Polling Systems with Simultaneous Arrivals

Hanoch Levy and Moshe Sidi

Abstract—**In this paper, we analyze polling systems with multiple types of simultaneous arrivals, namely, batches of customers may arrive at the different queues at an arrival epoch. We consider cyclic polling systems with N queues, general service time distribution in each queue and general switchover times. For both the exhaustive and the gated service disciplines we derive the necessary equations for computing the N expected waiting time figures. A "pseudo" conservation law for these systems is also derived. The analysis approach can be applied to other polling systems with multitype simultaneous arrivals.**

I. Introduction

QUEUEING systems with multiple types of simultaneous arrivals (correlated arrivals) have rarely been treated in the literature. The purpose of this paper is to introduce a complete analysis of the expected waiting times in cyclic polling systems with this type of arrivals.

The main reason for considering correlated arrivals is that, in practice, the arrival processes to the different queues of a queueing system are not necessarily independent. For instance, let us examine a communication system that can be modeled as a cyclic polling system. Consider a switching node with N input queues, N output channels and a single switch that connects queue i to channel i according to some polling policy. In such a system, when one message originated at the node has to be broadcast through a subset of the channels, its generation corresponds to an arrival of a copy of that message to each of the corresponding queues at the same time. Therefore, the arrivals are not independent in this system. Other examples in different systems can also be thought of.

Polling systems have been considered in numerous papers [1]–[28]. The only paper that considered a polling system in the context of correlated arrivals is that of Dou and Chang [6], where a *two-queue* discrete-time polling system with *zero switchover time* (walking time) and a *very restricted* structure of correlation between the arrivals to the two queues has been considered. The analysis in [6] is tailored to the specific correlation they assume and therefore it cannot be extended to general correlations or to systems having more than two queues.

In this paper, we consider continuous-time polling systems with N queues, general structure of correlation between the arrivals to the queues at an arrival epoch and arbitrary switchover times. In Section II, we present our model for correlated arrivals and in Section III, we use the buffer occupancy method [4], [5], [7], [11], [14], [22], [24] to derive the N expected waiting time figures for the gated and the exhaustive disciplines. In addition, we present a "pseudo" conservation law [2] for our system. In Section IV, we compare several special cases of the correlated arrivals polling system, discuss the computational aspects of the numerical method and examine the applicability of our analysis to other polling systems.

Paper approved by the Editor for Queueing and Network Performance of the IEEE Communications Society. Manuscript received July 11, 1988; revised February 20, 1990. The work of M. Sidi was supported by the Technion V. P. R. Fund—Bernstein Fund for Promotion of Research. This paper was presented at INFOCOM '89, Ottawa, Ont., Canada, 1989.

H. Levy is with the Computer Science Department, Tel-Aviv University, Tel-Aviv 69978, Israel.

M. Sidi is with the Electrical Engineering Department, Technion—Israel Institute of Technology, Haifa 32000, Israel.

IEEE Log Number 9144901.

II. Model Description, Main Assumptions, and Notations

The systems considered here consist of N infinite buffer queues and a single server with switchover periods (assumed not to be deterministically of zero length). The queues are indexed $1, 2, \cdots, N$ and for simplicity of notation, all references to queue index are implicitly assumed to be modulo N.

The service order is cyclic. After completing servicing queue i the server incurs a switchover period (whose length is associated with queue i) and moves to serve queue $i+1$. Two *service policies* are considered: i) *Exhaustive service,* in which when the server polls queue i it will serve this queue until its buffer empties. ii) *Gated service,* in which when the server polls queue i it will serve all the customers present in queue i at the polling instant. The period at which the server serves queue i is called a *service period* of queue i.

We model correlated arrivals as follows. Along the time axis, there are *arrival points.* The distribution of the arrival points is Poisson with parameter λ. At each arrival point, batches of customers arrive to the different queues according to some probability distribution. Specifically, let $K = (K_1, K_2, \cdots, K_N)$ be a random vector in which component K_i represents the number of customers arriving to queue i at an arrival point. The vector K is assumed to have the same joint distribution at each arrival point and this distribution is independent of previous or future arrival points. The joint probability distribution of the vector K, $\{\text{Prob}[K_1 = i_1, \cdots, K_N = i_N], i_j \geq 0, 0 \leq j \leq N\}$, is arbitrary.[1] Thus, we have a rather general structure of correlation between the simultaneous arrivals to the different queues at an arrival epoch. The special case of independent batch arrivals is represented by a distribution in which the only nonzero probabilities are of the form $\text{Prob}[K_1 = 0, \cdots, K_l = i_l, \cdots, K_N = 0]$ for some $l (i_l \geq 1)$. The case of single independent arrivals is represented by a distribution in which the only nonzero probabilities are of the form $\text{Prob}[K_1 = 0, \cdots, K_l = 1, \cdots, K_N = 0]$ for some l. Notice that with correlated arrivals the arrival rate of customers

[1] With the constraints $K_i \geq 0$ for every i, and K is not the zero vector.

to queue i is $\lambda_i = \lambda E[K_i]$. We let $K_{i,i} = E[K_i^2] - E[K_i]$ and for $j \neq i$ $K_{i,j} = E[K_i K_j]$. In addition, we let $z = (z_1, z_2, \cdots, z_N)$ and $K(z) = E[z_1^{K_1} z_2^{K_2} \cdots z_N^{K_N}]$.

The service time required by the customers of queue i (*type-i customers*) is a random variable B_i (*type-i service*) with a general distribution, having Laplace transform $B_i^*(s)$, mean b_i and second moment $b_i^{(2)}$. The offered load to queue i is defined as $\rho_i \triangleq \lambda_i b_i$ and the total system utilization is $\rho = \sum_{i=1}^N \rho_i$. The switchover period following the service of queue i is an independent random variable R_i, having a general distribution with Laplace transform $R_i^*(s)$, mean r_i and second moment $r_i^{(2)}$. We denote $r = \sum_{i=1}^N r_i$.

In the sequel, we use the buffer occupancy method to obtain the generating function, first and second moments of queue size distributions at polling instants. To that end, we let X_i^j denote the number of customers residing in queue j when queue i is polled. We define

$$F_i(z_1, z_2, \cdots, z_N) \triangleq E\left[z_1^{X_i^1} z_2^{X_i^2} \cdots z_N^{X_i^N}\right] \tag{2.1}$$

$$f_i(j) \triangleq E\left[X_i^j\right]; \qquad f_i(j,k) \triangleq \begin{cases} E[X_i^j X_i^k] & j \neq k \\ E\left[X_i^j\left(X_i^j - 1\right)\right] & j = k \end{cases}. \tag{2.2}$$

The number of customers arriving to queue j during the switchover period R_i is denoted by R_i^j. The number of customers arriving to queue j during a service period of queue i is denoted by A_i^j. The basic relations that hold in the polling system are

$$X_{i+1}^j = \begin{cases} X_i^j + A_i^j + R_i^j & i \neq j \\ A_i^i + R_i^i & i = j \end{cases} \quad \text{GATED} \tag{2.3}$$

$$X_{i+1}^j = \begin{cases} X_i^j + A_i^j + R_i^j & i \neq j \\ R_i^i & i = j \end{cases} \quad \text{EXHAUSTIVE.} \tag{2.4}$$

III. Analysis

A. The Gated Discipline

Since $R_i^*(\lambda - \lambda K(z))$ and $B_i^*(\lambda - \lambda K(z))$ are the joint generating functions of the number of customers arriving to the queues during the switchover period R_i and the service time B_i, respectively, we obtain from (2.3) [after some algebra]

$$F_{i+1}(z) = R_i^*(\lambda - \lambda K(z)) \cdot F_i(z_1, z_2, \cdots, z_{i-1}, B_i^*(\lambda - \lambda K(z)), z_{i+1}, \cdots, z_N). \tag{3.1}$$

By taking the respective derivatives of (3.1) at $z_1 = z_2 = \cdots = z_N = 1$ we can obtain sets of equations necessary to compute $f_i(j)$ and $f_i(j,k)$. In fact, it is easy to see that first moments are identical to that of the regular gated system (without correlations). Thus (see, e.g., [27])

$$f_i(i) = \frac{\lambda_i r}{1-\rho};$$
$$f_i(j) = \lambda_j \left[\sum_{k=j}^{i-1} r_k + r \sum_{k=j}^{i-1} \rho_k/(1-\rho)\right] \quad j \neq i. \tag{3.2}$$

For the second moments, we obtain

$$\begin{aligned} f_{i+1}(j,k) = {} & \lambda_j \lambda_k r_i^{(2)} + \lambda_k r_i f_i(j) + \lambda_j r_i f_i(k) \\ & + \lambda_j \lambda_k \left[2 r_i b_i + b_i^{(2)}\right] f_i(i) + f_i(j,k) \\ & + \lambda_k b_i f_i(i,j) + \lambda_j b_i f_i(i,k) \\ & + \lambda_j \lambda_k b_i^2 f_i(i,i) + \lambda[r_i + b_i f_i(i)] K_{j,k} \\ & i \neq j, i \neq k \end{aligned} \tag{3.3a}$$

$$\begin{aligned} f_{i+1}(i,k) = {} & \lambda_i \lambda_k r_i^{(2)} + \lambda_i r_i f_i(k) \\ & + \lambda_i \lambda_k \left[2 r_i b_i + b_i^{(2)}\right] f_i(i) \\ & + \lambda_i b_i f_i(i,k) + \lambda_i \lambda_k b_i^2 f_i(i,i) \\ & + \lambda[r_i + b_i f_i(i)] K_{i,k}, \qquad i \neq k \end{aligned} \tag{3.3b}$$

$$\begin{aligned} f_{i+1}(i,i) = {} & \lambda_i^2 r_i^{(2)} + \lambda_i^2 [2 r_i b_i + b_i^{(2)}] f_i(i) \\ & + \lambda_i^2 b_i^2 f_i(i,i) + \lambda[r_i + b_i f_i(i)] K_{i,i}. \end{aligned} \tag{3.3c}$$

Equations (3.3a–c) form a set of N^3 linear equations which can be solved numerically to yield the values of $f_i(i,i), (i = 1, \cdots, N)$ that are required for the derivation of expected waiting times. Note that an alternative way to derive (3.3a–c) is to use a direct computation by substituting (2.3) in (2.2) [20].

Expected Waiting Times: In the analysis of the regular gated system (single independent arrivals, see, e.g., [27]) the mean waiting time of a customer is given by

$$E[W_i] = \frac{h_i(i,i)}{h_i(i)} \cdot \frac{1 + \lambda_i b_i}{2\lambda_i} \tag{3.4}$$

where $h_i(i)$ and $h_i(i,i)$ play the role of $f_i(i)$ and $f_i(i,i)$ in that system.

Using (3.4) we next derive the waiting time in our system. Before doing so, some more notation is required. The customers arriving in an arrival point can be viewed as a collection of customer batches of different types. The (nonempty) group of type-i customers arriving in an arrival point is called a *type-i batch*. Let $\tilde{K}_i$ be the number of type-i customers in a type-i batch. Let p_i be the probability that any type-i customers arrive in an arrival point. Then we have

$$E[K_i] = p_i \cdot E\left[\tilde{K}_i\right]; \qquad E[K_i^2] = p_i \cdot E\left[\tilde{K}_i^2\right]. \tag{3.5}$$

To compute the waiting time of type-i customers in our system, which we denote by *System A*, we concentrate on the marginal behavior of queue i and consider *System B* which simulates System A as follows: A type-i customer in System B represents a type-i batch in System A. Thus, the waiting time of the first customer in a batch, in System A, is equal to the waiting time of an arbitrary customer in System B. We therefore have to solve for the waiting time in System B using (3.4). To do this we need to derive the parameters of System B, and substitute them into (3.4).

Let Z_i be the number of type-i batches residing in queue i when it is polled. Let $g_i(i) = E[Z_i]$ and $g_i(i,i) = E[Z_i^2 - Z_i]$. We have $E[X_i^1] = E[Z_i]E[\tilde{K}_i]$ and $E[X_i^2] = E[Z_i]E[\tilde{K}_i^2] + (E[Z_i^2] - E[Z_i])(E[\tilde{K}_i])^2$. Thus, $E[Z_i^2] = \{E[X_i^2] - E[Z_i](E[\tilde{K}_i] - (E[\tilde{K}_i])^2)\}/(E[\tilde{K}_i])^2$ and

$$\frac{g_i(i,i)}{g_i(i)} = \frac{1}{E[\tilde{K}_i]}\left[\frac{E[X_i^2]}{E[X_i^1]} - \frac{E[\tilde{K}_i^2]}{E[\tilde{K}_i]}\right]. \tag{3.6}$$

The arrival rate of type-i customers to System B is obviously $p_i\lambda$. The service time of a customer in that system consists of the sum of $\tilde{K}_i$ type-i service times in System A. Thus, the expected value of that service time is given by $E[\tilde{K}_i]b_i$. In addition, the terms $g_i(i)$ and $g_i(i,i)$ will substitute $h_i(i)$ and $h_i(i,i)$, respectively, in (3.4) Putting all this together, we derive the expected waiting time in System B, which is the expected waiting time of the first customer in a batch in System A [we use (3.5) and (3.6)]:

$$\begin{aligned} E[W_i^{\text{batch}}] &= \frac{g_i(i,i)}{g_i(i)} \cdot \frac{1 + \lambda p_i E[\tilde{K}_i] b_i}{2\lambda p_i} \\ &= \left[\frac{f_i(i,i)}{f_i(i)} - \frac{K_{i,i}}{E[K_i]}\right]\frac{1+\rho_i}{2\lambda_i}. \end{aligned} \tag{3.7}$$

Let W_i^b be the waiting time of a tagged customer within a batch of type i. By standard techniques we have $E[W_i^b] = b_i K_{i,i}/(2E[K_i])$. Thus, the total expected waiting time of a customer in a gated system with correlated arrivals is

$$\begin{aligned} E[W_i] &= E[W_i^b] + E[W_i^{\text{batch}}] \\ &= \frac{f_i(i,i)}{f_i(i)}\left[\frac{1+\rho_i}{2\lambda_i}\right] - \frac{K_{i,i}}{2\lambda_i E[K_i]}. \end{aligned} \tag{3.8}$$

B. The Exhaustive Discipline

The key in the analysis of the exhaustive discipline is the analysis of the queue size during a busy period. We concentrate on a service period in which queue i is served. For simplicity, assume that during the service period customers are served according to the last come–first served (nonpreemptive) policy. This assumption does not affect the results since the selection of customers for service is independent of their service time.

Let c_k be an arbitrary customer; we recursively define the *subbusy period initiated by* c_k as the period at which c_k and all the customers arriving during the subbusy period are served. Let P_i^j be a random variable denoting the number of customers arriving to queue j during one subbusy period of queue i. We define $P(z_1, z_2, \cdots, z_{i-1}, z_{i+1}, \cdots, z_N) \triangleq E[z_1^{P_i^1} z_2^{P_i^2} \cdots z_{i-1}^{P_i^{i-1}} z_{i+1}^{P_i^{i+1}} \cdots z_N^{P_i^N}]$. We have

$$\begin{aligned} &P(z_1, z_2, \cdots, z_{i-1}, z_{i+1}, \cdots, z_N) \\ &\quad = B_i^*(\lambda - \lambda K(z_1, z_2, \cdots, z_{i-1}, \\ &\qquad \cdot P(z_1, z_2, \cdots, z_{i-1}, z_{i+1}, \cdots, z_N), z_{i+1}, \cdots, z_N)). \end{aligned}$$

From (2.4) we have

$$\begin{aligned} F_{i+1}(z) &= R_i^*(\lambda - \lambda K(z))F_i(z_1, z_2, \cdots, z_{i-1}, \\ &\quad P(z_1, z_2, \cdots, z_{i-1}, z_{i+1}, \cdots, z_N), z_{i+1}, \cdots, z_N). \end{aligned} \tag{4.1}$$

By taking the respective derivatives of (4.1) at $z_1 = z_2 = \cdots = z_N = 1$ we can obtain sets of equations necessary to compute $f_i(j)$ and $f_i(j,k)$. In fact, it is easy to see that in this case also the first moments are identical to that of the regular exhaustive system (without correlations). Thus (see e.g., [27])

$$\begin{aligned} f_i(i) &= \frac{\lambda_i(1-\rho_i)r}{1-\rho}; \\ f_i(j) &= \lambda_j\left[\sum_{k=j}^{i-1} r_k + r\sum_{k=j+1}^{i-1}\rho_k/(1-\rho)\right] \quad j \neq i. \end{aligned} \tag{4.2}$$

For the second moments, we obtain

$$\begin{aligned} f_{i+1}(j,k) &= \lambda_j\lambda_k r_i^{(2)} + r_i\lambda_k f_i(j) \\ &\quad + r_i\lambda_j f_i(k) + \lambda_j\lambda_k\left[\frac{2r_i b_i}{1-\rho_i} + \frac{b_i^{(2)}}{(1-\rho_i)^3}\right]f_i(i) \\ &+ f_i(j,k) + \frac{b_i}{1-\rho_i}[\lambda_k f_i(i,j) + \lambda_j f_i(i,k)] + \frac{\lambda_j\lambda_k b_i^2 f_i(i,i)}{(1-\rho_i)^2} \\ &+ \frac{f_i(i)\lambda b_i}{1-\rho_i}\left[K_{j,k} + \frac{b_i}{1-\rho_i}[\lambda_k K_{i,j} + \lambda_j K_{i,k}] + \frac{\lambda_j\lambda_k b_i^2}{(1-\rho_i)^2}K_{i,i}\right] \\ &\quad + \lambda r_i K_{j,k} \qquad i \neq j,\ i \neq k,\ j \neq k \end{aligned} \tag{4.3a}$$

$$\begin{aligned} f_{i+1}(i,k) &= \lambda_i\lambda_k r_i^{(2)} \\ &\quad + r_i\lambda_i\left[f_i(k) + \frac{\lambda_k b_i f_i(i)}{1-\rho_i}\right] + \lambda r_i K_{i,k} \qquad i \neq k \end{aligned} \tag{4.3b}$$

$$f_{i+1}(i,i) = \lambda_i^2 r_i^{(2)} + \lambda r_i K_{i,i}. \tag{4.3c}$$

Equations (4.3a–c) form a set of N^3 linear equations which can be solved by numerical techniques to yield the values of $f_i(i,i), (i = 1, \cdots, N)$ which are required for the derivation of expected waiting times. A direct derivation of (4.3a–c) appears in [20].

Expected Waiting Times: In the analysis of the regular exhaustive system (single independent arrivals, see, e.g., [27]) the mean waiting time of a customer is given by

$$E[W_i] = \frac{h_i(i,i)}{2\lambda_i h_i(i)} + \frac{\lambda_i b_i^{(2)}}{2(1-\rho_i)} \tag{4.4}$$

where $h_i(i)$ and $h_i(i,i)$ play the role of $f_i(i)$ and $f_i(i,i)$ in that system.

Similarly to the gated system we may compute the mean waiting time of the first customer in a batch by considering a system in which each batch is viewed as a customer. In that system, we have to substitute $g_i(i), g_i(i,i), \lambda p_i, b_i^{(2)} E\tilde{K}_i + b_i^2 E[\tilde{K}_i^2 - \tilde{K}_i]$ and ρ_i for $h_i(i), h_i(i,i), \lambda_i, b_i^{(2)}$ and ρ_i in (4.4). This will yield

$$E\left[W_i^{\text{batch}}\right] = \frac{g_i(i,i)}{2\lambda p_i g_i(i)} + \frac{\lambda p_i \left(b_i^{(2)} E\left[\tilde{K}_i\right] + b_i^2 E\left[\tilde{K}_i^2 - \tilde{K}_i\right]\right)}{2(1-\rho_i)}.$$

From the analysis of the gated system (3.5) and (3.6) we have the values of $g_i(i,i)/g_i(i), E[\tilde{K}_i^2]$, and $E[\tilde{K}_i^2]$, and we may derive the mean waiting time of an arbitrary customer at queue i:

$$E[W_i] = E\left[W_i^{\text{batch}}\right] + E\left[W_i^b\right] = \frac{f_i(i,i)}{2\lambda_i f_i(i)} + \frac{\lambda_i b_i^{(2)}}{2(1-\rho_i)} + \frac{(2\rho_i - 1)K_{i,i}}{2\lambda_i(1-\rho_i)E[K_i]}.$$

C. A "Conservation Law"

In [2] a "pseudo" conservation law has been derived for polling systems with uncorrelated arrivals. This law can be generalized to systems with correlated arrivals by adding the term $\frac{1}{2}\lambda \sum_{i=1}^N \sum_{j=1}^N b_i b_j K_{i,j}/(1-\rho)$ to the right-hand side of the law presented in [2]. For details see [20].

IV. Discussion

A. Comparison of Different Systems

Consider a system in which $K_{i,i} = 0$ (for every i) and $K_{i,j}$ is either 0 or 1 (for every $i \neq j$), i.e., each queue gets at most one customer at any arrival point, but the arrivals to the different queues are correlated. It is seen from (3.8) and (4.5) that the expected waiting time in this system is greater than or equal to that of the regular system (no correlations). This dominance is attributed to the correlation of arrivals. Note, that though this result sounds "intuitive" it may not be simple to prove by other means.

Next, consider a system in which $K_{i,i} > 0$ (for every i) and $K_{i,j} = 0$ (for every $i \neq j$). This corresponds to independent batch arrivals. The comparison of this system to the regular system is not simple, and the question of whether batching always hurts performance is, therefore, left open in the context of this work. Nevertheless, based on numerous numerical examples, we conjecture that batching *does always hurt* performance.

Last, consider an example of an exhaustive system consisting of two queues with $b_1 = 0.7, b_2 = 0.3, b_1^{(2)} = 0.98, b_2^{(2)} = 0.18$. Switchover periods are identical: $r_1 = r_2 = r_1^{(2)} = r_2^{(2)} = 1$. The total arrival rate λ is a varying parameter and $E[K_1] = E[K_2]$ implying $\lambda_1 = \lambda_2$. The mean waiting time in both queues is depicted in Fig. 1 versus the total load ($\rho = \sum_{i=1}^2 \lambda_i b_i$). Continuous curves represent queue 1 and dotted curves represent queue 2. We consider three cases. The first is a regular (single independent arrivals) system (in which $E[K_1] = E[K_2] = 0.5$ and the correlation matrix is identically 0), whose curves are marked by the letter *I*. The second is a system with single correlated arrivals

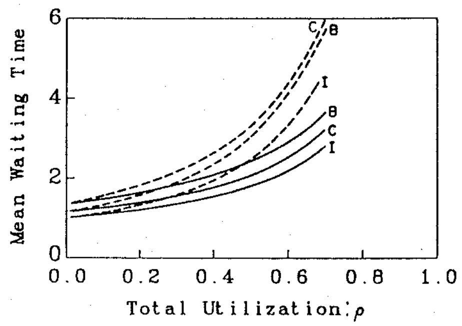

Fig. 1. Two stations.

(in which $E[K_1] = E[K_2] = 1, K_{1,2} = K_{2,1} = 1$, and $K_{1,1} = K_{2,2} = 0$) whose curves are marked by the letter *C*. The third is a system with independent batch arrivals (in which $E[K_1] = E[K_2] = 1, K_{1,2} = K_{2,1} = 0$, and $K_{1,1} = K_{2,2} = 1$), whose curves are marked by the letter *B*.

As seen in the figure, the expected waiting time in the independent single arrival system (regular system) is lower than that of both the independent batch arrival system and the correlated single arrival system. This is true for both queues. Another interesting feature is that in the independent batch arrival system, the "heavy" station (queue 1) enjoys lower waiting times than queue 2 at high loads, while for low loads the situation is reversed. The reason is that for low loads, the main factor affecting performance is batching, while for high loads, as is well known for exhaustive systems, heavy loaded stations receive better service.

B. Computational Aspects

The main computational issue is that of numerically solving (3.3a–c) and (4.3a–c). These equation sets can be efficiently solved using the successive approximation method (see, e.g., [15]). As shown in [16] for systems with single arrivals the successive approximation methods will converge and it will require $O(N^3 \log_\alpha \varepsilon)$ steps to converge to within relative accuracy ε. The logarithm base, in that expression, is $\alpha = (\rho - \rho_{\min})^2/(1-\rho_{\min})^2$ for the exhaustive system (where $\rho_{\min} = \min_i \rho_i$) and $\alpha = \rho^2$ for the gated system. The same holds for our system with correlated arrivals.

It is an open question whether the *station-time* method [1], [8], [12], [23] can be applied to polling systems with correlated arrivals. We conjecture (see [20]) that it cannot be applied since the *N* recent station times do not form a Markov chain.

C. Applicability of the Analysis to Other Polling Systems

In this paper, we concentrated on analyzing the "most traditional" polling systems, namely, the continuous time cyclic polling systems with either gated service or exhaustive service. The analysis, nevertheless, applies (after simple modifications) to most polling systems which have previously been analyzed under the assumption of independent single arrivals. These include mixed service, periodic polling (polling table) [1], random polling [13], discrete-time systems, zero switchover periods [12], [17], limited-1 service (here the analysis is limited to the conservation law and symmetric systems) [10], [21], [25], [26], binomial gated [18], and exhaustive [19], etc.

References

[1] J. E. Baker and I. Rubin, "Polling with a general-service order table," *IEEE Trans. Commun.*, vol. COM-35, pp. 283–288, Mar. 1987.

[2] O. J. Boxma and W. P. Groenendijk, "Pseudo-conservation laws in cyclic queues," *J. Appl. Prob.*, vol. 24, pp. 949–964, 1987.

[3] R. T. Carsten, E. E. Newhall, and M. J. M. Posner, "A simplified analysis of scan times in an asymmetrical newhall loop with exhaustive service," *IEEE Trans. Commun.*, vol. COM-25, pp. 951–957, Sept. 1977.

[4] R. B. Cooper, "Queues served in cyclic order: Waiting times," *Bell Syst. Tech. J.*, vol. 49, pp. 399–413, Mar. 1970.

[5] R. B. Cooper and G. Murray, "Queues served in cyclic order," *Bell Syst. Tech. J.*, vol. 48, pp. 675–689, Mar. 1969.

[6] C. Dou and J. F. Chang, "Polling of two queues with synchronized correlated inputs and zero walktime," in *Proc. IEEE INFOCOM'87*, San Francisco, CA, Mar. 1987, pp. 952–961.

[7] M. Eisenberg, "Queues with periodic service and changeover time," *Oper. Res.*, vol. 20, pp. 440–451, 1972.

[8] M. J. Ferguson and Y. J. Aminetzah, "Exact results for nonsymmetric token ring systems," *IEEE Trans. Commun.*, vol. COM-33, pp. 223–231, Mar. 1985.

[9] M. J. Ferguson, "Mean waiting time for a token ring with station dependent overheads," in *Local Area & Multiple Access Networks*, R. L. Pickholtz, Ed. Rockville, MD: Computer Science, ch. 3, pp. 43–67.

[10] S. W. Fuhrmann, "Symmetric queues served in cyclic order," *Oper. Res. Lett.*, vol. 4, no. 3, pp. 139–144, Oct. 1985.

[11] O. Hashida, "Analysis of multiqueue," *Rev. Elec. Commun., Lab.*, vol. 20, 3, 4, pp. 189–199, 1972.

[12] P. Humblet, "Source coding for communication concentrators," *Electron. Syst. Lab.*, Massachusetts Inst. Technol., Cambridge, ESL-R-798, Jan. 1978.

[13] L. Kleinrock and H. Levy, "The analysis of random polling systems," *Oper. Res.*, vol. 36, no. 5, pp. 716–732, Sept.–Oct. 1988.

[14] A. G. Konheim and B. Meister, "Waiting times and lines in system with polling," *J. Ass. Comput. Mach.*, vol. 21, pp. 470–490, 1974.

[15] H. Kushner, *Introduction to Stochastic Control.* New York: Holt, Rinehart and Winston, 1971.

[16] H. Levy, "Delay computation and dynamic behavior of non-symmetric polling systems," *Perform. Eval.*, vol. 10, no. 1, pp. 35–51, Oct. 1989.

[17] H. Levy and L. Kleinrock, "Polling systems with zero switch-over periods: A general method for analyzing the expected delay," *Perform. Eval.*, June 1987.

[18] H. Levy, "Analysis of cyclic-polling systems with binomial-gated service," in *Performance of Distributed and Parallel Systems*, T. Hasegawa, H. Takagi and Y. Takahashi, Ed. Amserterdam, The Netherlands: Elsvier Science, 1989, pp. 127–139.

[19] ——— "Optimization of polling systems: The fractional-exhaustive service method," Tech. Rep. Dep. Comput. Sci., Tel-Aviv Univ., Nov. 1988.

[20] H. Levy and M. Sidi, "Correlated arrivals in polling systems," in *Proc. INFOCOM'89*, Ottawa, Ont., Canada, April 1989, pp. 907–913, also an EE Publication, no. 676, Technion, June 1988.

[21] M. Nomura and K. Tsukamoto, "Traffic analysis on polling systems," *Trans. Inst. Electron. Commun. Eng. Japan*, vol. J61-B, no. 7, pp. 600–607, July 1978, (in Japaneese).

[22] I. Rubin and L. F. Demoraes, "Message delay analysis for polling and token multiple-access schemes for local communication networks," *IEEE Select. Areas Commun.*, vol. SAC-1, Nov. 1983.

[23] D. Sarkar and W. I. Zangwill, "Expected waiting time for nonsymmetric cyclic queueing systems—Exact results and applications," *Manage. Sci.*, vol. 35, no. 12, pp. 1463–1474, Dec. 1989.

[24] G. B. Swartz, "Polling in a loop system," *J. Assoc. Comp. Mach.*, vol. 27, pp. 42–59, 1980.

[25] H. Takagi, "Mean message waiting times in symmetric polling system," in *Performance' 84*, E. Gelenbe, Ed. New York: Elsevier Science, pp. 293–302.

[26] ———, "Mean message waiting times in symmetric multi-queue systems with cyclic service," *Perform. Eval.*, vol. 5, no. 4, pp. 271–277, 1985.

[27] ———, *Analysis of Polling Systems.* Cambridge, MA: M. I. T. Press, Apr. 1986.

[28] K. S. Watson, "Performance evaluation of cyclic service strategies — A survey," in *Performance '84*, E. Gelenbe, Ed. Amsterdam, The Netherlands: North-Holland, pp. 521–533.