THE ANALYSIS OF RANDOM POLLING SYSTEMS

LEONARD KLEINROCK
University of California, Los Angeles, California

HANOCH LEVY
AT&T Bell Laboratories, Holmdel, New Jersey

(Received August 1985; revisions received June 1986, February 1987; accepted September 1987)

In this paper, we analyze the behavior of *random polling systems.* The polling systems we consider consist of N stations, each equipped with an infinite buffer and a single server who serves them in some order. In contrast to previously studied polling systems, where the order of service used by the server is *periodic* (and usually *cyclic*), in the systems we consider the next station to be served after station i is determined by *probabilistic means.* More specifically, according to the model we consider in this paper, after serving station i, the server will poll (i.e., serve) station j ($j = 1, 2, \ldots, N$) with probability p_j. The main results of this paper are expressions for the expected response time in a random polling system operated under a variety of service disciplines. The results are compared to the response time in the equivalent cyclic polling systems. Also in this paper, we analyze the cycle time and the number of customers found in the system.

The queueing behavior of *polling systems* has been extensively investigated in the past. The "traditional" polling scheme that appears in the literature is a method by which a single server serves N stations: each generates its own stream of work requests (or customers) and each is equipped with an infinite queue to store its requests. According to this scheme, the N stations are served in a *cyclic order* in which the station served after station i is station $i + 1$ (modulo N); this is called the *cyclic polling scheme.*

In contrast to previous studies that dealt with (periodic and) *cyclic polling* schemes, our aim in this paper is to study the *random polling* scheme, where the polling order is not fixed. Rather, the next station polled is determined according to some random (memoryless) criterion. According to the specific scheme we investigate, the next station polled will be station j ($j = 1, 2, \ldots, N$) with probability p_j.

The traditional cyclic polling schemes have been successfully used to model systems where a *central* controller polls and serves many stations. A typical example is a time shared system where a single computer serves many terminals. In contrast, our work has been motivated by the wish to model *distributed* systems. In many of these distributed systems the control moves from one station to another according to some random criterion. As an example, consider a shared broadcast channel where the decision regarding "who will transmit next" is made in a distributed manner, and is based on some randomly behaving algorithms, rather than on a fixed order. The random schemes analyzed in this paper are believed to be a natural model for such distributed systems. As an example, the results reported in this paper were used in Levy (1984) to predict the expected delay in a Slotted ALOHA system.

The main objective of this paper is to analyze the *response time* (*waiting time* plus *service time*) observed in the random polling systems. Specifically, the random polling scheme is studied for three types of service policy: 1) exhaustive service, 2) gated service, and 3) limited service. In the *exhaustive* policy, when queue i is selected for service, the server will continue to serve this queue until the queue becomes empty. Thus, all customers found in the queue at the beginning of the service period, and those who arrive during the service period, are served in that period. In the *gated* policy when queue i is selected for service, the server will serve in that service period, all (and only) those customers found in queue i at the *beginning* of the service period. Thus, none of the customers arriving during the service period will be served during this period. In the limited service policy, the server will serve in a given service period exactly one customer (given that at least one customer is present at the polled station at the polling instant). The model is a *discrete time* model and the extension of the results to a continuous time model can be done in a similar way. As in the analysis of many cyclic polling systems, we allow the server to have a random length *switchover*

Subject classification: Queues: random polling.

0030-364X/88/3605-0716 $01.25

period between the service of one station and the next station. The length of a switchover period, in our model, is associated with the station served *prior* to the switchover period.

The main results of this paper are delay expressions for the different service policies. Under the assumption of a fully symmetric system, we are able to derive a closed form expression of the expected response time for all three types of service policies. Under the assumption of a nonsymmetric system, we derive the expected response time for both the exhaustive and the gated systems. In this case, we form a set of N^2 linear equations, the solution of which yields the expected response time in the system. Other important measures such as the number of customers found in the system, the cycle time and the buffer utilization are also derived in this paper. The approach used to analyze the exhaustive and gated systems is similar to approaches previously used to analyze the equivalent cyclic systems. The approach we use to derive the expected response time in the limited service system is partially new. The analysis is based on the assumption that the switchover periods are not (all) zero length. Nevertheless, the results obtained can be applied to systems with no switchover periods by considering the limits of these results when the lengths of the switchover periods approach zero.

The structure of this paper is as follows. After a detailed description of the system model (Section 2), the exhaustive scheme, the gated scheme and the limited service scheme are analyzed in Sections 3, 4 and 5, respectively. In Section 6 we discuss the application of our results to systems with zero length switchover periods. Finally, in Section 7, the expression for the expected response time of the three different policies are compared to each other and to the corresponding expressions in the cyclic polling systems. A glossary of notation is given in the Appendix.

1. Previous Work

Since the amount of work done in the area of polling systems is tremendous, we will mention only those references which are closely related to this paper. The *discrete time* models of cyclic polling with N stations, independent arrivals and nonzero switchover periods (the models to which our model is similar in assumptions) were studied first in the mid-1970s. Konheim and Meister (1974) analyzed the exhaustive service policy in the symmetric system; Swartz (1980), De Moraes (1981) and Rubin and De Moraes (1983) studied the nonsymmetric exhaustive system; and De Moraes (1981) and Rubin and De Moraes (1983) studied the nonsysmmetric gated system. Takagi (1985) studied the symmetric limited service system where, at most, one customer is served at a time.

Many ideas used in the analysis of discrete time polling systems are similar to those used in the analysis of the *continuous time* polling systems with Poisson arrivals. Cooper and Murray (1969) and Cooper (1970) studied the exhaustive and the gated schemes in systems with zero length switchover period. Systems with non-zero switchover periods were analyzed by Eisenberg (1972) (the exhaustive scheme) and Hashida (1972) (both the gated and the exhaustive schemes). Common to these studies (and to the discrete time studies) is the approach of analyzing customers' delays by computing the number of customers present in the system at polling instants. In more recent studies, Humblet (1978) and Ferguson and Aminetzah (1985) suggested a different approach to study the continuous time gated and exhaustive systems. Their approach is based on computing the length of the service period and results in an efficient method for calculating the delay in nonsymmetric systems. Nomura and Tsukamoto (1978) studied the symmetric limited service system where, at most, one customer is served at a time (the analysis is provided for systems with non-zero switchover periods).

Lastly, a tutorial of polling systems was recently written by Takagi and Kleinrock (1985a,b), which has since been published as a book by Takagi (1986). This tutorial summarizes the known results for polling systems and presents an organized derivation of most of the known results; it served as an excellent source for previous results, and guided us in the derivation of many of our results. Many of the references to polling systems not mentioned here (such as those which use different models or contain approximations) can be found in that tutorial. More recent results appear in Takagi (1987).

2. Model Description and General Notation

We consider a system with N infinite-buffer queues and one roving server. Time is slotted with the slot size equal to the (constant) service time of a customer, and all time units are normalized to this slot size. The time interval $(t-1, t)$ is called the tth slot. Customers who arrive during the tth slot are assumed to arrive at the end of the slot (i.e., at time $t-0$) and may first be served during the $t+1$st slot.

The arrival process to each queue consists of batches of customers. We denote by $X_i(t)$ the number of

customers arriving at station i during the tth slot, i.e., this is the size of the batch arriving at station i during the tth slot. For each queue i, the arrival sequence, $\{X_i(t): t = 1, 2, \ldots\}$ is assumed to be an independent and identically distributed sequence of random variables. The generating function, mean and variance of $X_i(t)$ are given by

$$P_i(z) \triangleq \mathrm{E}[z^{X_i(t)}];$$

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

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

where

$$P_i^{(1)}(1) \triangleq \frac{dP(z)}{dz}\bigg|_{z=1}, \quad P_i^{(2)}(1) \triangleq \frac{d^2P(z)}{dz^2}\bigg|_{z=1}.$$

The polling policy is the following: after completing the service of queue i (the period during which the server continuously serves a queue is called a *service period*), the server incurs a *switchover period.* (If a selected queue contains no customers at its polling instant, the length of the service period is zero and a switchover period will still be incurred in moving to the next queue.) During this period, none of the queues is served, and it may be considered as the time required to switch from queue i to the next queue to be served. The length of the switchover period has a distribution that depends only on the queue previously served (in this case, i). At the end of the switchover period, the server picks, in a random fashion, the next queue to be served. The *polling policy* is memoryless such that queue j is selected to be served next with probability p_j. As described in the Introduction, three types of service *policies* are considered in this paper: *exhaustive, gated* and *limited service.*

Three types of epochs are of interest: the time at which the server starts serving queue i for the mth time, the time at which this service periods ends, and the time when the switchover period, succeeding this service period, terminates. The mth period at which queue i is served is called the *mth service period of queue i.* The switchover period succeeding the mth service period of queue i is called the *mth switchover period of queue i.* Let us use the following notation:

$\tau_i(m) \triangleq$ the instant at which the mth *service* period of queue i *starts.*

$\tau_i(m) \triangleq$ the instant at which the mth *service* period of queue i *terminates.*

$\bar{\tau}_i(m) \triangleq$ the instant at which the mth *switchover* period of queue i *terminates.*

Similarly, the instant at which the server starts the nth service period (independent of the station polled), the instant at which the server finishes the nth service period, and the instant at which the server finishes the nth switchover period are, respectively, denoted by $\tau(n)$, $\tau(n)$ and $\bar{\tau}(n)$. Note that $\bar{\tau}(n) = \tau(n+1)$.

The length of the mth switchover period of queue i is $\bar{\tau}_i(m) - \tau_i(m)$. For each queue, we assume that the sequence of switchover periods associated with it, $\{\bar{\tau}_i(m) - \tau_i(m): m = 1, 2, \ldots\}$, is a sequence of independent and identically distributed random variables. The generating function, mean and variance of $\bar{\tau}_i(m) - \tau_i(m)$ are given by:

$$\begin{aligned} R_i(z) &\triangleq \mathrm{E}[z^{\bar{\tau}_i(m)-\tau_i(m)}] \\ r_i &\triangleq \mathrm{E}[\bar{\tau}_i(m) - \tau_i(m)] = R_i^{(1)} \\ \delta_i^2 &\triangleq \mathrm{Var}[\bar{\tau}_i(m) - \tau_i(m)] \\ &= R_i^{(2)}(1) + R_i^{(1)}(1) - [R_i^{(1)}]^2. \end{aligned} \tag{1}$$

It is assumed that not all the switchover periods are of zero length. This means that there exists i such that $R_i(z) \neq 1$ (and thus $r_i > 0$).

The number of customers in the system is denoted as:

$$L_i(t) \triangleq \text{number of customers at queue } i \text{ at time } t;$$

$$\mathbf{L}(t) \triangleq [L_1(t), L_2(t), \ldots, L_N(t)].$$

Note that the process $\mathbf{L}$ embedded at the polling instants is Markovian (although the process $\mathbf{L}(t)$ by itself is not).

The generating function of the number of customers found in the system at the mth polling instant is:

$$F_m\|z_1, z_2, \ldots, z_N) \triangleq \mathrm{E}\left[\prod_{j=1}^{N} z^{L_j(\tau(m))}\right]. \tag{1}$$

Assuming equilibrium conditions, we may define the limiting generating function as

$$F(z_1, z_2, \ldots, z_N) \triangleq \lim_{m\to\infty} F_m(z_1, z_2, \ldots, z_N).$$

Similarly, the limiting marginal generating function for $L_i(\tau(m))$ when m approaches infinity is denoted by

$$F_i(z) \triangleq \lim_{m\to\infty} \mathrm{E}[z^{L_i(\tau(m))}] = F(1, \ldots, 1, z, 1, \ldots, 1).$$

In addition, let L_i be a random variable representing the number of customers at station i at an arbitrary instant when the system is in equilibrium. Similarly, let L_i^* be a random variable representing the number

of customers at station i at an arbitrary *polling instant* when the system is in equilibrium.

3. Analysis of the Exhaustive Service Policy

3.1. Number of Customers at Polling Instants: Derivation of the Generating Function

We start our study by analyzing the number of customers found in the exhaustive system at the polling instants. To calculate $F(z_1, z_2, \ldots, z_N)$, we express $F_{m+1}(z_1, z_2, \ldots, z_N)$ in terms of $F_m(z_1, z_2, \ldots, z_N)$. This is done by conditioning the calculation on the specific queue served during the mth service period. Let this queue be the ith queue.

The time interval of interest is the interval $[\underline{\tau}(m), \bar{\tau}(m)]$ which consists of the concatenation of the mth service period, $[\underline{\tau}(m), \tau(m)]$ and the mth switchover period, $[\tau(m), \bar{\tau}(m)]$. Since station i is the station served in the mth service period, there exists some (unique) n such that $\underline{\tau}_i(n) = \underline{\tau}(m)$, $\tau_i(n) = \tau(m)$ and $\bar{\tau}(n) = \bar{\tau}(m)$. Thus, the periods of interest are the nth service period of station i and the nth switchover period of queue i. First, consider the service period of station i. The length of this period, given by $\tau_i(n) - \underline{\tau}_i(n)$, corresponds to the gambler's ruin time (i.e., the time from an initial capital to zero capital) in the well known *gambler's ruin problem* (a short description of this problem and its solution may be found in Konheim 1980). The generating function of this time is expressed in terms of the number of customers present at station i at the polling instant

$$\mathrm{E}[w^{\tau_i(n)-\underline{\tau}_i(n)}] = \mathrm{E}[\{\Theta_i(w)\}^{L_i(\underline{\tau}_i(n))}] \tag{2}$$

where $\Theta_i(\cdot)$ is the generating function of the ruin time when the gambler's initial capital is one unit, and where the moments of this ruin time are given by

$$\Theta_i(1) = 1,$$

$$\Theta_i^{(1)}(1) = \frac{1}{1-\mu_i},$$

$$\Theta_i^{(2)}(1) = \frac{\mu_i}{(1-\mu_i)^2} + \frac{\sigma_i^2}{(1-\mu_i)^3}.$$

Now, to calculate the number of customers in the system we follow the analysis of the discrete time cyclic exhaustive system (Konheim and Meister; Swartz; Rubin and De Moraes; and Takagi). The approach (which was used by earlier authors, e.g., Cooper and Murray, for the continuous time system) is to express the generating function of the number of customers found in the system when station $i+1$ is polled as a function of the generating function of the number of customers found in the system when station i is polled. It is easy to adapt this analysis to our system, yielding the corresponding expression (see Levy 1984):

$$\begin{aligned} &F_{m+1}(z_1, z_2, \ldots, z_N \mid A_i) \\ &= R_i\left(\prod_{j=1}^{N} P_j(z_j)\right) \\ &\quad \cdot F_m\left(z_1, z_2, \ldots, z_{i-1}, \Theta_i\left(\prod_{\substack{j=1 \\ (j\neq i)}}^{N} P_j(z_j)\right), \right. \\ &\qquad \left. z_{i+1}, \ldots, z_N\right) \end{aligned} \tag{3}$$

where A_i is the event that queue i was polled at the previous (in this case, the mth) service period.

Now, unconditioning (3), letting m approach infinity and assuming that the system reaches equilibrium we obtain:

$$\begin{aligned} &F(z_1, z_2, \ldots, z_N) \\ &= p_1 \cdot R_1\left(\prod_{j=1}^{N} P_j(z_j)\right) \\ &\quad \cdot F\left(\Theta_1\left(\prod_{\substack{j=1 \\ (j\neq 1)}}^{N} P_j(z_j)\right), z_2, z_3, \ldots, z_N\right) \\ &\quad + p_2 \cdot R_2\left(\prod_{j=1}^{N} P_j(z_j)\right) \\ &\quad \cdot F\left(z_1, \Theta_2\left(\prod_{\substack{j=1 \\ (j\neq 2)}}^{N} P_j(z_j)\right), z_3, \ldots, z_N\right) \\ &\quad + \cdots + p_N \cdot R_N\left(\prod_{j=1}^{N} P_j(z_j)\right) \\ &\quad \cdot F\left(z_1, z_2, z_3, \ldots, \Theta_N\left(\prod_{\substack{j=1 \\ (j\neq N)}}^{N} P_j(z_j)\right)\right). \end{aligned} \tag{4}$$

3.2. Number of Customers at Polling Instants: Mean and Variance

Next, we compute from (4) the mean and variance of the number of customers found in the system at polling instants. Let the partial derivatives of

$F_m(z_1, z_2, \ldots, z_N)$ be denoted as follows:

$$f_m(i) \triangleq \left.\frac{\partial F_m(z_1, \ldots, z_N)}{\partial z_i}\right|_{\vec{z}=\vec{1}}$$

$$f_m(i,j) \triangleq \left.\frac{\partial^2 F_m(z_1, \ldots, z_N)}{\partial z_i \partial z_j}\right|_{\vec{z}=\vec{1}}; \quad i,j = 1, 2, \ldots, N$$

where $\vec{z} \triangleq (z_1, z_2, \ldots, z_N)$ and $\vec{1}$ corresponds to the vector $(1, 1, \ldots, 1)$. Similarly, we define $f_m(i \mid k)$ and $f_m(i, j \mid k)$ to be the corresponding derivatives, conditioned on station k being served during the previous service period. We also define $f(i), f(i, j), f(i \mid k)$ and $f(i, j \mid k)$ to be, respectively, the limits of these derivatives (when the limits exist) when m approaches infinity. Using this notation,

$$\mathrm{E}[L_i^*] = f(i),$$

$$\mathrm{Var}[L_i^*] = f(i, i) + f(i) - \{f(i)\}^2. \tag{5}$$

Differentiating (4) with respect to the z_i's, to calculate the terms $f(i), i = 1, 2, \ldots, N$, yields a set of N linear equations of the form

$$f(j) = \frac{\mu_j}{p_j} \cdot \left(\sum_{i=1}^{N} p_i r_i + \sum_{\substack{i=1 \\ i \neq j}}^{N} \frac{p_i f(i)}{1 - \mu_i} \right).$$

The solution of this equation set (see Levy 1984) is

$$\mathrm{E}[L_j^*] = f(j) = \frac{(1 - \mu_j)\mu_j \sum_{i=1}^{N} p_i r_i}{p_j (1 - \sum_{i=1}^{N} \mu_i)} \tag{6}$$

which is the expected length of queue j at polling instants.

For the special case $p_i = 1/N$ for each i, we find that (6) is equal to the equivalent expression in the exhaustive service cyclic polling system (Swartz). In the case of a fully symmetric system, i.e., where $\mu_i = \mu$, $\sigma_i^2 = \sigma^2$, $r_i = r$, $\delta_i^2 = \delta^2$ and $p_i = 1/N$ for each i, the expected queue length is given by

$$\mathrm{E}[L_j^*] = f(j) = \frac{Nr\mu(1 - \mu)}{1 - N\mu}.$$

Note that this result for the random polling system is exactly the same as the well known result (Konheim and Meister; Swartz) for the cyclic polling exhaustive system.

Next, to derive $\mathrm{Var}[L_i^*]$ we must calculate $f(i, i)$. Differentiation of (4) twice with respect to the z_i's (see Levy 1984) yields a set of N^2 linear equations that may be solved by numerical methods. Note that the equivalent equation set for the cyclic polling systems (e.g., Rubin and De Moraes) consists of N^3 linear equations, and thus, for some numerical techniques it will be easier to solve the random polling system. However, it seems that the efficient techniques for solving these equation sets are iterative ones (see e.g., Levy 1986 for an analysis of the successive substitution method when applied to these sets). This specific reduction from N^3 to N^2 does not reduce the computational complexity for these techniques. The reason is that to solve the larger set of equations, a vector of N^3 components is computed in each iteration, with each component requiring $O(1)$ operations, while in the smaller set, a vector of N^2 components is computed in each iteration, however each component requires $O(N)$ operations per iteration (see the summations in 7a and 7b); therefore, the overall computation per iteration required in both cases is $O(N^3)$.

When the switchover period and the arrival process are assumed to be identical for all stations, this set becomes

$$\begin{aligned} f(j,k) &= \sum_{\substack{i=1 \\ (i \neq j) \\ (i \neq k)}}^{N} (a + b[f(j) + f(k)] + cf(i) \\ &\quad + d[f(i,j) + f(i,k)] + f(j,k) + d^2 f(i,i)) \cdot p_i \\ &\quad + (a + b[f(j) + df(k)]) \cdot p_k \\ &\quad + (a + b[f(k) + df(j)]) \cdot p_j \qquad j \neq k \end{aligned} \tag{7a}$$

$$\begin{aligned} f(j,j) = \sum_{\substack{i=1 \\ (i=j)}}^{N} \Big\{ & a + r(\sigma^2 - \mu) + 2bf(j) \\ & + \left(\frac{\sigma^2 - \mu}{1 - \mu} + c \right) f(i) + f(j,j) \\ & + 2\,df(i,j) + d^2 f(i,i) \Big\} \cdot p_i \\ & + p_j [a + r(\sigma^2 - \mu)] \end{aligned} \tag{7b}$$

where

$$a \triangleq \mu^2(\delta^2 + r^2), \quad b \triangleq r\mu,$$

$$c \triangleq \mu^2 \left[\frac{2r}{1 - \mu} + \frac{1}{(1 - \mu)^2} + \frac{\sigma^2}{(1 - \mu)^3} \right],$$

$$d \triangleq \frac{\mu}{1 - \mu}.$$

In the case of fully symmetric stations, (7a) and (7b) can be solved analytically (see Levy 1984). This yields

the following solution for $f(i, i)$, $i = 1, \ldots, N$:

$$f(i,i) = \frac{\delta^2\mu^2 N(1-\mu)}{1-N\mu} + \frac{\sigma^2 rN[1-(N+1)\mu+(2N-1)\mu^2]}{(1-N\mu)^2} - \frac{Nr\mu(1-\mu)}{1-N\mu} + \frac{N^2r^2\mu^2(1-\mu)^2}{(1-N\mu)^2} + \frac{Nr^2\mu^2(N-1)(1-\mu)}{(1-N\mu)^2}. \quad (8)$$

From (5) and (8) we can now calculate the second moment and the variance of the number of customers at polling instants:

$$\mathrm{E}[\{L_i^*\}^2] = \frac{\delta^2\mu^2 N(1-\mu)}{1-N\mu} + \frac{\sigma^2 rN[1-(N+1)\mu+(2N-1)\mu^2]}{(1-N\mu)^2} + \frac{N^2r^2\mu^2(1-\mu)^2}{(1-N\mu)^2} + \frac{Nr^2\mu^2(N-1)(1-\mu)}{(1-N\mu)^2} \quad (9)$$

$$\mathrm{Var}[L_i^*] = \frac{\delta^2\mu^2 N(1-\mu)}{1-N\mu} + \frac{\sigma^2 rN[1-(N+1)\mu+(2N-1)\mu^2]}{(1-N\mu)^2} + \frac{Nr^2\mu^2(N-1)(1-\mu)}{(1-N\mu)^2}.$$

3.3. Service Time, Intervisit Time and Cycle Time

Let S_i be a random variable denoting the length of the service period of queue i. The intervisit period of queue i is defined to be the period between two consecutive services of queue i. A cycle of queue i, consists of a service period followed by an intervisit period. Let I_i and C_i be random variables representing the length of the intervisit period and the length of the cycle, respectively. The length of a service period is given by $\tau_i(m) - \underline{\tau}_i(m)$, the length of an intervisit period is given by $\underline{\tau}_i(m+1) - \tau_i(m)$, and the length of the cycle is given by $\underline{\tau}_i(m+1) - \underline{\tau}_i(m)$. These measures are called the *service time*, the *intervisit time*, and the *cycle time* of station i, respectively. In addition we define

$$S_i(z) \triangleq \mathrm{E}[z^{\tau_i(m)-\underline{\tau}_i(m)}],$$

$$I_i(z) \triangleq \mathrm{E}[z^{\underline{\tau}_i(m+1)-\tau_i(m)}],$$

$$C_i(z) \triangleq \mathrm{E}[z^{\underline{\tau}_i(m+1)-\underline{\tau}_i(m)}].$$

It is easy to see that the behavior of these periods in our system is very similar to their behavior in the system where the queues are served in cyclic fashion (polling system). In both systems, a service period of queue i is followed by an intervisit period of queue i, and this is followed by another service period of queue i, etc. Thus, the relation between the variables representing the service time, the intervisit time, the cycle time and the number of customers present at the polling instants are identical for both systems. The relevant expressions (see (2) in this paper and (3.36a), (3.36b), (3.39b), (3.40a) and (3.40b) in Takagi 1986) are

$$\mathrm{E}[w^{\tau_i(n)-\underline{\tau}_i(n)}] = \mathrm{E}[\{\Theta_i(w)\}^{L_i(\underline{\tau}_i(n))}],$$

$$\mathrm{E}[S_i] = \mathrm{E}[L_i^*] \cdot \Theta_i^{(1)}(1),$$

$$\mathrm{Var}[S_i] = \frac{\mathrm{Var}[L_i^*]}{(1-\mu_i)^2} + \frac{\sigma_i^2\mathrm{E}[L_i^*]}{(1-\mu_i)^3}$$

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

$$\mathrm{E}[L_i^*] = \mu_i\mathrm{E}[I_i],$$

$$\mathrm{Var}[L_i^*] = \mu_i^2\mathrm{Var}[I_i] + \sigma_i^2\mathrm{E}[I_i]$$

$$C_i(z) = I_i[\Theta_i(z)],$$

$$\mathrm{E}[C_i] = \mathrm{E}[I_i] \cdot \Theta_i^{(1)}(1),$$

$$\mathrm{Var}[C_i] = \frac{\mathrm{Var}[I_i]}{(1-\mu_i)^2} + \frac{\mathrm{E}[I_i]\sigma_i^2}{(1-\mu_i)^3}.$$

Using these relations and (6) we get the expected value and variance of the cycle time (the expressions for the expected value and variance of the service time and of the intervist time can similarly be derived and may be found in (4.32) and (4.36) of Levy 1984):

$$\mathrm{E}[C_i] = \frac{\sum_{j=1}^N p_j r_j}{p_i(1-\sum_{j=1}^N \mu_j)},$$

$$\mathrm{Var}[C_i] = \frac{1}{\mu_i^2(1-\mu_i)^2} \cdot \left[\mathrm{Var}[L_i^*] + \frac{\sigma_i^2(\mu_i^2+\mu_i-1)\sum_{j=1}^N p_j r_j}{p_i(1-\sum_{j=1}^N \mu_j)}\right] \quad (10)$$

In the fully symmetric system these expressions become

$$\mathrm{E}[C_i] = \frac{Nr}{1 - N\mu},$$

$$\mathrm{Var}[C_i] = \frac{1}{1-\mu} \cdot \left[\frac{N\delta^2}{1-N\mu} + \frac{(N-1)Nr^2}{(1-N\mu)^2} + \frac{N^2\sigma^2 r}{(1-N\mu)^2}\right].$$

3.4. The Waiting Times and the Number of Customers at Arbitrary Times

Let $Q_i(z)$ denote the generating function of the number of customers found at queue i at an arbitrary time, when the system is in equilibrium: $Q_i(z) \triangleq \mathrm{E}[z^{L_i}]$. This generating function can be related to the generating functions $F_i(z)$ and $P_i(z)$ as follows (for details see (3.51) in Takagi 1986 regarding the derivation of a similar relation for the cyclic system);

$$Q_i(z) = \frac{p_i(1 - \sum_{j=1}^{N} \mu_j)}{\sum_{j=1}^{N} p_j r_j} \cdot \left[z\frac{F_i(z) - 1}{z - P_i(z)} + \frac{1 - F_i(z)}{1 - P_i(z)}\right]. \tag{11}$$

From (11) one may get:

$$\mathrm{E}[L_i] = \frac{\mathrm{E}[\{L_i^*\}^2]}{2\mathrm{E}[L_i^*]} + \frac{\sigma_i^2}{2}\left(\frac{1}{1-\mu_i} - \frac{1}{\mu_i}\right). \tag{12}$$

In the case of fully symmetric stations, we use (6) and (9) to get

$$\mathrm{E}[L_i] = \frac{1}{2} \cdot \left[\frac{\delta^2\mu}{r} + \frac{\sigma^2}{1-N\mu} + \frac{Nr\mu(1-\mu)}{1-N\mu} + \frac{(N-1)r\mu}{1-N\mu}\right]. \tag{13}$$

Next we calculate the waiting times (in queue) and response times observed in the system. Let c_j be an arbitrary customer. Recalling that customers arrive at the system in batches, we realize that the waiting time of c_j consists of the sum of two independent random variables:

1. The waiting time of the first customer in the batch in which c_j arrives.
2. The service time of all the customers which arrive together with c_j (the same batch) and are served ahead of c_j. Recall that all service times are equal to the slot size.

Let $\tilde{W}_i$ denote the waiting time of the first customer served in a batch (for a batch that arrives to queue i) and let $\tilde{W}_i(z)$ be the generating function of $\tilde{W}_i$. Let V_i be the number of customers who arrive together with c_j to queue i (in the same batch) and who are served before c_j, and let $V_i(z)$ be the generating function of V_i. Let W_i denote the waiting time of an arbitrary customer served in station i, and $W_i(z)$ denote the generating function of W_i. Let T_i denote the response time (waiting plus service time) of an arbitrary customer served in station i and $T_i(z)$ be the corresponding generating function.

The generating function of the waiting time observed by an arbitrary customer can be calculated from $\tilde{W}_i(z)$ and $V_i(z)$: $W_i(z) = \tilde{W}_i(z) \cdot V_i(z)$.

It is straightforward to calculate $V_i(z)$ from the generating function of the batch size, $P_i(z)$, and from its first moment, μ_i (see Takagi 1986, Equation 3.8a):

$$V_i(z) = \frac{1 - P_i(z)}{\mu_i(1 - z)}.$$

The generating function of the waiting time for a first customer in a batch, $\tilde{W}_i(z)$, can be calculated from the generating function of the idle period length, $I_i(z)$, and from the expected cycle length, $\mathrm{E}[C_i]$, as follows:

$$\tilde{W}_i(z) = \frac{1}{\mathrm{E}[C_i]} \cdot \frac{I_i(z) - 1}{z - P_i(z)}.$$

The derivation of this expression can be found in Takagi (1986) (Equations 3.57a and 3.57b) for the cyclic system and can be shown to hold for our system as well. Using (10), we get $\tilde{W}_i(z)$ in terms of $I_i(z)$ and the system parameters

$$\tilde{W}_i(z) = \frac{p_i(1 - \sum_{j=1}^{N} \mu_j)}{\sum_{j=1}^{N} p_j r_j} \cdot \frac{I_i(z) - 1}{z - P_i(z)}.$$

Since $I_i(z)$ can be calculated from $F_i(z)$, this equation actually expresses $\tilde{W}_i(z)$ in terms of $F_i(z)$.

Next, to calculate the expected value of the response time observed by an arbitrary customer in queue i we apply Little's result to (12) and (13). This yields

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

which in the case of fully symmetric stations becomes

$$\mathrm{E}[T_i] = \frac{1}{2} \cdot \left[\frac{\delta^2}{r} + \frac{\sigma^2}{(1-N\mu)\mu} + \frac{Nr(1-\mu)}{1-N\mu} + \frac{(N-1)r}{1-N\mu}\right].$$

3.5. Conditions for Steady State

The scope of this paper is too limited to supply a detailed analysis for the convergence of the system variables to steady state. Nevertheless, since our main results regard the system *moments* (first and second) at steady state, we now substantiate the conditions under which these results hold. The steady state moments derived in this section are all expressed in terms of $f(i)$ and $f(i, j)$. Therefore, it is sufficient to find conditions under which $f_m(i)$ and $f_m(i, j)$ are guaranteed to reach steady state.

The expressions for $f_m(i)$ can be obtained by unconditioning (3) and differentiating it with respect to z_i. This yields

$$f_{m+1}(j) = \frac{\mu_j}{p_j} \cdot \left(\sum_{i=1}^{N} p_i r_i + \sum_{\substack{i=1 \\ i \neq j}}^{N} \frac{p_i f_m(i)}{1 - \mu_i} \right).$$

This relation, which transforms $f_m(i)$ to $f_{m+1}(i)$, is shown in Levy (1986) to be a contraction mapping provided that $\sum_{i=1}^{N} \mu_i < 1$. Thus, under these conditions, $f_m(i)$ is guaranteed to converge independently of the initial values $f_0(i)$ and the first moments of the number of customers present in the system at polling instants are guaranteed to reach equilibrium.

The expressions for $f_m(i, j)$ are obtained in a similar manner (unconditioning (3) and twice differentiating it). The resulting relation which transforms $f_m(i, j)$ to $f_{m+1}(i, j)$ is also a contraction mapping under the condition $\sum_{i=1}^{N} \mu_i < 1$. This condition provides that $f_m(i, j)$ will reach steady state. We may therefore conclude that all our results regarding the system moments at steady state hold if $\sum_{i=1}^{N} \mu_i < 1$.

4. Analysis of the Gated Service Policy

As in the exhaustive system, the key to this analysis is the generating function of the number of customers found in the system at the end of a switchover period. This is $F_m(z_1, z_2, \ldots, z_N)$, as defined in (1).

For the gated policy, the length of the service period of station i is simply the number of customers found in queue i at the polling instant

$$\tau_i(m) - \underline{\tau}_i(m) = L_i(\underline{\tau}_i(m)).$$

Thus, the generating function of the number of customers arriving during this period is given by

$$\mathrm{E}\left[\prod_{j=1}^{N} \{P_j(z_j)\}^{\tau_i(m) - \underline{\tau}_i(m)} \right] = \mathrm{E}\left[\prod_{j=1}^{N} \{P_j(z_j)\}^{L_i(\underline{\tau}_i(m))} \right].$$

Thus, (3) is replaced by

$$F_{m+1}(z_1, z_2, \ldots, z_N \mid A_i) = R_i\left(\prod_{j=1}^{N} P_j(z_j) \right) \cdot F_m\left(z_1, z_2, \ldots, z_{i-1}, \prod_{j=1}^{N} P_j(z_j), z_{i+1}, \ldots, z_N\right)$$

and (4) is replaced by

$$F(z_1, z_2, \ldots, z_N) = \sum_{i=1}^{N} p_i \cdot F(z_1, \ldots, z_N \mid A_i) \tag{14a}$$

where

$$F(z_1, \ldots, z_N \mid A_i) = R_i\left(\prod_{j=1}^{N} P_j(z_j) \right) \cdot F\left(z_1, z_2, \ldots, z_{i-1}, \prod_{j=1}^{N} P_j(z_j), z_{i+1}, \ldots, z_N\right). \tag{14b}$$

Defining the moments of $L_i^*(f(i), f(i, j), f(i \mid k)$ and $f(i, j \mid k))$ as in the exhaustive model and differentiating (14b), we get the following set of equations:

$$f(i \mid i) = r_i \mu_i + \mu_i f(i)$$

$$f(j \mid i) = r_i \mu_j + \mu_j f(i) + f(j) \quad i \neq j.$$

The solution of these equations (see Levy 1984) is

$$\mathrm{E}[L_j^*] = f(j) = \frac{\mu_j}{p_j} \cdot \frac{\sum_{i=1}^{N} p_i r_i}{1 - \sum_{i=1}^{N} \mu_i}.$$

When $p_i = 1/N$ for every i this is identical to the equivalent expression in the gated system where the polling is done in a cyclic fashion (Rubin and De Moraes). In the case of fully symmetric stations $\mathrm{E}[L_j^*]$ is

$$\mathrm{E}[L_j^*] = f(j) = \frac{N r \mu}{1 - N\mu}. \tag{15}$$

To find the variance of L_i^* we differentiate (14b) twice. This gives the following set of equations:

$$f(j, k \mid i) = \mu_j \mu_k (\delta_i^2 + r_i^2) + r_i \mu_k f(j) + r_i \mu_j f(k) + f(i) \mu_j \mu_k (2r_i + 1) + f(j, k) + \mu_j f(i, k) + \mu_k f(i, j) + \mu_j \mu_k f(i, i) \quad i \neq j, i \neq k, j \neq k \tag{16a}$$

$$f(j, j \mid i) = \mu_j^2(\delta_i^2 + r_i^2) + r_i(\sigma_j^2 - \mu_j) + 2r_i\mu_j f(j) + f(i)[\sigma_j^2 - \mu_j + \mu_j^2(r_i + 1)] + f(j, j) + 2\mu_j f(i, j) + \mu_j^2 f(i, i) \quad i \neq j \tag{16b}$$

$$f(j, k \mid j) = \mu_j\mu_k(\delta_j^2 + r_j^2) + r_j\mu_j f(k) + f(j)\mu_j\mu_k(2r_j + 1) + \mu_j f(j, k) + \mu_j\mu_k f(j, j) \quad j \neq k \tag{16c}$$

$$f(j, i \mid j) = \mu_j^2(\delta_j^2 + r_j^2) + r_j(\sigma_j^2 - \mu_j) + f(j)[\sigma_j^2 - \mu_j + \mu_j^2(2r_j + 1)] + \mu_j^2 f(j, j). \tag{16d}$$

These equations, together with the relation $f(j, k) = \sum_{i=1}^{N} p_i \cdot f(j, k \mid i)$, form a set of N^2 linear equations that can be solved by numerical methods (Levy 1986) to yield the solution of $f(i, i)$ for $i = 1, 2, \ldots, N$.

In the case of fully symmetric stations this set of equations can be solved analytically (see Levy 1984) to yield

$$f(i, i) = \frac{\sigma^2 rN[1 - (N - 1)\mu]}{(1 + \mu)(1 - N\mu)^2} + \frac{(\delta^2 - r^2)N\mu^2}{(1 + \mu)(1 - N\mu)} + \frac{(\mu + 2r)N^2 r\mu^2}{(1 + \mu)(1 - N\mu)^2} - \frac{\mu Nr}{(1 + \mu)(1 - N\mu)} - \frac{\mu^2 Nr}{(1 + \mu)(1 - N\mu)^2}. \tag{17}$$

Now, using (15) and (17) we get

$$\mathrm{Var}[L_i^*] = \frac{\delta^2\mu^2 N}{(1 + \mu)(1 - N\mu)} + \frac{\sigma^2 rN[1 - (N - 1)\mu]}{(1 + \mu)(1 - N\mu)^2} + \frac{(N - 1)N\mu^2 r^2}{(1 + \mu)(1 - N\mu)^2}.$$

This is the variance of the number of customers found in queue i at polling instants.

To calculate the cycle time, note that in a gated system the generating function of the cycle length is related to the generating function of the number of customers found in queue i at polling instants as

$$F_i(z) = C_i[P_i(z)]. \tag{18}$$

From (18) the mean and the variance of the cycle time can be easily calculated:

$$\mathrm{E}[C_i] = \frac{\mathrm{E}[L_i^*]}{\mu_i} = \frac{1}{p_i} \cdot \frac{\sum_{j=1}^{N} p_i r_i}{1 - \sum_{i=1}^{N} \mu_i};$$

$$\mathrm{Var}[C_i] = \frac{\mathrm{Var}[L_i^*]}{\mu_i^2} - \frac{\mathrm{E}[L_i^*] \cdot \sigma_i^2}{\mu_i^3}.$$

In the fully symmetric case these become

$$\mathrm{E}[C_i] = \frac{Nr}{1 - N\mu},$$

$$\mathrm{Var}[C_i] = \frac{N\delta^2}{(1 + \mu)(1 - N\mu)} + \frac{N^2\sigma^2 r}{(1 + \mu)(1 - N\mu)^2} + \frac{(N - 1)Nr^2}{(1 + \mu)(1 - N\mu)^2}.$$

Next the generating function of the number of customers found in queue i at arbitrary moments may be calculated. This is done by using expressions that relate the number of customers found in the system at arbitrary moments to the cycle time and to the number of customers found in the system at polling instants. These expressions have been derived for the cyclic polling system (see, for example, (5.14) and (5.15a) in Takagi 1986) and can be easily shown to hold for our system too. These relations are

$$Q_i(z) = \frac{1}{\mathrm{E}[C_i]} \cdot \frac{F_i[P_i(z)] - F_i(z)}{P_i(z) - z} \cdot \frac{(1 - z)P_i(z)}{1 - P_i(z)};$$

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

From these, we can now calculate the expected value of the number of customers found in queue i at arbitrary moments. For a fully symmetric system this value is

$$\mathrm{E}[L_i] = \frac{\delta^2\mu}{2r} + \frac{\sigma^2}{2(1 - N\mu)} + \frac{Nr\mu(1 + \mu)}{2(1 - N\mu)} + \frac{(N - 1)r\mu}{2(1 - N\mu)}. \tag{19}$$

To calculate the waiting time in the system, we again recall a relation from the analysis of the cylic polling, gated service system (5.18 in Takagi 1986):

$$\tilde{W}_i(z) = \frac{z[C_i(z) - F_i(z)]}{\mathrm{E}[C_i] \cdot (z - P_i(z))}$$

which is also valid for our system. From this expression the expected waiting time of a first customer in a batch can be calculated.

Lastly, application of Little's result to (19) yields the expected response time for an arbitrary customer in a symmetric system:

$$\mathrm{E}[T_i]=\frac{\delta^2}{2r}+\frac{\sigma^2}{2\mu(1-N\mu)}+\frac{Nr(1+\mu)}{2(1-N\mu)}+\frac{(N-1)r}{2(1-N\mu)}.$$

The conditions under which the moments of the system variables reach equilibrium are identical to those of the exhaustive system (namely, $\sum_{i=1}^{N}\mu_i<1$). The arguments supporting this claim are identical to those provided in Section 3.5.

5. Analysis of the Limited Service Policy

5.1. The Expected Response Time in a Symmetric System

As in the previous analysis, the key to this analysis is the generating function of the number of customers found in the system at *polling instants.* This is $F_m(z_1, z_2, \ldots, z_N)$, as defined in (1).

To express $\mathrm{F}_{m+1}(z_1, z_2, \ldots, z_N)$ in terms of $F_m(z_1, z_2, \ldots, z_N)$, we condition $F_{m+1}(z_1, z_2, \ldots, z_N)$ on the station polled during the mth cycle:

$$\begin{aligned}
&\mathrm{F}_{m+1}(z_1, z_2, \ldots, z_N \mid A_i)\\
&\quad=R_i\left(\sum_{j=1}^{N}P_j(z_j)\right)\cdot\left(\prod_{j=1}^{N}P_j(z_j)\right)\\
&\qquad\cdot\frac{1}{z_i}\left[F_m(z_1, z_2, \ldots, z_N)\right.\\
&\qquad\qquad\left.-F_m(z_1, \ldots, 0, \ldots, z_N)\right]\\
&\qquad+R_i\left(\prod_{j=1}^{N}P_j(z_j)\right)\cdot F_m(z_1, \ldots, 0, \ldots, z_N) \qquad (20)
\end{aligned}$$

where $F_m(\mathrm{z}_1, \ldots, 0, \ldots \mathrm{z}_N)$ is $F_m(z_1, z_2, \ldots, z_N)$ where the ith element equals zero. The first term of this expression represents the situation where queue i is not empty when polled, so queue j "builds up" during the service slot by a factor of $P_j(\mathrm{z}_j)$ and one customer is removed from the ith buffer. The second term represents the situation where queue i is empty when polled, so no service period follows this polling instant. In both terms, the factor $R_i(\prod_{j=1}^{N}P_j(z_j))$ represents the queueing build up during the switchover period prior to the $m+1$st polling instant.

From (20), and under equilibrium conditions, we get the following relation

$$\begin{aligned}
&F(z_1, \ldots, z_N)\\
&\quad=p_1\cdot R_1\left(\prod_{j=1}^{N}P_j(z_j)\right)\\
&\qquad\cdot\left[\left(\prod_{j=1}^{N}P_j(z_j)\right)\cdot F(z_1, z_2, \ldots, z_N)\cdot\frac{1}{z_1}\right.\\
&\qquad\qquad\left.+\left(1-\frac{1}{z_1}\prod_{j=1}^{N}P_j(z_j)\right)F(0, z_2, \ldots, z_N)\right]\\
&\qquad+p_2\cdot R_2\left(\prod_{j=1}^{N}P_j(z_j)\right)\\
&\qquad\cdot\left[\left(\prod_{j=1}^{N}P_j(z_j)\right)\cdot F(z_1, \ldots, z_N)\cdot\frac{1}{z_2}\right.\\
&\qquad\qquad\left.+\left(1-\frac{1}{z_2}\prod_{j=1}^{N}P_j(z_j)\right)F(z_1, 0, z_3, \ldots, z_N)\right]\\
&\qquad+\cdots+p_N\cdot R_N\left(\prod_{j=1}^{N}P_j(z_j)\right)\\
&\qquad\cdot\left[\left(\prod_{j=1}^{N}P_j(z_j)\right)\cdot F(z_1, \ldots, z_N)\cdot\frac{1}{z_N}\right.\\
&\qquad\qquad\left.+\left(1-\frac{1}{z_N}\prod_{j=1}^{N}P_j(z_j)\right)F(z_1, \ldots, z_{N-1}, 0)\right]. \qquad (21)
\end{aligned}$$

In the following, we analyze the fully symmetric system. The analysis approach partially follows the approach used in the analysis of the limited service cycle system (originally reported in Takagi and Kleinrock 1983); here we extend it to derive the expected response time in the system (this extension was later used by Takagi (1985) to derive the expected response time in the cyclic system). Assuming symmetry and substituting $z_1=z_2=\ldots=z_N=z$ in (21) we get

$$\begin{aligned}
&F(z, z, \ldots, z)\\
&\quad=\frac{1}{z}R(\{P(z)\}^N)\cdot\{P(z)\}^N F(z, z, \ldots, z)\\
&\qquad+R(\{P(z)\}^N)\cdot\left(1-\frac{\{P(z)\}^N}{z}\right)\\
&\qquad\cdot F(0, z, z, \ldots, z) \qquad (22)
\end{aligned}$$

where we have used the observation that $F(0, z, z, \ldots, z)=F(z, 0, z, \ldots, z)=\ldots=F(z, z, \ldots, z, 0)$

due to symmetry. From (22) we have

$$F(z, z, \ldots, z) = \frac{R(\{P(z)\}^N) \cdot (z - \{P(z)\}^N) \cdot F(0, z, z, \ldots, z)}{z - R(\{P(z)\}^N) \cdot \{P(z)\}^N}. \quad (23)$$

Next, we substitute $z_1 = z$ and $z_2 = z_3 = \ldots = z_N = 1$ into (21). This yields

$$F(z, 1, 1, \ldots, 1) = \frac{1}{N} \cdot \left[R(P(z)) \cdot P(z) \cdot F(z, 1, 1, \ldots, 1) \frac{1}{z} + R(P(z)) \cdot \left(1 - \frac{P(z)}{z}\right) \cdot F(0, 1, 1, \ldots, 1) \right] + \frac{N-1}{N} \left[R(P(z)) \cdot P(z) \cdot F(z, 1, 1, \ldots, 1) + R(P(z)) \cdot (1 - P(z)) \cdot F(z, 0, 1, 1, \ldots, 1) \right] \quad (24)$$

where we have used the symmetry observations:

$$F(0, 1, 1, \ldots, 1) = F(1, 0, 1, \ldots, 1) = \ldots = F(1, 1, 1, \ldots, 1, 0),$$

$$F(z, 0, 1, 1, \ldots, 1) = F(z, 1, 0, 1, \ldots, 1, 1) = \cdots = F(z, 1, 1, \ldots, 1, 1, 0),$$

$$p_1 = p_2 = \cdots = p_N = \frac{1}{N}.$$

From (24) get

$$F(z, 1, 1, \ldots, 1) = \frac{(N-1)zR(P(z)) \cdot (1 - P(z)F(z, 0, 1, 1, \ldots, 1)}{Nz - R(P(z)) \cdot P(z) \cdot (1 + (N-1)z)} + \frac{R(P(z)) \cdot (z - P(z)) \cdot F(0, 1, 1, \ldots, 1)}{Nz - R(P(z)) \cdot P(z) \cdot (1 + (N-1)z)}. \quad (25)$$

The next step is to calculate the probability that an arbitrary queue is empty at polling instants. This probability is given by $f_0 \triangleq F(0, 1, 1, \ldots, 1)$. From (23) we may calculate f_0 (see Appendix C.1 in Levy 1984):

$$f_0 = \frac{1 - N\mu - Nr\mu}{1 - N\mu}. \quad (26)$$

Next the expected queue length at polling instants is calculated using two simple relations; the first is

$$\left.\frac{\partial F(z, z, \ldots, z)}{\partial z}\right|_{z=1} = N \cdot \left.\frac{\partial F(z, 1, 1, \ldots, 1)}{\partial z}\right|_{z=1}. \quad (27)$$

This relation simply states that (at polling instants) the expected number of customers in the whole system is N times the expected number of customers in queue i ($i = 1, 2, \ldots, N$). This observation is true due to symmetry. The second relation is

$$\left.\frac{\partial F(0, z, z, \ldots, z)}{\partial z}\right|_{z=1} = (N-1) \cdot \left.\frac{\partial F(z, 0, 1, 1, \ldots, 1)}{\partial z}\right|_{z=1} \quad (28)$$

which is also true due to symmetry.

For convenience let us introduce the additional notation: $f_1 \triangleq \partial F(z, 0, 1, 1, \ldots, 1)/\partial z\,|_{z=1}$.

Differentiating (23) and using (28) we show (see Appendix C.2 in Levy 1984)

$$\left.\frac{\partial F(z, z, \ldots, z)}{\partial z}\right|_{z=1} = \frac{(N-1)(1 - N\mu)f_1}{1 - N\mu - Nr\mu} + \frac{Nr\sigma^2}{2(1 - N\mu)(1 - N\mu - Nr\mu)} + \frac{N^2\mu^2\delta^2}{2(1 - N\mu - Nr\mu)} + \frac{Nr\mu}{2}. \quad (29)$$

Differentiating (25) and using (28) we show (see Appendix C.3 in Levy 1984)

$$N \cdot \left.\frac{\partial F(z, 1, 1, \ldots, 1)}{\partial z}\right|_{z=1} = \frac{-(N-1)N\mu f_1}{1 - N\mu - Nr\mu} + \frac{\mathscr{A}}{2(1 - N\mu)(1 - N\mu - Nr\mu)} \quad (30)$$

where $\mathscr{A} = N \cdot [N^2 r\mu^3 + N\mu^2(1 - N\mu)(\delta^2 - r^2) - 2Nr\mu^2 + (\sigma^2 + \mu)Nr]$.

Now, using (27) we equate (29) to (30) and solve (see Appendix C.4. of Levy 1984) for f_1:

$$f_1 = \frac{(\sigma^2 + \mu)Nr}{2(1 - N\mu)}. \quad (31)$$

Substituting (31) back into (30) we finally get an expression for the expected queue length at polling instants

$$\left.\frac{\partial F(z, 1, 1, \ldots, 1)}{\partial z}\right|_{z=1}$$
$$= \frac{(N-1)(\sigma^2+\mu)r}{2(1-N\mu-Nr\mu)} + \frac{r\sigma^2}{2(1-N\mu)(1-N\mu-Nr\mu)} + \frac{N\mu^2\delta^2}{2(1-N\mu-Nr\mu)} + \frac{r\mu}{2}. \quad (32)$$

Having calculated the expected queue length of queue i at polling instants, we next calculate the expected queue length of queue i right after an arbitrary customer leaves this queue. Let us denote

$G_i \triangleq \mathrm{E}[L_i(t \mid t$ is a service completion time at queue $i)]$.

First, since the queue chosen to be polled at a given polling instant is independent of the system status, it is clear that

$$\mathrm{E}[L_i(t \mid \text{ is a service starting time at queue } i)] = \frac{1}{1-f_0} \cdot \left.\frac{\partial F(z, 1, 1, \ldots, 1)}{\partial z}\right|_{z=1}.$$

Second, we have

$$G_i = \mathrm{E}[L_i(t \mid t \text{ is a service starting time at queue } i)] + \mu - 1.$$

Thus, from these two relations and from (26) and (32) we get

$$G_i = \frac{(N-1)(\sigma^2+\mu)(1-N\mu)}{2N\mu(1-N\mu-Nr\mu)} + \frac{\sigma^2}{2N\mu(1-N\mu-Nr\mu)} + \frac{N\mu^2\delta^2}{2(1-N\mu-Nr\mu)} + \frac{\mu\delta^2}{2r} + \frac{1-N\mu}{2N} + \mu - 1. \quad (33)$$

This is the expected queue length at station i right after an arbitrary customer leaves this station.

Next, using G_i, the expected response time (waiting time plus service time) of an arbitrary customer is calculated. To calculate $\mathrm{E}[T_i]$ we investigate the number of customers left in queue i behind an arbitrary tagged customer, say c_j. These customers are of two types:

1. Customers who arrive together with c_j (in the same batch) but who are queued behind c_j.
2. Customers who arrive to queue i during the response time of c_j.

Let $\tilde{V}_i$ be the number of customers arriving to queue i together with c_j (the same batch) but queued behind c_j; then, the following relation is a direct result of the above observation:

$$G_i = \mathrm{E}[\tilde{V}_i] + \mu \cdot \mathrm{E}[T_i]. \quad (34)$$

To find $\mathrm{E}[\tilde{V}_i]$, we assume that c_j arrives at slot t and condition on the number of customers arriving during that slot: $\mathrm{E}[\tilde{V}_i \mid X_i(t) = k] = (k-1)/2$. The probability that c_j arrives in a batch of size k is given by:

$$\frac{k \cdot \Pr[X_i(t) = k]}{\sum_{l=1}^{\infty} l \cdot \Pr[X_i(t) = l]}.$$

Thus, unconditioning $\mathrm{E}[\tilde{V}_i]$ yields

$$\mathrm{E}[\tilde{V}_i] = \sum_{k=1}^{\infty} \frac{k-1}{2} \cdot \frac{k \cdot \Pr[X_i(t) = k]}{\sum_{l=1}^{\infty} l \cdot \Pr[X_i(t) = l]} = \frac{\sigma^2 + \mu^2 - \mu}{2\mu}. \quad (35)$$

Substituting (35) and (33) into (34) we finally get the expected response time of an arbitrary customer in the system

$$\mathrm{E}[T_i] = \frac{\delta^2}{2r} + \frac{\sigma^2}{2\mu(1-N\mu-Nr\mu)} + \frac{Nr\sigma^2}{2\mu(1-N\mu-Nr\mu)} + \frac{(N-1)r}{2(1-N\mu-Nr\mu)} + \frac{N\delta^2\mu}{2(1-N\mu-Nr\mu)}.$$

5.2. The Probability of an Empty Buffer

An important measure of a queueing system is the fraction of time that the system is empty. In this subsection we are interested in calculating the probability that a buffer is empty at some specific instants.

The probability that a buffer is empty at polling instants was calculated above as:

$$\Pr[\text{queue } i \text{ is empty at polling instants}] = F(0, 1, 1, \ldots, 1) = \frac{1-N\mu-Nr\mu}{1-N\mu}.$$

From this measure it is now easy to calculate the probability that a buffer is empty at switchover times.

A *switchover time* is the instant at which a switchover period starts. Thus, the mth switchover time is denoted by $\tau(m)$. Let s_0 denote the probability that buffer i is empty at switchover times:

$$s_0 \triangleq \lim_{m\to\infty} \Pr[L_i(\tau(m)) = 0]; \quad i = 1, 2, \ldots, N.$$

Since every polling instant is the end of a switchover period, the probability that buffer i is empty at a polling instant is related to the probability that this buffer is empty at the beginning of the preceding switchover period as follows:

$$\begin{aligned}&\Pr[L_i(\tau(m)) \\ &\quad = 0 \mid \tau(m), \tau(m)] \\ &\quad = \Pr[L_i(\tau(m)) = 0]\left(\sum_{t=\tau(m)+1}^{\tau(m)} \Pr[X_i(t) = 0]\right). \qquad (36)\end{aligned}$$

Now, since the arrival process at station i is independent of t, and since in the symmetric case it is also independent of i, the following notation can be used:

$$x_0 \triangleq \Pr[X_i(t) = 0]; \quad i = 1, 2, \ldots, N.$$

Letting $m \to \infty$ in (36), substituting s_0 and unconditioning (36) yields

$$\begin{aligned}&\lim_{m\to\infty} \Pr[L_i(\tau(m)) = 0] \\ &\quad = \lim_{m\to\infty} \Pr[L_i(\tau(m)) = 0] \cdot R(x_0) \qquad (37)\end{aligned}$$

and finally, from (37) and (26) we have the probability that a buffer is empty at an arbitrary switchover instant;

$$\Pr[\text{queue } i \text{ is empty at switchover instants}] = \frac{1 - N\mu - Nr\mu}{(1 - N\mu) \cdot R(x_0)}.$$

6. Systems with Zero Length Switchover Periods

The analysis provided above is based on the assumption that at least one of the switchover periods is not deterministically of zero length. A natural question to ask is how our results relate to systems where *all* the switchover periods are of zero length (which we denote below as systems with *zero reply intervals*). The problem of relating cyclic polling systems with zero reply intervals to systems with non-zero reply intervals has been raised by several authors (e.g., Eisenberg, pp. 441, Humblet, pp. 166 and Takagi 1986, pp. 142). Nevertheless, the problem was not addressed in any of those references in much detail, and thus, we discuss it below.

At first observation it seems that the analysis method used in our paper does not apply for systems with zero reply intervals. The reason is that when such a system empties, the server polls the queues infinitely many times in zero time. The expressions for the moments of the number of customers in the system at polling instants (e.g., (6) and (8) for the exhaustive system) shrink to zero and the use of these expressions for calculating the expected delay (e.g., (12) and (13)) is not feasible.

Nevertheless, a more careful examination shows that by *properly* taking limits on the *distribution* of the reply interval one can analyze systems with zero reply intervals using our analysis. The main idea is that the analysis approach is valid for any reply interval distribution which is not completely concentrated at zero. This is true since under such conditions the server will not poll the queues infinitely many times at a certain epoch t; rather, eventually it will "depart" from time t and poll the system again at time $t + k$ (for some $k > 0$).

Proper limits for the reply interval distribution should be taken so as to guarantee that in the limit the system will behave as a system with zero reply intervals. The crucial properties of the system with zero reply intervals are: 1) the server does not go idle unless the (whole) system becomes empty, and 2) when the system is empty and the server is idle, the server will be ready to start serving as soon as any customers arrive to the system. This behavior may be achieved by reply intervals which are of length 0 and 1 with probabilities $(1 - p)$ and p, respectively, and where p approaches zero. Under this distribution we have $r = p$ and $\delta^2 = p(1 - p)$, and thus the limit of $\delta^2/2r$ (which is a term in the expressions for the expected delay in all symmetric systems) exists and satisfies $\lim_{p\to 0} \delta^2/2 = \frac{1}{2}$.

Note that the use of other distributions for deriving the limits may be improper. For example, consider the case where the reply interval takes on the values 0 and k with probabilities $(1 - p)$ and p, respectively. The limiting behavior of this system will be such that after the system completely empties, the server takes a "vacation" whose length is k slots, and thus, is not always ready to serve customers as soon as they arrive at the empty system. The expected delay in this system (which behaves like a system with vacation periods) is obviously higher than that in the system with zero reply intervals. In the case of symmetric stations this difference is expressed in the term $\delta^2/2r$, which under these conditions, satisfies $\lim_{p\to 0} \delta^2/2r = k/2$.

The numerical stability of this procedure does not seem to be a problem. This is true although we deal with computing the values of two variables (namely, $f(i)$ and $f(i, i)$ in the exhaustive and gated systems) that vanish to zero. The reason is that in the computation of the expected delay only their ratio appears (see, e.g., (12)), which should not lead to numerical difficulties provided that the relative error in computing each of them is small enough. To examine this issue, we used the proposed procedure (namely, $r_i = p$ and $\delta_i^2 = p(1 - p)$ and letting p approach zero) for several fully symmetric systems and found good agreement between the expected delay values computed numerically to the ones derived by taking limits on (13).

It is important to mention that, to the best of our knowledge, cyclic polling systems with zero reply intervals have not been analyzed previously under the discrete time model (in contrast, such a treatment was given to the continuous time models, e.g., Cooper and Murray 1969, and Cooper). Thus the method suggested here is the only one currently available for analyzing these systems. A more detailed analysis of systems with zero length switchover periods may be found in Levy and Kleinrock (1987).

7. Comparison of the Results and Discussion

In this section, we compare the expected response time in the random polling systems. These results are also compared to the expected response time observed in the corresponding cyclic polling systems.

For the (discrete time) cyclic polling system, we assume the same arrival process as for the random system. The expected response time in the cyclic system was derived by Konheim and Meister for the exhaustive service policy, by Rubin and De Moraes for the gated service policy and by Takagi (1985) for the limited service policy. (Note that the original expressions derived by Konheim and Meister and extended by Swartz differ from our expressions in two aspects: First, those models assume that arrivals occur at the beginning of a slot while we assume that the arrivals occur at the end of the slot. Second, Konheim and Meister calculate the expected waiting time of the first customer in a batch, while we calculate the expected response time of an arbitrary customer. Note also that the results derived by Takagi (1986) are smaller than ours by one unit because our expressions include the customer service time while his expressions do not.) The expressions for the expected response time in the exhaustive, gated and limited service systems can be found in a unified form in Takagi (1986) in Equations 3.61b, 5.21b and 6.64, respectively. These results and the results derived in this paper are summarized in Table I.

Looking at the stability conditions, we see that both the gated and the exhaustive system are stable (under both types of polling methods) as long as $N\mu < 1$. On the other hand, the limited service scheme is stable only as long as $N\mu(1 + r) < 1$. This result is intuitive since at least one switchover period (whose expected length is r) is associated with every customer served.

Comparison of polling methods shows that for all

Table I
Expected Response Time in the Different Systems

Service Method	Polling Method	
	Cyclic	Random
Exhaustive	$\frac{\delta^2}{2r} + \frac{\sigma^2}{2\mu(1 - N\mu)} + \frac{Nr(1 - \mu)}{2(1 - N\mu)}$	$\frac{\delta^2}{2r} + \frac{\sigma^2}{2\mu(1 - N\mu)} + \frac{Nr(1 - \mu)}{2(1 - N\mu)} + \frac{(N - 1)r}{2(1 - N\mu)}$
Gated	$\frac{\delta^2}{2r} + \frac{\sigma^2}{2\mu(2 - N\mu)} + \frac{Nr(1 + \mu)}{2(1 - N\mu)}$	$\frac{\delta^2}{2r} + \frac{\sigma^2}{2\mu(1 - N\mu)} + \frac{Nr(1 + \mu)}{2(1 - N\mu)} + \frac{(N - 1)r}{2(1 - N\mu)}$
Limited	$\frac{\delta^2}{2r} + \frac{(1 + Nr)\sigma^2}{2\mu(1 - N\mu - Nr\mu)} + \frac{N\delta^2\mu}{2(1 - N\mu - Nr\mu)}$	$\frac{\delta^2}{2r} + \frac{(1 + Nr)\sigma^2}{2\mu(1 - N\mu - Nr\mu)} + \frac{N\delta^2\mu}{2(1 - N\mu - Nr\mu)} + \frac{(N - 1)r}{2(1 - N\mu - Nr\mu)}$

three service policies the expected response time of the random polling scheme is greater than the expected response time of the corresponding cyclic polling scheme. This observation is quite intuitive due to the random behavior of the server in the random polling system. Note also that when the number of stations is $N = 1$, then the response time in the random system is identical to the response time in the cyclic system.

The difference between the mean response time of the random polling system and that of the corresponding cyclic system is $(N - 1)r/2(1 - N\mu)$ for the exhaustive and gated schemes, and $(N - 1)r/2(1 - N\mu - Nr\mu)$ for the limited service scheme. In the cases of the exhaustive and the gated systems this difference is exactly the expected length of a period consisting of $(N - 1)/2$ service periods plus $(N - 1)/2$ switchover periods. This difference can be explained as follows. Let t_0 be the time of an arbitrary arrival to queue i. Let t_1 be the time when the server first starts polling after t_0. Let t_2 be the first time the server polls queue i after t_0. The period between t_1 and t_2 consists, on average, of $(N - 1)/2$ service periods and $(N - 1)/2$ switchover periods in the cyclic system. On the other hand, this period consists, on average, of $N - 1$ service periods and $N - 1$ switchover periods in the random system (this value can be easily calculated by noticing that the number of times the server polls the system unit it hits queue i has a geometrically-shifted distribution with parameter $1/N$). Thus, the difference between the expected length of this period in the random system, and the corresponding period in the cyclic system consists of $(N - 1)/2$ service and switchover periods. Therefore, in the exhaustive and gated schemes the difference in the expected response time between the cyclic and the random systems can be attributed to the period between t_1 and t_2.

Comparing the exhaustive service to the gated service (in both types of polling methods) we see that the expected response time in the gated system is higher. The difference in performance between the exhaustive and the gated schemes is the same for both types of polling methods.

A more difficult task is to compare the limited service to the gated system. In Theorem 1, it is shown that the expected response time observed in the limited service system is greater than or equal to the expected response time in the gated system.

Theorem 1. *In the stable random polling system, the expected response time in the limited service scheme is greater than or equal to the expected response time in the gated service scheme.*

Proof. Let Δ be a function representing the difference between the expected response time in the limited service system to the expected response time in the gated system, namely:

$$\begin{aligned}\Delta &\triangleq T_{\text{RANDOM;LIMITED}} - T_{\text{RANDOM;GATED}} \\ &= \frac{(1 + Nr)\sigma^2}{2\mu(1 - N\mu - Nr\mu)} + \frac{N\delta^2\mu}{2(1 - N\mu - Nr\mu)} \\ &\quad + \frac{(N-1)r}{2(1 - N\Sigma - Nr\mu)} \\ &\quad - \left[\frac{\sigma^2}{2\mu(1 - N\mu)} + \frac{Nr(1+\mu)}{2(1 - N\mu)} + \frac{(N-1)r}{2(1 - N\mu)}\right]\end{aligned}$$

To prove the theorem, one has to show that $\Delta \geq 0$ for any arbitrary distribution of the switchover period and for any arrival process. By observing that the moments of any discrete nonnegative random variable X (i.e., a variable that takes on the values 0, 1, 2, ...) obey $\text{Var}(X) \geq \text{E}[X](1 - \text{E}[X])$ (because $\text{E}[X^2] \geq \text{E}[X]$), we have: $\sigma^2 \geq \mu(1 - \mu)$ and $\delta^2 \geq r(1 - r)$. Also, a sufficient condition for stability is easily shown to be $Nr\mu + N\mu < 1$. Thus, it is only required to show that $\Delta \geq 0$ for N, r, δ^2, μ and σ^2 such that: $N \geq 1$, $Nr\mu + N\mu < 1$, $\sigma^2 \geq \mu(1 - \mu)$ and $\delta^2 \geq r(1 - r)$.

First we prove the claim for $\sigma^2 = \mu(1 - \mu)$ and $\delta^2 = r(1 - r)$. Rewriting Δ we have

$$\begin{aligned}\Delta &= \frac{(N-1)r}{2} \cdot \left[\frac{1}{1 - N\mu - Nr\mu} - \frac{1}{1 - N\mu}\right] \\ &\quad + \frac{\sigma^2}{2\mu} \cdot \left[\frac{1}{1 - N\mu - Nr\mu} - \frac{1}{1 - N\mu}\right] \\ &\quad + \frac{Nr\sigma^2}{2\mu(1 - N\mu - Nr\mu)} + \frac{N\mu\delta^2}{2(1 - N\mu - Nr\mu)} \\ &\quad - \frac{Nr(1+\mu)}{2(1 - N\mu)}. \end{aligned} \tag{38}$$

Using simple inequalities and substituting $\sigma^2 = \mu(1 - \mu)$ and $\delta^2 = r(1 - r)$ into (38) we have

$$\begin{aligned}\Delta &\geq \frac{Nr}{2} \cdot \left[\frac{(1-\mu)\mu}{(1 - N\mu) \cdot (1 - N\mu - Nr\mu)}\right. \\ &\qquad \left. + \frac{1-\mu}{1 - N\mu - Nr\mu} + \frac{(1-r)\mu}{1 - N\mu - Nr\mu} - \frac{1+\mu}{1 - N\mu}\right] \\ &\geq \frac{Nr}{2} \cdot \left[\frac{\mu}{(1 - N\mu - Nr\mu)} + \frac{1-\mu}{1 - N\mu - Nr\mu}\right. \\ &\qquad \left. + \frac{(1-r)\mu}{1 - N\mu - Nr\mu} - \frac{1+\mu}{1 - N\mu}\right] \\ &= \frac{Nr}{2} \cdot \left[\frac{(N-1)r\mu + 2Nr\mu^2}{(1 - N\mu) \cdot (1 - N\mu - Nr\mu)}\right] \geq 0.\end{aligned}$$

Once the claim $\Delta \geq 0$ is proven for $\sigma^2 = \mu(1-\mu)$ and $\delta^2 = r(1-r)$, it is now easy to prove it also for $\sigma^2 \geq \mu(1-\mu)$ and $\delta^2 \geq r(1-r)$. This can be shown by observing that Δ is monotonically nondecreasing both in σ^2 and in δ^2.

Note that the proof of the above result implies an equivalent inequality for the cyclic polling system, namely

$$T_{\text{CYCLIC;LIMITED}} \geq T_{\text{CYCLIC;GATED}}.$$

While the expected response time in the cyclic polling systems was derived in Takagi (1985), this inequality for discrete time systems has not been proven previously (note, however, that the inequality was established for continuous time systems; see for example, Fuhrmann 1985 and Takagi 1985).

We can therefore conclude that for both polling schemes the mean response time increases as we go down the table, i.e.,

$$T_{\text{RANDOM;EXHAUSTIVE}} \leq T_{\text{RANDOM;GATED}} \leq T_{\text{RANDOM;LIMITED}}$$

$$T_{\text{CYCLIC;EXHAUSTIVE}} \leq T_{\text{CYCLIC;GATED}} \leq T_{\text{CYCLIC;LIMITED}}$$

and the mean response time increases as we go across the table, that is,

$$T_{\text{CYCLIC};x} \leq T_{\text{RANDOM};x}$$

where x is any of the service policies.

8. Summary

We have analyzed the performance of random polling systems under three service policies: exhaustive, gated and limited. We derived closed form expressions for the expected response time in all three systems under the assumption of full symmetry. For the nonsymmetric exhaustive and gated systems our analysis yields a set of N^2 linear equations the solution of which directly gives the expected response time in the system. Also derived in this paper are expressions for the number of customers in the system, cycle time, intervisit time and buffer utilization.

Appendix: Glossary of Notation

(All time units are measured in slots rather than seconds).

A_i The event that queue i was polled in the previous service period.

c_i The ith customer.

C_i, $C_i(z)$ The length of a cycle (for a system in equilibrium) and its generating function, respectively.

$F(z_1, z_2, \ldots, z_N)$ The generating function of the number of customers found in the system at polling instants.

$F_i(z)$ The generating function of the number of customers found at queue i at polling instants.

I_i, $I_i(z)$ The length of an idle period (in equilibrium) and its generating function, respectively.

L_i^* The number of customers found in queue i at polling instants (system in equilibrium.)

$L_i(t)$, L_i The number of customers in queue i at time t and in equilibrium, respectively.

p_i The probability that station i is polled at a given polling instant.

$P_i(z)$ The generating function of $X_i(t)$.

$Q_i(z)$ The generating function of the number of customers found in queue i at arbitrary moments (in equilibrium).

r_i The expected length of the switchover period associated with station i.

$R_i(z)$ The generating function of the length of the switchover period associated with station i.

S_i, $S_i(z)$ The length of a service period (system in equilibrium) and its generating function, respectively.

T_i, $T_i(z)$ The response time of an arbitrary customer arriving to station i (system in equilibrium) and its generating function, respectively.

V_i The number of customers arriving together (in the same batch) with a tagged customer to queue i and which are served in front of the tagged customer.

$V_i(z)$ The generating function of V_i.

W_i, $W_i(z)$ The waiting time of an arbitrary customer arriving to station i (in equilibrium) and its generating function, respectively.

x_0 The probability that no customer arrives at queue i at time t (symmetric system).

$X_i(t)$ The number of arrivals to queue i at time t.

δ_i^2 The variance of the length of the switchover period associated with station i.

μ_i $\mathrm{E}[X_i(t)]$.

σ_i^2 $\mathrm{Var}[X_i(t)]$.

$\tau(m)$, $\tau(m)$, $\bar{\tau}(m)$ The instants at which the mth service period of the system starts, the mth service period of the system terminates, and the mth switchover period of the system terminates, respectively.

$\tau_i(m)$, $\tau_i(m)$, $\bar{\tau}_i(m)$ The instants at which the mth service period of queue i starts, the mth service period of queue i terminates, and the

mth switchover period of queue i terminates, respectively.

Acknowledgment

This research was supported in part by the Defense Advanced Research Projects Agency of the Department of Defense under contract MDA 903-82-C-0064. This paper was written while Hanoch Levy was at the University of California, Los Angeles, California. We would like to thank an anonymous referee and the associate editor for making us aware of the issue of polling systems with zero length switchover periods.

References

COOPER, R. B. 1970. Queues Served in Cyclic Order: Waiting Times. *Bell Syst. Tech. J.* **49,** 399–413.

COOPER, R. B., AND G. MURRAY. 1969. Queues Served in Cyclic Order, *Bell Syst. Tech. J.* **48,** 675–689.

DE MORAES, L. F. M. 1981. Message Queueing Delays in Polling Schemes with Applications to Data Communications Networks. UCLA-ENG-8106, Department of System Science, School of Engineering and Applied Science, University of California, Los Angeles, Calif. (May).

EISENBERG, M. 1972. Queues with Periodic Service and Changeover Time. *Opns. Res.* **20,** 440–451.

FERGUSON, M. J., AND Y. J. AMINETZAH. 1985. Exact Results for Nonsymmetric Token Ring Systems. *IEEE Trans. Commun.* **COM-33,** 223–331.

FUHRMANN, S. W. 1985. Symmetric Queues Served in Cyclic Order. *Opns. Res. Lett.* **4,** No. 3, October.

HASHIDA, O. 1972. Analysis of Multiqueue. *Rev. Elect. Commun. Lab.* **20,** 189–199.

HUMBLET, P. 1978. Source Coding for Communication Concentrators. Electronic Systems Laboratories, Massachusetts Institute of Technology, Cambridge, ESL-R-798 (January).

KONHEIM, A. G. 1980. Mathematical Models for Computer Data Communication. In *Case Studies in Mathematical Modeling*, pp. 256–334, W. E. Boyce (ed.). Pitman Advanced Publishing Program, Boston, Mass.

KONHEIM, A. G., AND B. MEISTER. 1974. Waiting Lines and Times in a System with Polling. *J. Assoc. Comput. Mach.* **21,** 470–490.

LEVY, H. 1984. Non-Uniform Structures and Synchronization Patterns in Shared-Channel Communication Networks. CSD-840049, Computer Science Department, University of California, Los Angeles, Ph.D. dissertation (August).

LEVY, H. 1986. Delay Computation and Dynamic Behavior of Non-Symmetric Polling Systems. AT&T Bell Laboratories, Holmdel, N.J., February 1986. Submitted for publication.

LEVY, H., AND L. KLEINROCK. 1987. Polling Systems with Zero Switch-Over Periods: A General Method for Analyzing the Expected Delay. AT&T Bell Laboratories, Holmdel, N.J. (June).

NOMURA, M., AND K. TSUKAMOTO. 1978. Traffic Analysis of Polling Systems (in Japanese). *Trans. Inst. Electron. Commun. Eng. Jap.* **J61-B(7),** 600–607.

RUBIN, I., AND L. F. M. DE MORAES. 1983. Message Delay Analysis for Polling and Token Multiple-Access Schemes for Local Communication Networks. *IEEE. J. Select. Areas Commun.* **SAC-1,** 935–947.

SWARTZ, G. B. 1980. Polling in a Loop System. *J. Assoc. Comput. Mach.* **27,** 42–59.

TAKAGI, H. 1985. Mean Message Waiting Times in Symmetric Multi-Queue Systems with Cyclic Service. *Perform. Eval.* **5,** 271–277.

TAKAGI, H. 1986. *Analysis of Polling Systems*, MIT Press, Cambridge, Mass.

TAKAGI, H. 1987. A Survey of Queueing Analysis of Polling Systems. In *Proceedings of Third International Conference on Data Communication Systems and Their Performance*, Rio De Janeiro (June).

TAKAGI, H., AND L. KLEINROCK. 1983. A Tutorial on the Analysis of Polling Systems, University of California at Los Angeles (December). (A preliminary draft of Takagi and Kleinrock 1985a, b, and Takagi 1986).

TAKAGI, H., AND L. KLEINROCK. 1985a. Analysis of Polling Systems. IBM Japan Science Institute, TR87-0002 (January).

TAKAGI, H., AND L. KLEINROCK. 1985b. Analysis of Polling Systems. University of California at Los Angeles, CSD-850005 (February).