6–11.

BERRY, R. E., AND CHANDY, K. M. 1983. Performance models of token ring local area networks. *Perform. Eval. Rev.* Special issue, *Proceedings of the 1983 ACM SIGMETRICS Conference on Measurement and Modeling of Computer Systems* (Aug.), pp. 266–274.

BERTSEKAS, D., AND GALLAGER, R. 1987. *Data Networks.* Prentice-Hall, Englewood Cliffs, N.J.

BHARUCHA-REID, A. T. 1960. *Elements of the Theory of Markov Processes and Their Applications.* McGraw-Hill, New York.

BOXMA, O. J. 1984. Two symmetric queues with alternating service and switching times. In *Performance '84*, E. Gelenbe, Ed. Elsevier North-Holland, Amsterdam, pp. 409–431.

BOXMA, O. J. 1986. Models of two queues: A few new views. In *Teletraffic Analysis and Computer Performance Evaluation*, O. J. Boxma, J. W. Cohen, and H. C. Tijms, Eds. Elsevier North-Holland, Amsterdam, pp. 75–98.

BOXMA, O. J., AND GROENENDIJK, W. P. 1986. Pseudo-conservation laws in cyclic-service systems. Rep. OS-R8606, Centre for Mathematics and Computer Science, Amsterdam.

BOXMA, O. J., AND GROENENDIJK, W. P. 1987. Two queues with alternating service and switching times. Rep. OS-R8712, Centre for Mathematics and Computer Science, Amsterdam.

BOXMA, O. J., AND GROENENDIJK, W. P. 1988. Waiting times in discrete-time cyclic-service systems. *IEEE Trans. Commun. COM-36*, 2 (Feb.), 164–170.

BOXMA, O. J., AND MEISTER, B. 1986. Waiting-time approximations for cyclic-service systems with switchover times. *Performance '86 and ACM SIGMETRICS 1986, Perform. Eval. Rev. 14*, 1, 254-262. Also in *Perform. Eval. 7*, 4 (Nov.), 299–308, 1987.

BOXMA, O. J., AND MEISTER, B. 1987. Waiting-time approximations in multi-queue systems with cyclic service. *Perform. Eval. 7*, 1 (Feb.), 59–70.

BRADLOW, H. S., AND BYRD, H. F. 1987. Mean waiting time evaluation of packet switches for centrally controlled PBX's. *Perform. Eval. 7*, 4 (Nov.), 309–327.

BRODETSKY, G. L. 1973. Analysis of a model with cyclic servicing of packages of orders. *Eng. Cybern. 11*, 3, 412–422.

BRODETSKY, G. L., AND VINNITSKII, V. P. 1973. Some characteristics of systems with cyclic data processing. *Cybernetics 9*, 3 (May–June), 407–413.

BUX, W. 1981. Local-area subnetworks: A performance comparison. *IEEE Trans. Commun. COM-29*, 10 (Oct.), 1465–1473. Reprinted in *Advances in Local Area Networks*, K. Kümmerle, F. A. Tobagi, and J. O. Limb, Eds. IEEE Press, N.Y., 1987, pp. 363–382.

BUX. W. 1984. Performance issues in local-area networks. *IBM Syst. J. 23*, 4, 351–374.

BUX, W. 1987. Modeling token ring networks—A survey. IBM Res. Rep. RZ 1615, IBM Zürich Research Laboratory, Rüschlikon, Switzerland.

BUX, W., AND TRUONG, H. L. 1983. Mean-delay approximation for cyclic service queueing systems. *Perform. Eval. 3*, 3 (Aug.), 187–196. Also, Token-ring performance: Mean-delay approximation, In *Proceedings of the 10th International Teletraffic Congress*, pp. 3.3.3-1–3.3.3-7.

BUX, W., CLOSS, F. H., KUEMMERLE, K., KELLER, H. J., AND MUELLER, H. R. 1983. Architecture and design of a reliable token-ring network. *IEEE J. Selected Areas Commun. SAC-1*, 5 (Nov.), 756–765. Reprinted in *Advances in Local Area Networks*, K. Kümmerle, F. A. Tobagi, and J. O. Limb, Eds. IEEE Press, N.Y., 1987, pp. 67–80.

CARSTEN, R. T., AND POSNER, M. J. M. 1978. Simplified statistical models of single and multiple Newhall loops. *1978 National Telecommunications Conference* (Birmingham, Ala., Dec. 3–6), pp. 44.5.1–44.5.7, IEEE, New York.

CARSTEN, R. T., NEWHALL, E. E., AND POSNER, M. J. M. 1977. A simplified analysis of scan times in an asymmetrical Newhall loop with exhaustive service. *IEEE Trans. Commun. COM-25*, 9 (Sept.), 951–957.

CHOU, W. 1978. Terminal response time on polled teleprocessing networks. In *Proceedings of Computer Network Symposium* (Dec.).

CHU, W. W., AND KONHEIM, A. G. 1972. On the analysis and modeling of a class of computer communication systems. *IEEE Trans. Commun. COM-20*, 3 (June), Part 2, 645–660.

COFFMAN, E. G. JR., AND GILBERT, E. N. 1986. A continuous polling system with constant service times. *IEEE Trans. Inf. Theory IT-33*, 4 (July), 584–591.

COFFMAN, E. G. JR., AND GILBERT, E. N. 1987. Polling and greedy servers. *Queueing Syst. 2*, 2 (July), 115–145.

COFFMAN, E. G., JR., AND HOFRI, M. 1982. On the expected performance of scanning disks. *SIAM J. Comput. 11*, 1 (Feb.), 60–70.

COFFMAN, E. G., JR., AND HOFRI, M. 1986. Queueing models of secondary storage devices. *Queueing Syst. 1*, 2 (Sept.), 129–168.

COFFMAN, E. G., JR., FAYOLLE, G. AND MITRANI, I. 1988. Two queues with alternating service periods. In *Performance '87*, P.-J. Courtois and G. Latouche, Eds. Elsevier North-Holland, Amsterdam, pp. 227–239.

COHEN, J. W. 1988. A two-queue model with semi-exhaustive alternating service. In *Performance '87*, P.-J. Courtois and G. Latouche, Eds. Elsevier North-Holland, Amsterdam, pp. 19–37.

COHEN, J. W., AND BOXMA, O. J. 1981. The M/G/1 queue with alternating service formulated as a Riemann-Hilbert problem. In *Performance '81*, F. J. Kylstra, Ed. North-Holland Publishing Co., Amsterdam, pp. 181–199.

COHEN, J. W., AND BOXMA, O. J. 1983. *Boundary Value Problems in Queueing System Analysis.* Mathematics Studies 79. North-Holland Publishing Co., Amsterdam.

COLVIN, M. A., AND WEAVER, A. C. 1986. Performance of single access classes on the IEEE 802.4 token bus. *IEEE Trans. Commun. COM-34*, 12 (Dec.), 1253–1256.

CONWAY, R. W., MAXWELL, W. L., AND MILLER, L. W. 1967. *Theory of Scheduling.* Addison-Wesley, Reading, Mass.

COOPER, R. B. 1970. Queues served in cyclic order: Waiting times. *Bell Syst. Tech. J. 49*, 3 (Mar.), 399–413.

COOPER, R. B. 1981. *Introduction to Queueing Theory.* 2d ed. Elsevier North-Holland, New York.

COOPER, R. B., AND MURRAY, G. 1969. Queues served in cyclic order. *Bell Syst. Tech. J. 48*, 3 (Mar.), 675–689.

DARROCH, J. N., NEWELL, G. F., AND MORRIS, R. W. J. 1964. Queues for a vehicle-actuated traffic light. *Oper. Res. 12*, 6 (Nov.–Dec.), 882–895.

DAVIES, P., AND GHANI, F. A. 1983. Access protocols for an optical-fibre ring network. *Comput. Commun. 6*, 4 (Aug.), 185–191.

DE MORAES, L. F. M. 1981. Message queueing delays in polling schemes with applications to data communication networks. Ph.D. dissertation, UCLA-ENG-8106, Dept. of System Science, School of Engineering and Applied Science, Univ. of California, Los Angeles, Calif.

DE MORAES, L. F. M. 1987. Comments on "Analysis and applications of a multiqueue cyclic service system with feedback." Tech. Rep., Instituto Militar de Engenharia-RJ, Rio de Janeiro, Brazil.

DE MORAES, L. F. M., AND RUBIN, I. 1983. On the performance of gated and exhaustive polling under unbalanced traffic. In *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '83)* (San Diego, Calif., Nov 28–Dec. 1). IEEE, New York, pp. 1054–1059.

DE MORAES, L. F. M., AND RUBIN, I. 1984. Analysis and comparison of message queueing delays in token-rings and token-buses local area networks. In *Proceedings of the IEEE International Conference on Communications (ICC '84)* (Amsterdam, May 14–17). IEEE, New York, pp. 130–135.

DOSHI, B. T. 1986. Queueing systems with vacations—A survey. *Queueing Syst. 1*, 1 (June), 29–66.

DOU, C., AND CHANG, J.-F. 1987. Polling of two queues with synchronized correlated inputs and zero walktime. In *Proceedings of IEEE INFOCOM '87* (San Francisco, Mar. 31–Apr. 2). IEEE, New York, pp. 952–961.

DUKHOVNYI, I. M. 1969. A single-channel queueing system with alternating priority. *Prob. Inf. Transm. 5*, 2 (Apr.–June), 47–55.

DUKHOVNYY, I. M. 1979. An approximate model of motion of urban passenger transportation over annular routes. *Eng. Cybern. 17*, 1, 161–162.

EISENBERG, M. 1971. Two queues with changeover times. *Oper. Res. 19*, 2 (Mar.–Apr.), 386–401.

EISENBERG, M. 1972. Queues with periodic service and changeover times. *Oper. Res. 20*, 2 (Mar.–Apr.), 440–451.

EISENBERG, M. 1979. Two queues with alternating service. *SIAM J. Appl. Math. 36*, 2 (Apr.), 287–303.

EVERITT, D. 1986. Simple approximations for token rings. *IEEE Trans. Commun. COM-34*, 7 (July), 719–721.

FERGUSON, M. J. 1986a. Mean waiting time for a token ring with station dependent overheads. In *Local Area & Multiple Access Networks*, R. L. Pickholtz, Ed. Computer Science Press, Rockville, Md., chap. 3, pp. 43–67. Also Mean waiting time for a token ring with nodal dependent overheads. In *11th International Teletraffic Congress*, M. Akiyama, Ed. Elsevier North-Holland, Amsterdam, pp. 4.2A-1-1–4.2A-1-7.

FERGUSON, M. J. 1986b. Computation of the variance of the waiting time for token rings. *IEEE J. Selected Areas Commun. SAC-4*, 6 (Sept.), 775–782.

FERGUSON, M. J., AND AMINETZAH, Y. J. 1985. Exact results for nonsymmetric token ring systems. *IEEE Trans. Commun. COM-33*, 3 (Mar.), 223–231.

FINE, M., AND TOBAGI, F. A. 1984. Demand assignment multiple access schemes in broadcast bus local area networks. *IEEE Trans. Comput. C-33*, 12 (Dec.), 1130–1159.

FISCHER, M. J. 1977. Analysis and design of loop service systems via a diffusion approximation. *Oper. Res. 25*, 2 (Mar.–Apr.), 269–278.

FUHRMANN, S. W. 1985. Symmetric queues served in cyclic order. *Oper. Res. Lett. 4*, 3 (Oct.), 139–144.

FUHRMANN, S. W. 1987. Inequalities for cyclic service systems with limited service disciplines. In *Proceedings of IEEE/IEICE Global Telecommunications Conference 1987* (Tokyo, Nov. 15–18). IEEE, New York, pp. 182–186.

FUHRMANN, S. W., AND COOPER, R. B. 1985a. Application of decomposition principle in M/G/1 vacation model to two continuum cyclic queueing models—Especially token-ring LANs. *AT&T Tech. J. 64*, 5 (May–June), 1091–1099.

FUHRMANN, S. W., AND COOPER, R. B. 1985b. Stochastic decompositions in the M/G/1 queue with generalized vacations. *Oper. Res. 33*, 5 (Sept.–Oct.), 1117–1129.

FUHRMANN, S. W., AND WANG, Y. T. 1987. Analysis of cyclic service systems with limited service: Bounds and approximations. Tech. Paper, AT&T Bell Laboratories, Holmdel, N.J.

FUHRMANN, S. W., AND WANG, Y. T. 1988. Mean waiting time approximations of cyclic service systems with limited service. In *Performance '87*, P.-J. Courtois and G. Latouche, Eds. Elsevier North-Holland, Amsterdam, pp. 253–265.

FUKAGAWA, Y., YANARU, T., AND YOSHIDA, S. 1986a. Multiqueues system with controlled and cyclic service. *Trans. Inst. Electron. Commun. Eng. Jpn J69-A*, 1 (Jan.), 25–31 (in Japanese).

FUKAGAWA, Y., YANARU, T., AND YOSHIDA, S. 1986b. A cyclic time of multiqueues with priority service rule. *Trans. Inst. Electron. Commun. Eng. Jpn J69-A*, 1 (Jan.), 159–160 (in Japanese).

FUKAGAWA, Y., YANARU, T., AND YOSHIDA, S. 1987. An approximate analysis for a multiqueue with a non-preemptive priority and cyclic service. *Trans. Inst. Electron. Inf. Commun. Eng. J70-A*, 9 (Sept.), 1351–1354 (in Japanese).

GERSHT, A. M., AND MARBUKH, V. V. 1975. Queueing systems with readjustment. *Eng. Cybern. 13*, 4, 55–65.

GIANINI, J., AND MANFIELD, D. 1988. Cyclic multiqueue systems with two priority classes and exhaustive service. In *Data Communication Systems and Their Performance*. L. F. M. de Moraes, E. de Souza e Silva and L. F. G. Soares, Eds. Elsevier North-Holland, Amsterdam, pp. 511–526. *Perform. Eval. 8*, 2 (Apr.), 93–115.

GOLD, Y. I., AND FRANTA, W. R. 1983. An efficient collision-free protocol for prioritized access-control of cable or radio channels. *Comput. Networks 7*, 2 83–98.

GOYAL, A., AND DIAS, D. 1988. Performance of priority protocols on high speed token ring networks. In *Data Communication Systems and Their Performance*, L. F. M. de Moraes, E. de Souza e Silva, and L. F. G. Soares, Eds. Elsevier North-Holland, Amsterdam, pp. 25–34.

GROENENDIJK, W. P. 1988. Waiting-time approximations for cyclic-service systems with mixed service strategies. In *Proceedings of the 12th International Teletraffic Congress* (Torino, Italy).

HALFIN, S. 1975. An approximate method for calculating delays for a family of cyclic-type queues. *Bell Syst. Tech. J. 54*, 10 (Dec.), 1733–1754.

HAMMOND, J. L., AND O'REILLY, P. J. P. 1986. *Performance Analysis of Local Computer Networks*, Addison-Wesley, Reading, Mass.

HASHIDA, O. 1970. Gating multiqueues served in cyclic order. *Syst. Comput. Controls 1*, 1, 1–8. Also in *Trans. Inst. Electron. Commun. Eng. Jpn 53-A*, 1 (Jan.), 43–50 (in Japanese).

HASHIDA, O. 1972. Analysis of multiqueue. *Rev. Elect. Commun. Lab. 20*, 3–4 (Mar.–Apr.), 189–199.

HASHIDA, O. 1981. A study of multiqueues in communication control. Ph.D. dissertation, Univ. of Tokyo, Tokyo (in Japanese).

HASHIDA, O., AND KAWASHIMA, K. 1979. Analysis of a polling system with single user at each terminal. *Trans. Inst. Electron. Commun. Eng. Jpn J62-B*, 12 (Dec.), 1097–1102 (in Japanese). Also in *Rev. Electric. Commun. Lab. 29*, 3–4 (Mar.–Apr.), 245–253, 1981.

HASHIDA, O., AND OHARA, K. 1972. Line accommodation capacity of a communication control unit. *Rev. Electr. Commun. Lab. 20*, 3–4 (Mar.–Apr.), 231–239.

HASHIMOTO, K., AND OHMAE, Y. 1986. Performance on the multiple ring networks. *Trans. Inf. Process. Soc. Jpn 27*, 10 (Oct.), 1003–1010 (in Japanese).

HAWKES, A. G. 1963. Queueing at traffic intersection. In *Proceedings of the 2nd International Symposium on the Theory of Traffic Flow* (London).

HAYES, J. F. 1981. Local distribution in computer communications. *IEEE Commun. Mag. 19*, 2 (Mar.), 6–14. Also in *Advances in Local Area Networks*, K. Kümmerle, F. A. Tobagi, and J. O. Limb, Eds. IEEE Press, New York, 1987, pp. 302–317.

HAYES, J. F. 1984. *Modeling and Analysis of Computer Communications Networks*. Plenum Press, New York.

HAYES, A., AND JALALI-NADUSHAN, A. 1986. Numerical solution to limited service polling models. *Comput. Comm. 9*, 4 (Aug.), 171–176.

HAYES, J. F., AND SHERMAN, D. N. 1972. A study of data multiplexing techniques and delay performance. *Bell Syst. Tech. J. 51*, 9 (Nov.), 1983–2011.

HEYMAN, D. P. 1983. Data-transport performance analysis of Fasnet. *Bell Syst. Tech. J. 62*, 8 (Oct.), 2547–2560.

HOFRI, M. 1986a. Queueing systems with a procrastinating server. *Performance '86 and ACM SIGMETRICS 1986, Perform. Eval. Rev. 14*, 1 (May), 245–253.

HOFRI, M. 1986b. Two queues and one server with threshold switching. In *Teletraffic Analysis and Computer Performance Evaluation*, O. J. Boxma, J. W. Cohen, and H. C. Tijms, Eds. Elsevier North-Holland, Amsterdam, pp. 409–428.

HUMBLET, P. A. 1978. Source coding for communication concentrators. Rep. ESL-R-798, Electronic Systems Laboratory, Massachusetts Institute of Technology, Cambridge, Mass.

IBE, O. C. 1987. Performance comparison of explicit and implicit token-passing networks. *Comput. Commun. 10*, 2 (Apr.), 59–69.

IBE, O. C., AND CHEN, P.-Y. 1985. Hybrinet: A hybrid bus and token ring network. *Comput. Networks ISDN Syst. 9*, 3 (Mar.), 215–221.

IBE, O. C., AND CHENG, X. 1986. Analysis of polling systems with single-message buffers. In *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '86)* (Houston, Tex., Dec. 1–4). IEEE, New York, pp. 939–943.

IISAKU, S., MIKI, N., NAGAI, N., AND HATORI, K. 1980. A consideration of multiqueues with nonexhaustive and cyclic service. *Trans. Inst. Electron. Commun. Eng. Jpn J63-B*, 11 (Nov.), 1102–1109 (in Japanese).

IISAKU, S., MIKI, N., NAGAI, N., AND HATORI, K. 1981a. Two queues with alternating service and walking time. *Trans. Inst. Electron. Commun. Eng. Jpn J64-B*, 4 (Apr.), 342–343 (in Japanese).

IISAKU, S., MIKI, N., NAGAI, N., AND HATORI, K. 1981b. Relationship of the number of waiting customers to waiting time of multiques with nonexhaustive and cyclic service. *Trans. Inst. Electron. Commun. Eng. Jpn J64-B*, 8 (Aug.), 891–892 (in Japanese).

JAISWAL, N. K. 1968. *Priority Queues*. Academic Press, New York.

KAMAL, A. E., AND HAMACHER, V. C. 1986. Approximate analysis of non-exhaustive multiserver polling systems with applications to local area networks. Tech. Rep. Dept. of Computing Science, Univ. of Alberta, Edmonton, Alberta, Canada.

KANADA, Y., AND IKEDA, T. 1987. On a multi-ring local area network. *Trans. Inst. Electron. Inf. Commun. Eng. J70-B*, 8 (Aug.), 903–911 (in Japanese).

KARVELAS, D., AND LEON-GARCIA, A. 1986. Performance of integrated packet voice/data token-passing rings. *IEEE J. Selected Areas Commun. SAC-4*, 6 (Sept.), 823–832.

KAYE, A. R. 1972. Analysis of a distributed control loop for data transmission. In *Proceedings of the Symposium on Computer-Communications Networks and Teletraffic*. Polytechnic Institute of Brooklyn, Brooklyn, N.Y., pp. 47–58.

KAYE, A. R., AND RICHARDSON, T. G. 1973. A performance criterion and traffic analysis for polling systems. *INFOR, Can. J. Oper. Res. Inf. Process. 11*, 2 (June), 93–112.

KIMURA, G., AND TAKAHASHI, Y. 1986. Diffusion approximation for a token ring system with nonexhaustive service. *IEEE J. Selected Areas Commun. SAC-4*, 6 (Sept.), 794–801.

KIMURA, G., AND TAKAHASHI, Y. 1987a. Traffic analysis for token ring systems with limited service. *Trans. Inst. Electron. Inf. Commun. Eng. J70-B*, 3 (Mar.), 327–336 (in Japanese).

KIMURA, G., AND TAKAHASHI, Y. 1987b. An approximation for a token ring system with priority classes of messages. *J. Inf. Process. 10*, 2, 86–91.

KLEINROCK, L. 1976. *Queueing Systems*. Vol. 2, *Computer Applications*. Wiley, New York.

KLEINROCK, L., AND SCHOLL, M. O. 1980. Packet switching in radio channels: New conflict-free access schemes. *IEEE Trans. Commun. COM-28*, 7 (July), 1015–1029.

KOBAYASHI, H., AND KONHEIM, A. G. 1977. Queueing models for computer communications system analysis. *IEEE Trans. Commun. COM-25*, 1 (Jan.), 2–29.

KONHEIM, A. G. 1976. Chaining in a loop system. *IEEE Trans. Commun. COM-24*, 2 (Feb.), 203–210.

KONHEIM, A. G. 1980. Mathematical models for computer data communication. In *Case Studies in Mathematical Modeling*, W. E. Boyce, Ed. Pitman Advanced Publishing Program, Boston, Mass., pp. 256–334.

KONHEIM, A. G., AND MEISTER, B. 1974. Waiting lines and times in a system with polling. *J. ACM 21*, 3 (July), 470–490.

KRUSKAL, J. B. 1969. Work-scheduling algorithms: A non-probabilistic queueing study (with possible application to No. 1 ESS). *Bell Syst. Tech. J. 48*, (Nov.), 2963–2974.

KSCHISCHANG, F. R., AND MOLLE, M. L. 1987. The helical window token ring. Tech. Rep. Dept. of Electrical Engineering, Univ. of Toronto, Toronto, Canada.

KUEHN, P. J. 1979. Multiqueue systems with nonexhaustive cyclic service. *Bell Syst. Tech. J. 58*, 3 (Mar.), 671–698.

KUEHN, P. J. 1981. Performance of ARQ-protocols for HDX transmission in hierarchical polling systems. *Perform. Eval. 1*, 1 (Jan.), 19–30.

KUROSAWA, K., AND TSUJII, S. 1981. Analysis on asymmetric polling systems with limiting service. In *Proceedings of the National Telecommunications Conference 1981* (New Orleans, La., Nov. 29–Dec. 3). IEEE, New York, pp. G.4.2.1–G.4.2.5.

LAM, S. S. 1983. Multiple access protocols. *Computer Communications.* Vol. I, *Principles*, W. Chou, Ed. Prentice Hall, Englewood Cliffs, N.J., pp. 114–155.

LEE, T. T. 1983. M/G/1/N queue with vacation time and limited service discipline. Tech. Memo. AT&T Bell Laboratories, Holmdel, N.J.

LEE, T. T. 1984. M/G/1/N Queue with vacation time and exhaustive service discipline. *Oper. Res. 32*, 4 (Jul.–Aug.), 774–784.

LEIBOWITZ, M. A. 1961. An approximate method for treating a class of multiqueue problems. *IBM J. Res. Dev. 5*, 3 (July), 204–209.

LEIBOWITZ, M. A. 1962. A note on some fundamental parameters of multiqueue systems. *IBM J. Res. Dev. 6*, 4 (Oct.), 470–471.

LEIBOWITZ, M. A. 1968. Queues. *Sci. Am. 219*, 2 (Aug.), 96–103.

LEVY, H. 1984. Non-uniform structures and synchronization patterns in shared channel communication networks. Ph.D. dissertation, UCLA-CSD-840049, Computer Science Dept., School of Engineering and Applied Science, Univ. of California, Los Angeles, Calif.

LEVY, H. 1986. Delay computation and dynamic behavior of nonsymmetric polling systems. Tech. Paper. AT&T Bell Laboratories, Holmdel, N.J.

MACK, C. 1957. The efficiency of *N* machines uni-directionally patrolled by one operative when walking time is constant and repair times are variable. *J. Roy. Stat. Soc.* Series B, *19*, 1, 173–178.

MACK, C., MURPHY, T., AND WEBB, N. L. 1957. The efficiency of *N* machines uni-directionally patrolled by one operative when walking time and repair times are constants. *J. Roy. Stat. Soc.* Series B, *19*, 1, 166–172.

MANFIELD, D. R. 1983. Analysis of a polling system with priorities. In *Proceedings of the IEEE Global Telecommunications Conference* (San Diego, Calif., Nov. 28–Dec. 1). IEEE, New York, pp. 1504–1508.

MANFIELD, D. R. 1985. Analysis of a priority polling system for two-way traffic. *IEEE Trans. Commun. COM-33*, 9 (Sept.), 1001–1006.

MAPP, G. E., AND MANFIELD, D. R. 1986. Performance analysis of priority polling systems with complex cycles. In *Proceedings of the International Council for Computer Communication 1986*, P. Kuehn, Ed. (Munich, Sept.). Elsevier North-Holland, Amsterdam, pp. 583–588.

MARTIN, J. 1967. *Design of Real-Time Computer Systems.* Prentice-Hall, Englewood Cliffs, N.J.

MASE, K. 1977. An approximate method for treating a queueing model of a class of time division multiplexing. *Trans. Inst. Electron. Commun. Eng. Jpn. J60-A*, 3 (Mar.), 332–333 (in Japanese).

MILLER, L. W. 1964. Alternating priorities in multiclass queues. Ph.D. dissertation, Dept. of Industrial Engineering, Cornell Univ., Ithaca, N.Y.

MORRIS, R. J. T., AND WANG, Y. T. 1984. Some results for multi-queue systems with multiple cyclic servers. In *Performance of Computer-Communication Systems*, H. Rudin and W. Bux, Eds. Elsevier North-Holland, Amsterdam, pp. 245–258.

MURATA, M., AND TAKAGI, H. 1987a. Two-layer modeling for local area networks. In *Proceedings of IEEE INFOCOM '87* (San Francisco, Mar. 31–Apr. 2). IEEE, New York, pp. 132–140.

MURATA, M., AND TAKAGI, H. 1987b. Performance of token ring networks with a finite capacity bridge. TRL Res. Rep., TR87-0027, IBM Japan, Tokyo.

NAHMIAS, S., AND ROTHKOPF, M. H. 1984. Stochastic models of internal mail delivery systems. *Manage. Sci. 30*, 9 (Sept.), 1113–1120.

NAIR, S. S. 1976. Alternating priority queues with non-zero switch rules. *Comput. Oper. Res. 3*, 337–346.

NEUTS, M. F., AND YADIN, M. 1968. The transient behavior of the queue with alternating priorities, with special reference to the waitingtimes. *Bull. Soc. Math. Belg. 20*, 343–376.

NISHIDA, T., MURATA, M., MIYAHARA, H., AND TAKASHIMA, K. 1983. Prioritized token passing method in ring-shaped local area networks. In *Proceedings of the IEEE International Conference on Communications (ICC '83)* (Boston, June 19–22). IEEE, New York, pp. D.2.2.1–D.2.2.6.

NISHIDA, T., MURATA, M., MIYAHARA, H., AND TAKASHIMA, K. 1986. An approximate analysis of a prioritized token passing method in ring-shaped local area networks. *Trans. Inst. Electron. Commun. Eng. Jpn E-69*, 1 (Jan.), 29–39.

NOGUCHI, K., SHIRATORI, N., AND NOGUCHI, S. 1984. Analysis of a multiqueue with finite buffers. *Queueing Theory and Its Applications*, Lecture Note 519. Research Institute for Mathematical Science, Kyoto Univ., Kyoto, Japan, pp. 114–136 (in Japanese).

NOMURA, M., AND TSUKAMOTO, K. 1978. Traffic analysis on polling systems. *Trans. Inst. Electron. Commun. Eng. Jpn. J61-B*, 7 (July), 600–607 (in Japanese).

OZAWA, T. 1987a. An analysis for multi-queueing systems with cyclic-service discipline: Models with exhaustive and gated services. IEICE Tech. Rep. The Institute of Electronics, Information and Communication Engineers of Japan, Tokyo, pp. 19–24 (in Japanese).

OZAWA, T. 1987b. Analysis of a single server model with two queues having different service disciplines. Tech. Rep. NTT Electrical Communication Laboratories, Musashino-shi, Tokyo.

PACK, C. D., AND WHITAKER, B. A. 1976. Multipoint private line (MPL) access delay under several interstation disciplines. *IEEE Trans. Commun. COM-24*, 3 (Mar.), 339–348.

PANG, J. W. M. 1985. Mean delay analysis for unidirectional broadcast structures for local area networks. Master's thesis, Dept. of Electrical Engineering, Faculty of Applied Science, Univ. of British Columbia, Vancouver, Canada.

PANG, J. W. M., AND DONALDSON, R. W. 1986. Approximate delay analysis and results for asymmetric token passing and polling networks. *IEEE J. Select. Areas Commun. SAC-4*, 6 (Sept.), 783–793.

PENNY, B. K., AND BAGHDADI, A. A. 1979a. Survey of computer communications loop networks: Part 1. *Comput. Commun. 2*, 4, 165–180.

PENNY, B. K., AND BAGHDADI, A. A. 1979b. Survey of computer communications loop networks: Part 2. *Comput. Commun. 2*, 5, 224–241.

PITTEL, B. G. 1973. Optimal control in a mass service system with several flows of demands. *Eng. Cybern. 11*, 6, 1029–1044.

RAITH, T. 1985. Performance analysis of multibus interconnection networks in distributed systems. In *11th International Teletraffic Congress*, M. Akiyama, Ed. Elsevier North-Holland, Amsterdam, pp. 4.2A-5-1–4.2A-5-7.

RAITH, T., AND TRAN-GIA, P. 1986. Performance analysis of polling mechanisms with receiver blocking. In *Proceedings of the International Council for Computer Communication 1986*, P. Kuehn, Ed. (Munich). Elsevier North-Holland, Amsterdam, pp. 577–582.

REGO, V. J. 1985. Cycle-time distributions in an adaptive token-passing bus network. In *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '85)* (New Orleans, La., Dec. 2–5). IEEE, New York, p. 48.2.

REGO, V. J., AND HUGHES, H. D. 1986. A comparison of two token-passing bus protocols. In *Proceedings of the ACM SIGCOMM '86 Symposium, Communications Architectures & Protocols* (Stowe, Vt., Aug. 5–7). ACM, New York, pp. 58–66.

REGO, V. J., AND NI, L. M. 1986. Cycle-time distributions in token-passing local networks. Tech. Paper. Dept. of Computer Sciences, Purdue Univ., West Lafayette, Ind.

REISER, M. 1982. Performance evaluation of data communications systems. *Proc. IEEE 70*, 2 (Feb.), 171–196.

ROBILLARD, P. N. 1974. An analysis of a loop switching system with multirank buffers based on the Markov process. *IEEE Trans. Commun. COM-22*, 11 (Nov.), 1772–1778.

RUBIN, I., AND DE MORAES, L. F. M. 1983. Message delay analysis for polling and token multiple-access schemes for local communication networks. *IEEE J. Select. Areas Commun. SAC-1*, 5 (Nov.), 935–947.

RUBIN, I., AND ZSAI, Z.-H. 1987a. Performance of random-access reservation schemes using a polling backbone for metropolitan and packet-radio networks. In *Proceedings of the IEEE International Conference on Communications '87* (Seattle, Wash., June 7–10). IEEE, New York, pp. 1427–1431.

RUBIN, I., AND ZSAI, Z.-H. 1987b. Performance of double-tier access-control schemes using a polling backbone for metropolitan and interconnected communication networks. *IEEE J. Select. Areas Commun. SAC-5*, 9 (Sept.), 1403–1417.

RUNNENBERG, J. TH. 1957. Machines served by a patrolling operator. Statistische Afdeling Rep. S221 (VP13), Mathematisch Centrum, Amsterdam.

SAATY, T. L. 1961. *Elements of Queueing Theory with Applications.* McGraw-Hill, New York. Reprinted by Dover Publications, New York, 1983.

SARKAR, D., AND ZANGWILL, W. I. 1987. Expected waiting time for nonsymmetric cyclic queueing systems—Exact results and applications. Tech. Paper. AT&T Bell Laboratories, Holmdel, N.J.

SAYDAM, T., AND SETHI, A. S. 1985. Performance evaluation of voice-data token ring LANs with random priorities. In *Proceedings of IEEE INFOCOM '85* (Washington, D.C., Mar.). IEEE, New York, pp. 326–332.

SCHAY, G. JR., 1962. Approximate methods for a multiqueueing problem. *IBM J. Res. Dev. 6*, 2 (Apr.), 246–249.

SCHOLL, M. O. 1976. Multiplexing techniques for data transmission over packet switched radio systems. Tech. Rep. UCLA-ENG-76123. Computer Science Dept., Univ. of California, Los Angeles, Calif.

SCHOLL, M., AND KLEINROCK, L. 1983. On the M/G/1 queue with rest periods and certain service-independent queueing disciplines. *Oper. Res. 31*, 4 (Jul.–Aug.), 705–719.

SCHOLL, M., AND POTIER, D. 1978. Finite and infinite source models for communication systems under polling. IRIA Rapport de Recherche 308. Institut de Rescherche en Informatique et en Automatique, Le Chesnay, France.

SCHWARTZ, M. 1977. *Computer-Communication Network Design and Analysis.* Prentice-Hall, Englewood Cliffs, N.J.

SCHWARTZ, M. 1987. *Telecommunication Networks: Protocols, Modeling and Analysis.* Addison-Wesley, Reading, Mass.

SERVI, L. D. 1985. Capacity estimation of cyclic queues. *IEEE Trans. Commun. COM-33*, 3 (Mar.), 279–282.

SERVI, L. D. 1986. Average delay approximation of M/G/1 cyclic service queues with Bernoulli schedules. *IEEE J. Select. Areas Commun. SAC-4*, 6 (Sept.), 813–822; correction in *SAC-5*, 3 (Apr.), p. 547, 1987.

SETHI, A. S., AND SAYDAM, T. 1985. Performance analysis of token ring local area networks. *Comput. Networks ISDN Syst. 9*, 3 (Mar.), 191–200.

SEVCIK, K. C., AND JOHNSON, M. J. 1987. Cycle time properties of the FDDI token ring protocol. *IEEE Trans. Softw. Eng. SE-13*, 3 (Mar.), 376–385.

SHEN, Z., MASUYAMA, S., MURO, S., AND HASEGAWA, T. 1985. Performance evaluation of prioritized token ring protocols. In *11th International Teletraffic Congress*, M. Akiyama, Ed. Elsevier North-Holland, Amsterdam, pp. 4.2A-3-1–4.2A-3-7.

SKINNER, C. E. 1967. A priority queueing system with server-walking time. *Oper. Res. 15*, 2 (Mar.–Apr.), 278–285.

SRINIVASAN, M. M. 1986. An approximation for mean waiting times in cyclic server systems with nonexhaustive service. Tech. Rep. 86-36. Dept. of Industrial and Operations Engineering, Univ. of Michigan, Ann Arbor, Mich.

SRINIVASAN, M. M., AND LEE, H.-S. 1987. An analysis of a prioritized polling mechanism for cyclic server systems. Tech. Rep. 87-6. Dept. of Industrial and Operations Engineering, Univ. of Michigan, Ann Arbor, Mich.

SRIVASTAVA, H. M., AND KASHYAP, B. R. K. 1982. *Special Functions in Queuing Theory and Related Stochastic Processes.* Academic Press, Orlando, Fla.

STIDHAM, S., JR. 1969. Optimal control of a signalized intersection. Part I: Introduction: Structure of intersection models; Part II: Determining the optimal switching policies; Part III: Descriptive stochastic models. Tech. Reps. 94, 95, and 96, Dept. of Operations Research, Cornell Univ., Ithaca, New York.

STIDHAM, S. 1972. Regenerative processes in the theory of queues, with applications to the alternating-priority queue. *Adv. Appl. Probab. 4*,'3 (Dec.), 542–577.

SUDA, T., MIYAHARA, H., AND HASEGAWA, T. 1980. Approximate analysis for a queueing system with probabilistic switching. *Trans. Inst. Electron. Commun. Eng. Jpn J63-D*, 1 (Jan.), 1–8 (in Japanese).

SWARTZ, G. B. 1980. Polling in a loop system. *J. ACM 27*, 1 (Jan.), 42–59.

SWARTZ, G. B. 1982. Analysis of a SCAN policy in a gated loop system. In *Applied Probability–Computer Science: The Interface*, vol. 2, R. L. Disney and T. J. Ott, Eds. Birkhäuser, Boston, pp. 241–252.

SYKES, J. S. 1969a. Analytical model of half-duplex interconnections of computers. *IEEE Trans. Commun. Technol. COM-17*, 2 (Apr.), 235–238.

SYKES, J. S. 1969b. Analysis of the communications aspects of an inquiry-response system. In *AFIPS Conference Proceedings, 1969 Fall Joint Computer Conference*, vol. 35 (Las Vegas, Nev., Nov. 18–20). AFIPS Press, Reston, Va., pp. 655–667.

SYKES, J. S. 1970. Simplified analysis of an alternating-priority queueing model with setup times. *Oper. Res. 18*, 6 (Nov.–Dec.), 1182–1192.

SZPANKOWSKI, W. AND REGO, V. 1987. Ultimate stability conditions for some multidimensional distributed systems. Tech. Rep. Dept. of Computer Science, Purdue Univ., West Lafayette, Ind.

TAKÁCS, L. 1962. *Introduction to the Theory of Queues.* Oxford University Press, New York.

TAKÁCS, L. 1968. Two queues attended by a single server. *Oper. Res. 16*, 3 (May–June), 639–650.

TAKAGI, H. 1984. Mean message waiting time in a symmetric polling systems. In *Performance '84*, E. Gelenbe, Ed. Elsevier North-Holland, Amsterdam, pp. 293–302.

TAKAGI, H. 1985a. On the analysis of a symmetric polling system with single-message buffers. *Perform. Eval. 5*, 3 (Aug.), 149–157.

TAKAGI, H. 1985b. Mean message waiting times in symmetric multi-queue systems with cyclic service. *Perform. Eval. 5*, 4 (Nov.), 271–277. Also in *Proceedings of the IEEE International Conference on Communications 1985* (Chicago, June 23–26). IEEE, New York, pp. 1154–1157.

TAKAGI, H. 1986. *Analysis of Polling Systems.* The MIT Press, Cambridge, Mass. Also, Takagi, H., and Kleinrock, L. Analysis of polling systems. JSI Res. Rep. TR87-0002. IBM Japan Science Institute, Tokyo, Jan. 1985.

TAKAGI, H. 1987a. Analysis and applications of a multi-queue cyclic service system with feedback. *IEEE Trans. Commun. COM-35*, 2 (Feb.), 248–250.

TAKAGI, H. 1987b. Exact analysis of round-robin scheduling of services. *IBM J. Res. Dev. 31*, 4 (July), 484–488.

TAKAGI, H. 1987c. Analysis of polling models with a mix of exhaustive and gated service disciplines. Tech. Rep. Tokyo Research Laboratory, IBM Japan, Tokyo.

TAKAGI, H. 1987d. Queueing analysis of polling systems with zero switchover times. TRL Res. Rep. TR87-0034. IBM Japan, Tokyo.

TAKAGI, H., AND MURATA, M. 1986. Queueing analysis of scan-type TDM and polling systems. In *Computer Networking and Performance Evaluation*, T. Hasegawa, H. Takagi, and Y. Takahashi, Eds. Elsevier North-Holland, Amsterdam, pp. 199–211.

TAKINE, T., TAKAHASHI, Y., AND HASEGAWA, T. 1986. Performance analysis of a polling system with single buffers and its application to interconnected networks. *IEEE J. Selected Areas Commun. SAC-4*, 6 (Sept.), 802–812.

TAKINE, T., TAKAHASHI, Y., AND HASEGAWA, T. 1987a. Analysis of a buffer relaxation polling system with single buffers. In *Proceedings of the Seminar on Queueing Theory and Its Applications* (Kyoto, Japan, May 11–13). Kyoto Univ., Kyoto, Japan. pp. 117–132.

TAKINE, T., TAKAHASHI, Y., AND HASEGAWA, T. 1987b. An analysis for interdeparture process of a polling system with single message buffer at each station. *Trans. Inst. Electron. Inf. Commun. Eng. 70-B*, 9 (Sept.), 989–998.

TAKINE, T., TAKAHASHI, Y., AND HASEGAWA, T. 1987c. Average message delay of an asymmetric single buffer polling system with round-robin scheduling of services. Tech. Rep. 87020. Dept. of Applied Mathematics and Physics, Kyoto Univ., Kyoto, Japan.

TAKINE, T., TAKAHASHI, Y., AND HASEGAWA, T. 1988. Analysis of an asymmetric polling system with single buffers. In *Performance '87*, P.-J. Courtois, and G. Latouche, Eds. Elsevier North-Holland, Amsterdam, pp. 241–251.

TOBAGI, F. A., AND FINE, M. 1983. Performance of unidirectional broadcast local area networks: Expressnet and Fasnet. *IEEE J. Select. Areas Commun. SAC-1*, 5 (Sept.), 913–926.

TOBAGI, F. A., AND KLEINROCK, L. 1976. Packet switching in radio channels: Part III—Polling and dynamic split-channel reservation multiple channel access. *IEEE Trans. Commun. COM-24*, 8 (Aug.), 832–845.

TOBAGI, F. A., BORGONOVO, F., AND FRATTA, L. 1983. Expressnet: A high-performance integrated-services local area network. *IEEE J. Selec. Areas Commun. SAC-1*, 5 (Sept.), 898–913. Also in *Advances in Local Area Networks*, K. Kümmerle, F. A. Tobagi, and J. O. Limb, Eds. IEEE, New York, 1987, pp. 171–189.

TRAN-GIA, P. 1988. Discrete-time analysis of polling systems with renewal inputs. In *Data Communication Systems and Their Performance*, L. F. M. de Moraes, E. de Souza e Silva, and L. F. G. Soares, Eds. Elsevier North-Holland, Amsterdam, pp. 495–510.

TRAN-GIA, P., AND RAITH, T. 1986. Multiqueue systems with finite capacity and nonexhaustive cyclic service. In *Computer Networking and Performance Evaluation*, T. Hasegawa, H. Takagi, and Y. Takahashi, Eds. Elsevier North-Holland, Amsterdam, pp. 213–225.

TROPPER, C. 1981. *Local Computer Network Technologies.* Academic Press, Orlando, Fla.

VINNITSKII, V. P., AND BRODETSKII, G. L. 1972. A cyclic queueing system. *Cybernetics 8*, 6 (Nov.–Dec.), 979–987.

WANG, D.-G. 1986. Analysis of cyclic service multi-queue systems for ring type local area networks. Master's thesis, Dept. of Electrical and Computer Engineering, North Carolina State Univ., Raleigh, N.C.

WANG, Y. T., AND MORRIS, R. J. T. 1985. Load sharing in distributed systems. *IEEE Trans. Comput. C-34*, 3 (Mar.), 204–217.

WATSON, K. S. 1984. Performance evaluation of cyclic service strategies—A survey. In *Performance '84*, E. Gelenbe, Ed. Elsevier North-Holland, Amsterdam, pp. 521–533.

WOLFF, R. W. 1988. *Stochastic Modeling and the Theory of Queues.* Prentice-Hall, Englewood Cliffs, N.J.

WU, R. M., AND CHEN, Y.-B. 1975. Analysis of a loop transmission system with round-robin scheduling of services. *IBM J. Res. Dev. 19*, 5 (Sept.), 486–493.

YADIN, M. 1970. Queueing with alternating priorities, treated as random walk on the lattice in the plane. *J. Appl. Probab. 7*, 196–218.

YAMAMOTO, T., OKADA, H., AND NAKANISHI, Y. 1984. Analysis of token-passing ring networks with transmission priority. *Trans. Inst. Electron. Commun. Eng. Jpn., J67-D*, 9 (Sept.), 989–996 (in Japanese).

YUE, O.-C. 1987. Performance of the timed token scheme in MAP. In *Proceedings of the IEEE/IEICE Global Telecommunications Conference 1987.* (Tokyo, Nov. 15–18). IEEE, New York, pp. 187–192.

YUEN, M. L. T., BLACK, B. A., NEWHALL, E. E., AND VENETSANOPOULOS, A. N. 1972. Traffic flow in a distributed loop switching system. In *Proceedings of the Symposium on Computer-Communications Networks and Teletraffic* (Apr. 4–6). Polytechnic Institute of Brooklyn, Brooklyn, N.Y., pp. 29–46.

ZHDANOV, V. S., AND SAKSONOV, E. A. 1979. Conditions of existence of steady-state modes in cyclic queueing systems. *Autom. Remote Control 40*, 2, Part 1 (Feb.), 176–184.

Received July 1987; final revision accepted January 1988.