

DISCRETE-TIME ANALYSIS OF POLLING SYSTEMS WITH RENEWAL INPUTS*

Phuoc TRAN-GIA

IBM Research Division, Zurich Research Laboratory, 8803 Rüschlikon,
Switzerland

The class of polling systems, i.e., multiqueue systems with cyclic service, plays an important role in the performance evaluation of various computer and communication systems, e.g., switching systems with distributed control and token-passing local-area networks. Although the behavior of users and the corresponding incoming traffic characteristics in such systems are increasingly complex, most analytical performance studies in the literature are based on the assumption of Poissonian input processes. In this paper, an approximate discrete-time analysis of polling systems with finite capacity of waiting places and nonexhaustive service is presented, considering general renewal input traffic. The analysis method is based on the evaluation of discrete convolution operations using fast convolution algorithms, e.g., the Fast Fourier Transform (FFT). To illustrate the accuracy of the approximation and its dependency on system parameters, numerical results are given.

1. INTRODUCTION

In modeling and performance evaluations of a broad spectrum of computer and communication systems, e.g., investigations of communication structures in switching systems with distributed control and token-passing local-area networks, etc. the class of polling models, i.e., multiqueue systems with cyclic service is often employed. Owing to the architecture of such systems and the respective new services, the traffic mix of subscribers and users on communication networks and user behavior are of high complexity. There is an increasing tendency for the need of more complex processes to describe traffic streams in such environments.

To include more realistic modeling elements in the class of polling systems, two main objectives are taken into account in this paper: i) the consideration of general renewal processes as inputs, and ii) the modeling component in respect of finite waiting capacity of devices or stations in polling systems. From the analytic viewpoint, this study takes advantage of discrete-time analysis methods, where effective convolution and transformation algorithms [e.g., the Fast Fourier Transform (FFT)] can be employed.

Polling systems have been the subject of numerous studies in the literature [2-8, 12-14, 16-19, 21, 22]. A survey on the analysis of polling systems can be found in Takagi and Kleinrock [21]. Various polling mechanisms, like cyclic or priority order (cf. [16]), and several service disciplines, e.g., exhaustive, nonexhaustive or

*The major part of this work was performed while the author was with the Institute of Communications Switching and Data Technics, University of Stuttgart, FRG.

gating, have been considered. Some of these studies take into account the switchover time, i.e., the time interval spent by the server in switching over from one queue to the succeeding one. In most of these studies, input processes are assumed to be Poisson, and the queues of the polling system to have infinite capacity.

An approximation of polling systems under symmetrical load conditions, constant switchover time and gating service is given by Leibowitz [14]. Cooper and Murray have considered a system with gating or exhaustive service and zero switchover time [5]. Their approach has been generalized by Eisenberg [6] and Hashida [7] to nonzero switchover time. The case of two queues with nonzero switchover time was taken into account by Boxma [2] where an exact solution has been derived. An approximate analysis technique for polling systems with nonexhaustive service and general switchover time has been developed by Kuehn [12, 13]. Approximate formulas for some general classes of polling systems can be found in Boxma and Meister [3] and Bux and Truong [4]. Morris and Wang [17] and Raith [18] have provided analytical approaches to deal with polling systems with multiple servers, while Raith and Tran-Gia [19] considered the influence of the backward-locking phenomenon caused by the receiving part of a polling system in a more general context. Polling systems with finite capacity of waiting places and Poisson input traffic have been analyzed approximately in Tran-Gia and Raith [22].

One of the main features of this study is the use of analysis methods operating in the discrete-time domain. The discrete-time approach is justified by the fact that the parameterization of model components is often based on data measured in terms of histograms. The discrete-time nature of model components can be registered in a number of modeling processes, e.g., in performance investigations of packet switching systems, time-slotted systems, etc.

There is a number of studies [1, 10, 11, 23, 24] which deal with the analysis of discrete-time models. In Kobayashi [10] a survey has been presented. Examples for the use of discrete-time analysis methods can be seen in the case of G/G/1 queues, where numerical solutions are given [1, 11, 23] based on the discrete-time form of the Lindley's relation [15, 20]. An interesting algorithm for calculation of the waiting-time distribution function of the GI/G/1 queue has been presented by Ackroyd [1], where methods used in signal processing theory (in both time and frequency domains) and fast convolution algorithms are employed. In [23], calculation algorithms can be found for the idle time and interdeparture distributions of the class of GI/G/1 queues with discrete-time arrival and service processes. A solution for this class of systems with general cyclic input processes was given in Tran-Gia and Rathgeb [24], considering models of semidynamic scheduling and routing mechanisms.

Methods of discrete-time analysis will be applied in this paper to obtain an approximation algorithm for the class of polling systems with renewal inputs. In Section 2, the model and its parameters will be described, while Sections 3 and 4 give an outline of the analysis and calculation of performance measures, respectively. Numerical results are shown in Section 5 to illustrate the accuracy of the calculation method and its dependency on system parameters.

2. MODEL DESCRIPTION

The basic structure and related parameters of the polling model considered here are illustrated in Figure 1. The model consists of g finite capacity queues, served nonexhaustively (or more precisely, limited service of one message per service) in a cyclic order by a single server with a generally distributed service time. After the

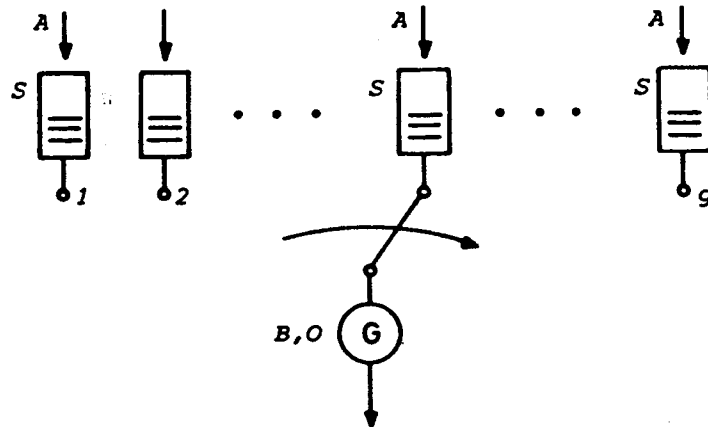


FIGURE 1
Polling system with general inputs and finite capacity.

service of a queue, the server will move to the succeeding queue. This switchover time, which models all overheads spent and procedures performed by the server to move to and scan the succeeding queue, is assumed to have a general distribution function. At the scanning epoch, i.e., at the end of the corresponding switchover time, the server will process one message in the queue, if there is at least one message waiting for service. If the queue is empty, the interscan period observed will consist of just the switchover time.

As previously mentioned, one of the main contributions of this study is the consideration of general input processes and their influence on the behavior of the polling system. Thus, the arrival processes are assumed to be general.

In principle, the analysis method presented in this paper can be applied to nonsymmetrical polling systems with queue-individual interarrival, service and switchover-time distribution functions. Also, individual sizes of the queue capacities can be chosen. However, to simplify the description of the analysis algorithm, and to focus our attention on the main objectives of the study, we shall restrict ourselves in the following to the case of symmetrical systems.

3. APPROXIMATE ANALYSIS IN THE DISCRETE-TIME DOMAIN

3.1. Discrete-Time Random Variables and Notation

In the context of this analysis, we consider the random variables to be of discrete-time nature, i.e., the time axis is conceived to be divided into intervals of unit length Δt . As a consequence, samples of these random variables are integer multiples of Δt ; the time discretization is equidistant. In real systems, Δt is often given in discrete-time form, e.g., as transmission time of a bit, byte or packet. The corresponding distributions can be obtained by means of measurements, and arranged in the form of histograms.

The following notation is used for functions belonging to a discrete-time random variable (r.v.) X :

$x(k) = \Pr(X = k), -\infty < k < +\infty$	distribution of X
$X(k) = \sum_{i=-\infty}^k x(i), -\infty < k < +\infty$	distribution function of X
$x_{ZT}(z) = \sum_{k=-\infty}^{+\infty} x(k) z^{-k}$	Z-transform of $x(k)$
EX	mean of X
c_x	coefficient of variation of X.

As indicated in Figure 1, we use the following notation:

g	number of queues in the polling system
S	queue capacity
A	r.v. for the interarrival time of the input process at a queue, distribution $a(k)$. Since $a(0)$ can have a nonzero value, batch arrival processes with geometrically distributed batch size can also be dealt with (cf. [23]).
B	r.v. for the service time, distribution $b(k)$
O	r.v. for the switchover time, distribution $o(k)$.

A sample path of the state process development in a queue chosen arbitrarily, say j , of the polling system is shown in Figure 2. We observe the cycle time seen from queue j , i.e., the time interval between two consecutive scanning instances at queue j . Similar to the approach provided in [12], two types of conditional cycle times can be distinguished, denoted by the following random variables:

C_0	r.v. for the cycle time, conditioned on an <i>empty</i> queue at the previous scanning instance (i.e., without service of queue j during the cycle considered)
C_1	r.v. for the cycle time, conditioned on a <i>nonempty</i> queue at the previous scanning instance (i.e., with service of queue j).

During a conditional cycle time, a number of messages may arrive according to the arrival process. Depending on the type of conditional cycle, we denote the arrival group as follows:

G_j	r.v. for the number of messages arriving during a cycle of type j (C_j), $j=0,1$.
-------	--

Since the process development is observed in the discrete-time domain, several events (e.g., arrival or end of service phase of messages, scanning events) can occur simultaneously. In those cases, a convention is made for events to be processed in the following order: i) end of service; ii) scanning, and iii) arrival.

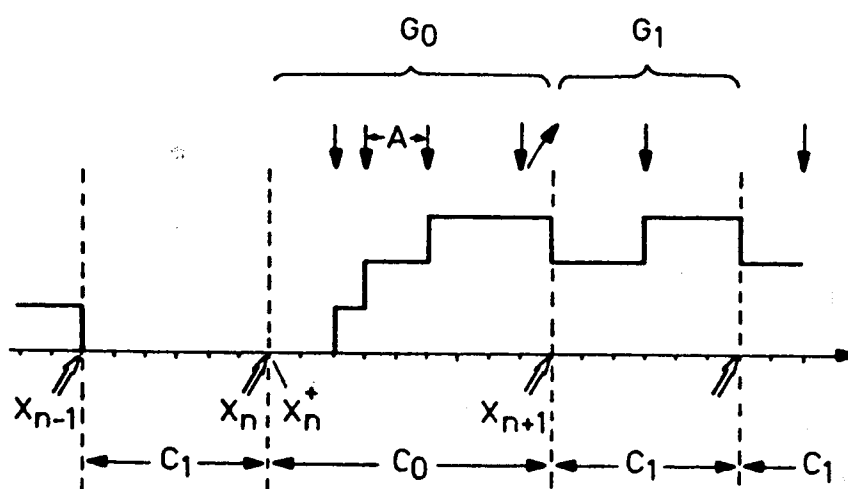


FIGURE 2
A sample path of the state process.

3.2. Markov-Chain State Probabilities

The state process of an observed queue is affected by two random processes: the arrival process of messages and the scanning process of the server, as depicted in Figure 2. Under the assumption that these two processes are not correlated, e.g., the scanning process sees the arrival process in the same way as an arbitrary outside observer, an approximation can be made to analyze the state process. We consider the regeneration points of a Markov chain embedded immediately prior to the scanning instants of a queue. The following random variables are used:

- X_n r.v. for the state of the queue observed (i.e., the number of messages in the queue) immediately *prior to* the n -th scanning epoch
- X_n^+ r.v. for the state of the queue observed immediately *after* the n -th scanning epoch.

Analogous to the definition of conditional cycle times, we define the following conditional r.v. for the state of the queue observed, depending on the previous value of X_n :

$$X_{n,0}^+ = X_n^+ | X_n = 0, \quad X_{n,1}^+ = X_n^+ | X_n > 0, \quad (3.1)$$

$$X_{n+1,0} = X_{n+1} | X_n = 0, \quad X_{n+1,1} = X_{n+1} | X_n > 0. \quad (3.2)$$

Thus, the relationships between these random variables and their distributions can be obtained:

i) $X_n = 0$

$$X_{n,0}^+ = 0 \quad (3.3)$$

$$x_{n,0}^+(k) = \delta(k) \quad (3.4)$$

$$X_{n+1,0} = \min(X_{n,0}^+ + G_0, S) = \min(G_0, S) \tag{3.5}$$

$$x_{n+1,0}(k) = \pi^S(x_{n,0}^+(k) \star g_0(k)) , \tag{3.6}$$

where

$$\delta(k - m) = \begin{cases} 1 & k = m \\ 0 & \text{otherwise} \end{cases} \tag{3.7}$$

with the operator π^m defined as

$$\pi^m(x(k)) = \begin{cases} x(k) & k < m \\ \sum_{i=m}^{\infty} x(i) & k = m \\ 0 & k > m \end{cases} \tag{3.8a}$$

and the \star -symbol denoting the discrete convolution operation:

$$a_3(k) = a_1(k) \star a_2(k) = \sum_{j=-\infty}^{+\infty} a_1(k - j) \cdot a_2(j) . \tag{3.8b}$$

ii) $X_n > 0$

$$X_{n,1}^+ = X_n - 1 \tag{3.9}$$

$$x_{n,1}^+(k) = \begin{cases} \frac{x_n(k+1)}{\Pr(X_n > 0)} = \frac{x_n(k+1)}{1 - x_n(0)} & k = 0, 1, \dots, S-1 \\ 0 & \text{otherwise} \end{cases} \tag{3.10}$$

$$X_{n+1,1} = \min(X_{n,1}^+ + G_1, S) \tag{3.11}$$

$$x_{n+1,1}(k) = \pi^S(x_{n,1}^+(k) \star g_1(k)) . \tag{3.12}$$

By eliminating the condition based on X_n , we arrive at

$$x_{n+1}(k) = x_n(0) \cdot x_{n+1,0}(k) + (1 - x_n(0)) \cdot x_{n+1,1}(k) , \quad k = 0, 1, \dots, S. \tag{3.13}$$

With Eqs. (3.4), (3.6), (3.10) and (3.12), a set of state equations describing the transition behavior of the Markov chain between two consecutive scanning epochs can be written as follows:

$$\begin{aligned} x_{n+1}(k) &= x_n(0) \cdot g_0(k) + \sum_{i=1}^{k+1} x_n(i) \cdot g_1(k - i + 1), \quad k = 0, \dots, S-1 \\ x_{n+1}(S) &= x_n(0) \sum_{i=S}^{\infty} g_0(i) + \sum_{i=1}^S x_n(i) \cdot \sum_{m=S-i+1}^{\infty} g_1(m) , \quad k = S. \end{aligned} \tag{3.14}$$

Under stationary conditions, i.e.,

$$X = \lim_{n \rightarrow \infty} X_n, \tag{3.15}$$

the equilibrium state equations of the Markov chain can be derived from Eq. (3.14):

$$\begin{aligned} x(k) &= x(0) \cdot g_0(k) + \sum_{i=1}^{k+1} x(i) \cdot g_1(k-i+1), \quad k = 0, \dots, S-1 \\ x(S) &= x(0) \sum_{i=S}^{\infty} g_0(i) + \sum_{i=1}^S x(i) \cdot \sum_{m=S-i+1}^{\infty} g_1(m), \quad k = S. \end{aligned} \tag{3.16}$$

To evaluate the equation system given in Eq. (3.16), it remains to calculate the group size distributions of $G_j, j=0,1$. Keeping in mind the assumption of no correlation between the scanning and arrival processes at a queue, we observe a cycle of type j and the arrival group G_j of messages arriving in this cycle. The time $F^{(k)}$ between the previous scanning epoch and the arrival epoch of the k -th message in this group is distributed according to

$$f^{(k)}(i) = a^*(i) \star \underbrace{a(i) \star \dots \star a(i)}_{(k-1)\text{-times}}, \tag{3.17a}$$

where $a^*(i)$ denotes the recurrence time distribution of the interarrival process in the discrete-time domain

$$a^*(i) = \frac{1}{EA} \left(1 - \sum_{n=0}^i a(n) \right), \quad i = 0, 1, \dots \tag{3.17b}$$

Assuming a cycle C_j of length m , the conditional arrival-group size distribution can be given as follows:

$$\begin{aligned} g_j(k|m) &= \text{Pr}(\text{group size is } k \mid \text{cycle } C_j \text{ is of length } m) \\ &= \text{Pr}(F^{(k)} < m \leq F^{(k+1)}) = \text{Pr}(F^{(k)} < m) - \text{Pr}(F^{(k+1)} < m) \end{aligned}$$

or

$$\begin{aligned} g_j(k|0) &= \delta(k) \\ g_j(k|m) &= \sum_{i=0}^{m-1} (f^{(k)}(i) - f^{(k+1)}(i)), \quad m = 1, 2, \dots \end{aligned} \tag{3.18}$$

Finally, the distribution of an arrival group G_j during a cycle C_j (distribution $c_j(m), m > 0$) can be obtained:

$$\begin{aligned} g_j(k) &= c_j(0) \cdot \delta(k) \\ &+ \sum_{m=1}^{\infty} c_j(m) \sum_{i=0}^{m-1} (f^{(k)}(i) - f^{(k+1)}(i)), \quad k = 0, 1, \dots, j = 0, 1. \end{aligned} \tag{3.19}$$

3.3. Cycle-Time Analysis

Define C_g to be the r.v. for the cycle-time segment (cf. [22]), i.e., the time interval between the scanning instants of two consecutive queues. In a symmetrical polling system under stationary conditions with the state probabilities obtained from Eq. (3.16), we obtain

$$c_S(k) = x(0) \cdot o(k) + (1 - x(0)) \cdot (o(k) \star b(k)) \quad (3.20)$$

or in Z-transform domain:

$$c_{S,ZT}(z) = o_{ZT}(z) \cdot (x(0) + (1 - x(0)) \cdot b_{ZT}(z)) \quad (3.21)$$

Under the assumption of independency between cycle-time segments, the Z-transforms of the conditional cycle times can be given as

$$c_{0,ZT}(z) = o_{ZT}(z) \cdot c_{S,ZT}^{g-1}(z) \quad (3.22)$$

$$c_{1,ZT}(z) = o_{ZT}(z) \cdot b_{ZT}(z) \cdot c_{S,ZT}^{g-1}(z) \quad (3.23)$$

Finally, we obtain the Z-transform of the cycle-time distribution:

$$\begin{aligned} c_{ZT}(z) &= c_{S,ZT}^g(z) \\ &= x(0) \cdot c_{0,ZT}(z) + (1 - x(0)) c_{1,ZT}(z). \end{aligned} \quad (3.24)$$

Equations (3.22)–(3.24) can be effectively evaluated using algorithms belonging to the Fast Fourier Transform (FFT) (based on the Discrete Fourier Transform (DFT), cf. [9, 23, 24]).

3.4. Discrete-Time Calculation Algorithm

The logical interrelation of the analysis steps in Subsections 3.2 and 3.3 is as follows. Assuming that the conditional cycle-time distributions are known, the Markov-chain state probabilities can be calculated according to Eqs. (3.16) and (3.17). To calculate the conditional cycle times, as formulated in Eqs. (3.20), (3.22) and (3.23), we need to know component $x(0)$ of the Markov-chain state probability vector. This fact leads to a numerical calculation algorithm based on an alternating evaluation of the cycle-time distributions and the state probability vector. This approximation algorithm is similar to the continuous-time algorithm presented in [22]. However, by taking advantage of fast convolution algorithms and the effective evaluation of Z-transform using FFT algorithms, the analysis in the discrete-time domain allows the calculation of the entire distributions (e.g., those of cycle times). This is in contrast, e.g., to using a number of moments (e.g., the two-moment matching method) to describe distribution functions.

The main steps of the algorithm are:

- i) Initialize the Markov-chain state probability vector and the conditional cycle-time distributions.
- ii) Calculate the conditional cycle-time distributions according to Eqs. (3.20), (3.22) and (3.23).

- iii) Calculate the arrival-group size distributions according to Eq. (3.17).
- iv) Calculate the Markov-chain state probability vector according to Eq. (3.16).
- v) Repeat steps ii), iii) and iv) until a convergence criterion is fulfilled. In the analysis here, the difference between two consecutive means of the state probability vector is used as convergence criterion.

4. SYSTEM CHARACTERISTICS

Using the equilibrium Markov-chain state probabilities obtained by the algorithm presented above, a number of performance measurements of interest can be derived. This will be discussed in this section.

4.1. Message Blocking Probability

Again, the assumption is made that the scanning and arrival processes are uncorrelated. We first consider the following conditional blocking probability defined by

$$B(k) = \Pr(\text{message is blocked} \mid X = k)$$

Probability for a message in an arrival group to be rejected, conditioned on a number $X = k$ of messages waiting in the queue observed at the previous scanning epoch.

Two cases have to be distinguished, $X = 0$ and $X > 0$. Derivation of the conditional blocking probability for case $X > 0$ will be described in detail.

For $X > 0$, i.e., the current cycle time is of type G_1 , we observe an arbitrarily chosen test message in the group arriving during the cycle. The probability for the test message to be in an arrival group of size i is $i \cdot g_1(i) / EG_1$. Assuming further the group size i , blocking will occur for $k - 1 + i > S$, where a fraction of $k - 1 + i - S$ of messages will be rejected. Accordingly, the probability of the test message being in the rejected fraction is $(k - 1 + i - S) / i$. Finally, for case $X > 0$ we obtain the following conditional blocking probability of messages:

$$\begin{aligned}
 B(k) &= \sum_{i=S-k+2}^{\infty} \frac{k-1+i-S}{i} \cdot \frac{i \cdot g_1(i)}{EG_1} \\
 &= \frac{1}{EG_1} \sum_{i=S-k+2}^{\infty} (k-1+i-S) \cdot g_1(i), \quad k = 1, \dots, S.
 \end{aligned} \tag{4.1}$$

The case of an empty queue at the previous scanning instant ($X = 0$) can be analyzed analogously:

$$B(0) = \sum_{i=S+1}^{\infty} \frac{i-S}{i} \cdot \frac{i \cdot g_0(i)}{EG_0} = \frac{1}{EG_0} \sum_{i=S+1}^{\infty} (i-S) \cdot g_0(i). \tag{4.2}$$

By eliminating the condition, we arrive at the blocking probability for messages arriving at a queue of the polling system

$$\begin{aligned}
 B_{\Lambda} &= \sum_{k=0}^S x(k) \cdot B(k) \\
 &= \frac{1}{EG_0} x(0) \sum_{i=S+1}^{\infty} (i-S) \cdot g_0(i) + \frac{1}{EG_1} \sum_{k=1}^S x(k) \sum_{i=S-k+2}^{\infty} (k-1+i-S) \cdot g_1(i).
 \end{aligned}
 \tag{4.3}$$

4.2. Arbitrary-Time State Probabilities

From the Markov-chain state probability vector, which describes the state process at regeneration points chosen prior to the scanning epochs, the state process seen by an arbitrary outside observer, which will be characterized by means of an arbitrary-time state probability vector, can be derived. Using this probability vector, further system characteristics, e.g., the mean waiting time for messages, can be obtained.

The state process will now be observed at an instant t^* chosen arbitrarily. Since we are operating in the discrete-time domain, t^* is conceived to be immediately prior to a discrete-time epoch (cf. Figure 3). The probability P_j^* ($j=0,1$) that the observation instant t^* lies in a conditional cycle of type C_j can be given as follows, using results of semi-Markov processes:

$$P_0^* = x(0) \frac{EC_0}{EC}
 \tag{4.4}$$

$$P_1^* = 1 - P_0^* = (1 - x(0)) \frac{EC_1}{EC} .
 \tag{4.5}$$

We further consider the distribution of the recurrence cycle time with r.v. C_j^* , which represents the discrete-time interval from the previous scanning instant upto the observation point t^* :

$$c_j^*(k) = \frac{1}{EC_j} \left(1 - \sum_{i=0}^k c_j(i) \right), \quad k = 0,1,\dots, \quad j = 0,1.
 \tag{4.6}$$

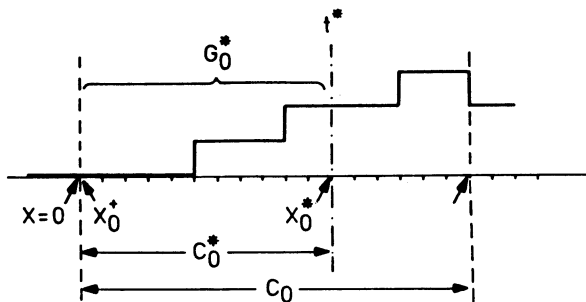


FIGURE 3
State process observation at an arbitrary instance.

Denoting G_j^* ($j=0,1$) to be the r.v. of the group of messages arriving during a recurrence cycle time of type C_j , we obtain a similar expression for the group size distribution as in Eq. (3.19):

$$g_j^*(k) = c_j^*(0) \cdot \delta(k) + \sum_{m=1}^{\infty} c_j^*(m) \sum_{i=0}^{m-1} (f^{(k)}(i) - f^{(k+1)}(i)), \quad k = 0,1,\dots, \quad j = 0,1. \tag{4.7}$$

As shown in Figure 3, the following r.v. for the state process in the queue being observed are introduced:

X^* r.v. for the state of the queue being observed at t^*

X, X^+ r.v. for the state of the queue observed immediately *prior to* and *after* the previous scanning epoch, respectively,

and, similar to the Markov-chain analysis [cf. Eqs. (3.1) and (3.2)], the random variables conditioned on the state prior to the previous scanning instant:

$$X_0^+ = X^+ | X = 0, \quad X_1^+ = X^+ | X > 0, \tag{4.8}$$

$$X_0^* = X^* | X = 0, \quad X_1^* = X^* | X > 0. \tag{4.9}$$

To calculate the probability vector for X^* , the following set of equations describing the interrelations between state random variables and corresponding distributions is required:

i) $X_n = 0$

$$X_0^+ = 0 \tag{4.10}$$

$$x_0^+(k) = \delta(k) \tag{4.11}$$

$$X_0^* = \min(X_0^+ + G_0^*, S) = \min(G_0^*, S) \tag{4.12}$$

$$x_0^*(k) = \pi^S(x_0^+(k) \star g_0^*(k)), \tag{4.13}$$

ii) $X_n > 0$

$$X_1^+ = X - 1 \tag{4.14}$$

$$x_1^+(k) = \begin{cases} \frac{x(k+1)}{\Pr(X > 0)} = \frac{x(k+1)}{1 - x(0)} & k = 0,1,\dots,S-1 \\ 0 & \text{otherwise} \end{cases} \tag{4.15}$$

$$X_1^* = \min(X_1^+ + G_1^*, S) \tag{4.16}$$

$$x_1^*(k) = \pi^S(x_1^+(k) \star g_1^*(k)) . \quad (4.17)$$

Hence, the arbitrary-time state probabilities can be given as

$$x^*(k) = P_0^* \cdot x_0^*(k) + P_1^* \cdot x_1^*(k) , \quad k = 0, 1, \dots, S. \quad (4.18)$$

Using Eqs. (4.4)–(4.6) and (4.10)–(4.17), we obtain the final set of equations to determine the arbitrary-time state probabilities:

$$\begin{aligned} x^*(k) &= \frac{EC_0}{EC} x(0) \cdot g_0^*(k) + \frac{EC_1}{EC} \sum_{i=1}^{k+1} x(i) \cdot g_1^*(k-i+1), \quad k = 0, \dots, S-1 \\ x^*(S) &= \frac{EC_0}{EC} x(0) \sum_{i=S}^{\infty} g_0^*(i) + \frac{EC_1}{EC} \sum_{i=1}^S x(i) \cdot \sum_{m=S-i+1}^{\infty} g_1^*(m), \quad k = S. \end{aligned} \quad (4.19)$$

4.3. Waiting Time of Accepted Messages

Having the state probabilities of a queue seen by an arbitrary outside observer, we can use Little's theorem to calculate the mean waiting time of *accepted* messages:

$$EW_{\Lambda} = \frac{EX^* \cdot EA}{1 - B_{\Lambda}} . \quad (4.20)$$

5. RESULTS AND APPROXIMATION ACCURACY

By means of the example of a symmetrical polling system with $g=5$ queues having finite capacities $S=5$, numerical results will be presented in this section to discuss the application of the calculation algorithm and the approximation accuracy. The switchover time is assumed to be deterministic equal to $0.5 EB$, and the discrete-time axis is scaled to $\Delta t = 1$. Time variables are normalized to the mean service time $EB = 10 \Delta t$. The offered traffic intensity is denoted by $\rho = EB/EA$.

To validate the approximate analysis, computer simulations are provided. The simulation results will be depicted with their 95 percent confidence intervals, calculated using the Student-t test technique.

To obtain a parametric representation of random process types, we consider here the interarrival and service times having distributions given by their two parameters, e.g., the mean and the coefficient of variation, whereby the negative binomial distribution is employed. Thus, for an r.v. X with mean EX and coefficient of variation c_X ,

$$x(k) = \binom{y+k-1}{k} p^y (1-p)^k , \quad 0 \leq p < 1 , \quad y \text{ real}, \quad (5.1)$$

where

$$p = \frac{1}{EX \cdot c_X} , \quad y = \frac{EX}{EX \cdot c_X - 1} , \quad EX \cdot c_X > 1 .$$

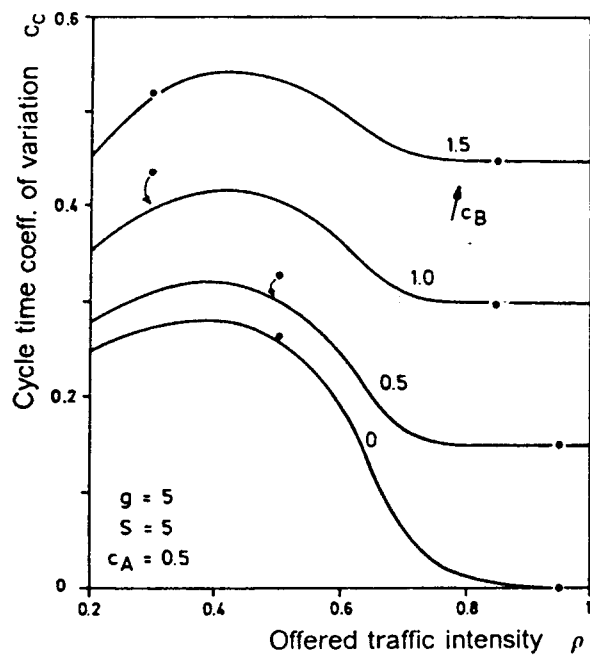
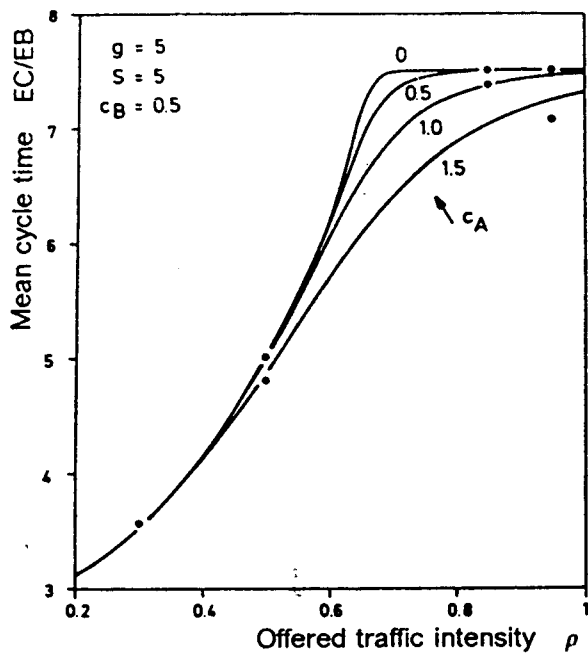


FIGURE 4

Influence of arrival process characteristics on cycle time.

FIGURE 5

Influence of service process on cycle time.

In Figures 4 and 5, the mean and the coefficient of variation of the cycle time are depicted as functions of the offered traffic intensity ρ for different types of interarrival and service processes. The choice of parameters in these figures is based on detailed studies showing that EC is mainly affected by the type of arrival processes, and c_C by the type of service processes. It should be noted here that the coefficients of variation of these discrete-time processes are chosen to be equivalent to the deterministic ($c_A, c_B = 0$), the Erlangian of fourth order ($c_A, c_B = 0.5$), the Markovian ($c_A, c_B = 1$), and the hyper-exponential ($c_A, c_B = 1.5$) distributions. With respect to the traffic level, the two lower and upper bounds of the mean cycle time can be observed (cf. Figure 4): i) the *empty cycle* at disappearing traffic intensities (sum of just switchover times), and ii) the *maximal cycle* at very high traffic levels (service at each queue during the polling cycle). For arrival processes with higher coefficients of variation, the blocking effect will become noticeable, which leads to smaller cycle lengths. As depicted in Figure 5, the cycle-time coefficient of variation has a maximum value which increases with increasing service-time coefficient of variation.

The mean waiting time and the blocking probability of messages are depicted in Figures 6 and 7, where attention is again devoted to the influence of the arrival process. The mean waiting time is drawn (Figure 6) for accepted messages. Hence, a crossover phenomenon can be recognized, which can be verified on considering the higher blocking probability, i.e., smaller number of accepted messages, for larger values of c_A (cf. Figure 7). For the case of disappearing traffic intensities, the waiting time of messages consists of just the recurrence time of empty cycles.

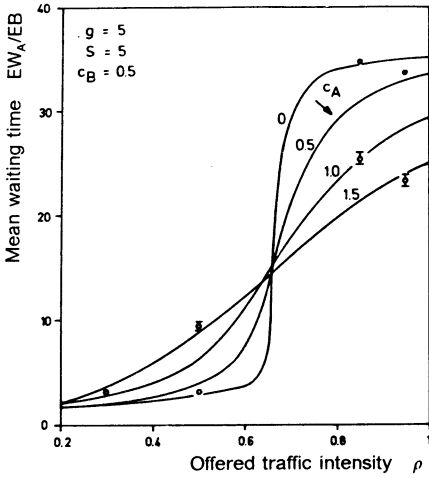


FIGURE 6

Mean waiting time versus offered traffic.

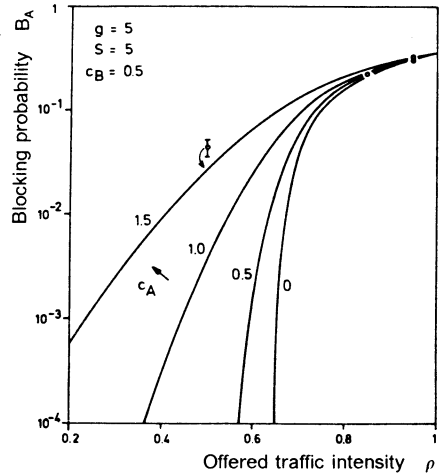


FIGURE 7

Influence of arrival process on blocking probability.

As mentioned above, the entire distributions of the random variables of interest can be obtained, using discrete-time analysis methods. An example is given in Figure 8, where the complementary cycle-time distribution function is depicted for different values of the service-time coefficient of variation. Two main effects can clearly be seen here: i) the step-wise functions, and ii) the geometric caudal characteristics of the discrete-time distribution functions obtained.

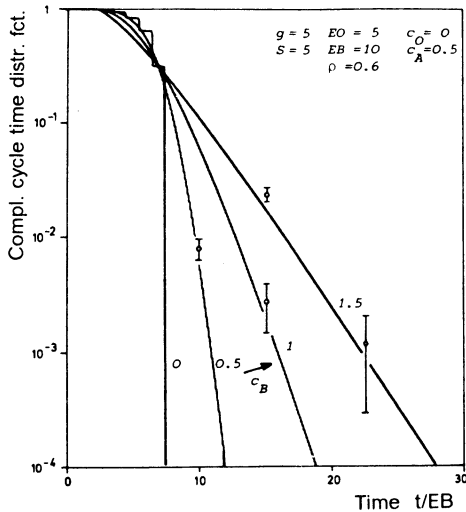


FIGURE 8

Complementary cycle-time distribution function.

As seen in the diagrams, with the exception of blocking probabilities at higher c_A values, the overall approximation accuracy for the system parameters given ($g=5$, $S=5$, $EO = 0.5 EB$) is good. However, two main restriction factors for the use of the approximation should be mentioned here: i) the accuracy of the approximate analysis is of decreasing tendency for smaller values of switchover time, and higher values of c_A , c_B , and ii) the computing efforts increase hyper-proportionally with the values of g and S , according to the lengths of probability vectors involved in convolution operations.

6. CONCLUSIONS

In this paper, an approximate algorithm for polling systems with finite capacity of waiting places and nonexhaustive service (or more precisely, limited service of one message per service) is presented. The analysis is done in the discrete-time domain, based on the evaluation of discrete convolution operations taking advantage of fast convolution algorithms, e.g., the Fast Fourier Transform. Attention is devoted to two essential modeling aspects: i) the consideration of general renewal input traffic, and ii) the assumption of finite capacity of waiting places in the system. Numerical examples are shown to illustrate the approximation accuracy of the analysis. Validation of the approximation is done by means of computer simulations. The class of polling models considered here can be applied in the performance analysis of a broad spectrum of models in computer and communication systems.

ACKNOWLEDGEMENTS

The author would like to thank M. Böpple for helpful discussions and valuable programming support, and Prof. P. Kühn (University of Stuttgart) for his interest in this work. Valuable comments and suggestions of his colleagues H. Ahmadi and M. Nassehi (IBM Zurich Research Laboratory) are greatly appreciated.

REFERENCES

- [1] Ackroyd M.H., "Computing the Waiting Time Distribution for the G/G/1 Queue by Signal Processing Methods", IEEE COM-28 (1980) 52-58.
- [2] Boxma O.J., "Two Symmetric Queues with Alternating Service and Switching Times", Proc. Performance '84, Paris, France, pp. 409-431.
- [3] Boxma O.J., and Meister B., "Waiting Time Approximations for Cyclic-Service Systems with Switch-Over Times", Proc. Joint Conf. Performance '86 and ACM Sigmetrics, Raleigh, NC, 1986, pp. 254-262.
- [4] Bux W., and Truong H.L., "Mean-Delay Approximation for Cyclic Service Queueing Systems", Perf. Eval. 3 (1981) 187-196.
- [5] Cooper R.B., and Murray G., "Queues served in Cyclic Order", Bell Syst. Tech. J. 48 (1969) 675-689.
- [6] Eisenberg M., "Queues with Periodic Service and Changeover Times", Operations Research 20 (1972) 440-451.
- [7] Hashida O., "Analysis of Multiqueue", Rev. El. Commun. Lab. 20 (1972) 189-199.
- [8] Hashida O., and Ohara K., "Line Accommodation Capacity of a Communication Control Unit", Rev. El. Commun. Lab. 20 (1972) 231-239.
- [9] Henrici P., "Fast Fourier Methods in Computational Complex Analysis", Siam Review 21 (1979) 481-527.
- [10] Kobayashi H., "Stochastic Modelling: Queueing Models; Discrete-Time Queueing Systems", in Part II, Louchard G., and Latouche G. (eds.), "Probability Theory and Computer Science" (Academic 1983).

- [11] Konheim A.G., "An Elementary Solution of the Queuing System GI/G/1", *SIAM J. Comp.* 4 (1975) 540-545.
- [12] Kuehn P.J., "Multiqueue Systems with Nonexhaustive Cyclic Service", *Bell Syst. Tech. J.* 58 (1979) 671-699.
- [13] Kuehn P.J., "Performance of ARQ-Protocols for HDX-Transmission in Hierarchical Polling Systems", *Perf. Eval.* 1 (1981) 19-30.
- [14] Leibowitz M.A., "An Approximate Method for Treating a Class of Multiqueue Problems", *IBM J. Res. Develop.* 5 (1961) 204-209.
- [15] Lindley D.V., "The Theory of Queues with a Single Server", *Proc. of the Cambridge Philosophical Society*, 48 (1952) 277-289.
- [16] Manfield D.R., "Analysis of a Polling System with Priorities", *Proc. Globecom'83*, San Diego, CA, 1983, paper 43.4.
- [17] Morris R.J.T., and Wang Y.T., "Some Results for Multiqueue Systems with Multiple Cyclic Servers", *Proc. 2nd Int. Symp. on the Performance of Comp. Comm. Systems*, Zurich, Switzerland, 1984, 245- 258.
- [18] Raith T., "Performance Analysis of Multibus Interconnection Networks in Distributed Systems", *Proc. 11th Int. Teletraffic Congress*, Sept. 1985, Kyoto, Japan, paper 4.2A-5.
- [19] Raith T., and Tran-Gia P., "Performance Analysis of Polling Mechanisms with Receiver Blocking", *Proc. 8th Int. Conf. on Comp. Comm. (ICCC)*, Munich, FRG, Sept. 1986, pp C5-1.
- [20] Smith W.L., "On the Distribution of Queuing Times", *Proc. of the Cambridge Philosophical Society*, 49 (1953) 449-461.
- [21] Takagi H., and Kleinrock L., "Analysis of Polling Systems", *Japan Science Inst. Research Report*, 1985.
- [22] Tran-Gia P., and Raith T., "Multiqueue Systems with Finite Capacity and Nonexhaustive Cyclic Service", *Proc. Int. Seminar on Computer Networking and Performance Evaluation*, Tokyo, Japan, Sept. 1985.
- [23] Tran-Gia P., "Discrete-Time Analysis for the Interdeparture Distribution of GI/G/1 Queues", *Proc. Seminar on Teletraffic Analysis and Computer Performance Evaluation*, Amsterdam, The Netherlands, June 1986.
- [24] Tran-Gia P., and Rathgeb E., "Performance Analysis of Semidynamic Scheduling Strategies in Discrete-Time Domain", *Proc. INFOCOM'87*, San Francisco, CA, 1987.