Contribution to the GI/NTG Conference on Measurement, Modelling and Evaluation of Computer Systems, October 1985, Dortmund, F.R.G.

Approximation for Finite Capacity Multiqueue Systems

P. Tran-Gia
T. Raith

Institute of Communications Switching and Data Technics
University of Stuttgart
Federal Republic of Germany

ABSTRACT

In performance investigations of token ring local area networks, switching systems with distributed control etc., the class of polling models, i.e. multiqueue systems with cyclic service is often employed. In this paper, an approximate analysis method for this class of models will be presented, whereby realistic modelling assumptions like the finiteness of queue capacities and nonsymmetrical load conditions are taken into account. The method of imbedded Markov chain is used for the analysis, whereby the special case of Markovian as well as the case of general service time are successively considered. The latter case is analyzed in conjunction with a moment matching approach for the cycle time. The approximation accuracy is validated by means of computer simulations. Numerical results are shown in order to illustrate the accuracy of the calculation method and its dependency on system parameters.

1. INTRODUCTION

Polling models are often used in performance investigations of communication and computer systems, e.g. switching systems with distributed control structures or local area networks with token passing protocols. Most of investigations with varying degree of model complexity consider queues having infinite capacities, where several aproximation techniques for the system analysis are proposed.

In the literature, multiqueue systems served by a single server have been the subject of numerous studies [1-11]. A number of modelling approaches considering various polling mechanisms like cyclic or priority order and several service disciplines, e.g., exhaustive, nonexhaustive or gating are considered. Some of these studies take into account the switchover time, i.e. the time interval spent by the server to switch over from one queue to the succeeding one. In most of the investigations the queues are assumed to have infinite capacity and the analysis is often derived using the imbedded Markov chain technique [1,2,3,4,5,8].

Multiqueues systems with cyclic polling strategy, symmetrical load conditions, constant switchover time and gating service were approximately analysed by Leibowitz [7]. In [1,2] Cooper and Murray have considered a cyclic polling system with gating or exhaustive service and zero switchover time. The case of two queues with general switchover time was taken into account by Eisenberg [3]. The approach of Cooper and Murray in [1] has been generalized by Eisenberg [4] and Hashida [5] to non-zero switchover time. An approximation technique for cyclic queues with nonexhaustive service and general switchover time has been developed by Kuehn [8]. Morris and Wang [12] and Raith [13] provided analytical approaches to deal with polling systems with multiple servers. An exact solution for a system with two queues and nonzero switchover time was presented by Boxma [10]. A survey on polling system analysis, where various system classes are considered, was provided by Takagi and Kleinrock [11].

In all practical systems, nonsymmetrical load conditions are often observed and all buffers are of finite capacity. In order to obtain a realistic model of a nonsymmetrically loaded or partly overloaded system, in which blocking of incoming messages may occur, the finite capacity of some overloaded queues must be taken into account. For the investigation of such finite capacity systems, established analysis methods for infinite multiqueue systems, which usually operate with generating functions or in Laplace domain, do not lead to closed expressions or effective algorithms for system characteristics of interest.

In the following, an approximate analysis method for polling systems with finite queue capacity and nonexhaustive cyclic service is presented, whereby a numerical algorithm is developed. According to the type of the service time distribution function (Markovian or general), two numerical schemes for the cycle time calculation in conjunction with an iteration method are derived. The accuracy of the approximative method will be illustrated by means of numerical results, which have been obtained for a wide range of system parameters.

2. MODEL DESCRIPTION

In Fig. 1 the queueing model of a single server polling system is depicted. The model consists of a number g of finite capacity queues. At each queue it is assumed that messages arrive according to a Poisson process. We consider a single server which scans the queues in a cyclic order and serves it nonexhaustively, i.e., if there are messages waiting, one message will be served per server visit instant. After the service at queue j is finished the server will move to the succeeding queue j+1. If no message is waiting in queue j the observed interscan period consists just of the switchover time. The switchover time models all overheads spent and all procedures performed by the server to move to and to scan the succeeding queue. In order to investigate nonsymmetrical systems, the arrival rate, the mean service time, the mean switchover time and the queue capacity can individually be chosen for each queue. Furthermore, service times and switchover times are assumed to be generally distributed and the type of their distribution functions can also individually be chosen for each queue.

The following symbols and random variables (r.v.) are used in this paper :

- g number of queues in the system
- λ_j arrival rate of messages offered to queue j
- T_{hj} r.v. for the service time of messages in queue j
- T_{uj} r.v. for the switchover time corresponding to queue j
- S_j capacity of queue j

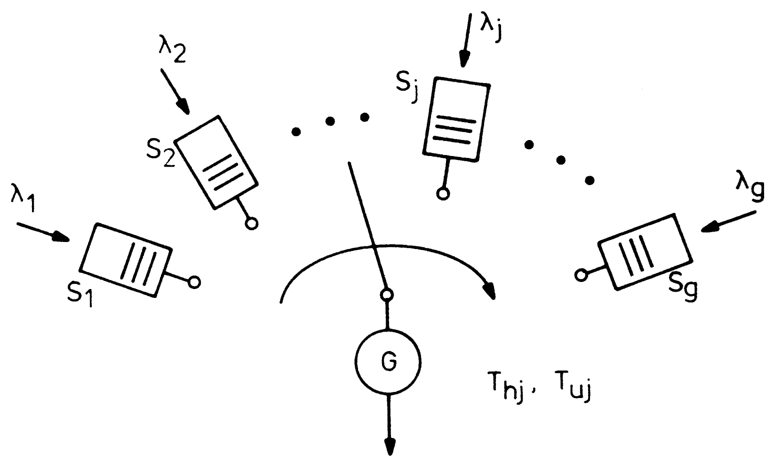

Fig. 1: Single server polling system with finite queue capacity

3. ANALYSIS

In this section, a numerical algorithm for an approximate analysis of finite capacity multiqueue systems will be derived. Basically, the analysis draws upon approaches presented in [6,8], using the technique of the imbedded Markov chain. However, some modifications must be provided in order to take into account the blocking effect and the finiteness of queue capacities.

In subsection 3.1 general equations for the state analysis and system characteristics will be derived, while in subsections 3.2 and 3.3 the special case of Markovian server and subsequently, the case of generally distributed service time will be investigated.

For a random variable (e.g. r.v. T) we use the following notations

- $F(t)$ probability distribution function
- $f(t)$ probability density function
- $\Phi(s)$ Laplace-Stieltjes-Transform of $F(t)$
- $\overline{T}$ mean value of T
- $c[T]$ coefficient of variation of T.

3.1 General Equations for the Analysis

3.1.1 Markov Chain State Probabilities

A particular queue j is considered in the following, which is observed at polling instants. Let t_n be the time of the n-th scanning epoch and let $X^{(n)}(0^-)$ be the number of messages in this queue at time t_n^-, i.e. just prior to the n-th scanning epoch, we define the Markov chain state probabilities

$$P_{k,j}^{(n)} = \Pr\{ X^{(n)}(0^-) = k \} , \quad k = 0,1,\ldots,S_j \tag{3.1}$$

and the steady state probabilities of the Markov chain are defined from

$$P_{k,j} = \Pr \{ X(0^-) = k \} , \quad k = 0,1,\ldots,S_j \tag{3.2}$$

For ease of reading, the subscript j indicating the observed queue will be suppressed, e.g., the notation P_k will be used instead of $P_{k,j}$.

In order to calculate the transition probabilities of the Markov chain

$$p_{jk} = \Pr \{ X^{(n+1)}(0^-) = k \mid X^{(n)}(0^-) = j \}, \tag{3.3}$$

we observe the system state $X^{(n)}(t)$ of the queue at time t_n+t. Considering the pure birth process in the queue between two consecutive scanning epochs, i.e. during a scanning cycle, the state probabilities at time t_n+t denoted by

$$P_k^{(n)}(t) = \Pr \{ X^{(n)}(t) = k \} , \quad k = 0,1,\ldots,S_j \tag{3.4}$$

can be obtained as follows

$$\begin{aligned} P_k^{(n)}(t) &= P_0^{(n)} a_k(t) + \sum_{i=0}^{k} P_{i+1}^{(n)} a_{k-i}(t) , \quad k = 0,1,\ldots,S_j-1 \\ P_{S_j}^{(n)}(t) &= P_0^{(n)} \sum_{i=S_j}^{\infty} a_i(t) + \sum_{i=0}^{S_j-1} P_{i+1}^{(n)} \sum_{m=S_j-i}^{\infty} a_m(t) \end{aligned} \tag{3.5}$$

where

$$a_m(t) = \frac{(\lambda_j t)^m}{m!} e^{-\lambda_j t} . \tag{3.6}$$

Using the consideration of conditional cycle time [8], where the following random variables (r.v.) are defined

- T_C r.v. for the cycle time with respect to the observed queue j
- T_C' r.v. for a cycle, conditioning on an empty queue at the previous scanning instant (i.e., without service of queue j during the cycle)
- T_C'' r.v. for a cycle, conditioning on an non-empty queue at the previous scanning instant (i.e., with service of queue j during the cycle),

the state equations which implicitly contain the transition probabilities can be written as

$$P_k^{(n+1)} = P_0^{(n)} \int_0^\infty a_k(t)\, f_{C'}(t)dt + \sum_{i=1}^{k+1} P_i^{(n)} \int_0^\infty a_{k-i+1}(t)\, f_{C''}(t)dt \qquad k=0,1,\dots,S_j-1$$

$$P_{S_j}^{(n+1)} = P_0^{(n)} \sum_{i=S_j}^{\infty} \int_0^\infty a_i(t)\, f_{C'}(t)dt + \sum_{i=1}^{S_j} P_i^{(n)} \sum_{m=S_j-i+1}^{\infty} \int_0^\infty a_m(t)\, f_{C''}(t)dt. \tag{3.7}$$

Defining the arrival probabilities, i.e. the probabilities for m arrivals during a conditional cycle of type T_C' or T_C''

$$b_m' = \int_0^\infty a_m(t)\, f_{C'}(t)\, dt$$
$$b_m'' = \int_0^\infty a_m(t)\, f_{C''}(t)\, dt \tag{3.8}$$

we obtain from (3.7) the following set of Markov chain state equations

$$P_k^{(n+1)} = P_0^{(n)}\, b_k' + \sum_{i=1}^{k+1} P_i^{(n)}\, b_{k-i+1}'' \qquad , k=0,1,\dots,S_j-1$$
$$P_{S_j}^{(n+1)} = P_0^{(n)} \sum_{i=S_j}^{\infty} b_i' + \sum_{i=1}^{S_j} P_i^{(n)} \sum_{m=S_j-i+1}^{\infty} b_m'' . \tag{3.9}$$

Eqn. (3.9) will be used for the numerical calculation of the steady state probabilities $\{P_k\}$. It remains to determine the arrival probabilities $\{b_m'\}$ and $\{b_m''\}$ in (3.8), which is the subject of subsections 3.2 and 3.3.

3.1.2 Conditional Cycle Time

Define $T_{E,j}$ to be the random variable for the time interval between the scanning epochs of queue j and (j+1), i.e., the segment of the cycle time corresponding to queue j, with the Laplace-Stieltjes-Transform (LST)

$$\Phi_{E,j}(s) = \Phi_{uj}(s) \cdot ((1-P_{0,j})\, \Phi_{hj}(s) + P_{0,j}). \tag{3.10}$$

Under the assumption of independence between $T_{E,j}$, j=1,2,...,g, the LST of the conditional cycle times can be given as follows

$$\Phi_{C',j}(s) = \Phi_{uj}(s) \prod_{\substack{k=1\\k\neq j}}^{g} \Phi_{E,k}(s) \tag{3.11a}$$

$$\Phi_{C'',j}(s) = \Phi_{uj}(s)\, \Phi_{hj}(s) \prod_{\substack{k=1\\k\neq j}}^{g} \Phi_{E,k}(s) . \tag{3.11b}$$

Eqns.(3.10) and (3.11a,b) yield the first two moments of the conditional cycle times, thus

$$\bar{T}'_{C,j} = \bar{T}_{uj} + \sum_{\substack{k=1\\k\neq j}}^{g} \bar{T}_{E,k} \quad ,VAR[T'_{C,j}] = VAR[T_{uj}] + \sum_{\substack{k=1\\k\neq j}}^{g} VAR[T_{E,k}]$$

$$\bar{T}''_{C,j} = \bar{T}_{uj} + \bar{T}_{hj} + \sum_{\substack{k=1\\k\neq j}}^{g} \bar{T}_{E,k} \quad ,VAR[T''_{C,j}] = VAR[T_{uj}] + VAR[T_{hj}] + \sum_{\substack{k=1\\k\neq j}}^{g} VAR[T_{E,k}] \ , \qquad \text{... (3.12)}$$

where

$$\bar{T}_{E,j} = \bar{T}_{uj} + (1-P_{0,j})\,\bar{T}_{hj} \qquad (3.13)$$

$$VAR[T_{E,j}] = VAR[T_{uj}] + (1-P_{0,j})VAR[T_{hj}] + P_{0,j}(1-P_{0,j})\,\bar{T}_{hj}^2 \ .$$

In general eqns. (3.11a,b) which determine the conditional cycle times can be used in conjunction with the state equations (3.9) for the calculation of the Markov chain state probabilities.

The main idea in the calculation method presented in this paper is to develop an alternating calculation algorithm to obtain the Markov chain state probabilities $\{P_k\}$ and the cycle times $\{T'_C, T''_C\}$, which fulfill the eqns. (3.9) and (3.11a,b).

3.1.3 Arbitrary Time State Probabilities

In order to calculate system characteristics, e.g. blocking probability for messages or mean waiting time in the queue, it is useful to obtain first the arbitrary time state probabilities (cf.[15]). Again, the subscript j indicating the observed queue is suppressed in this subsection.

Define $\{P^*_k, k=0, 1, ..., S_j\}$ to be the arbitrary time state probabilities, i.e. the distribution of the number of messages in the considered queue j at an arbitrary observation instant.

The time interval from the last scanning epoch until the observation point is the backward recurrence time with the probability density function (pdf)

$$f^r_{C'}(t) = (1 - F_{C'}(t)) \,/\, \bar{T}'_C \qquad (3.14a)$$

and

$$f^r_{C''}(t) = (1 - F_{C''}(t)) \,/\, \bar{T}''_C \ . \qquad (3.14b)$$

The arrival probabilities during the backward recurrence time T'^r_C and T''^r_C can be given as

$$b'^*_m = \int_0^\infty a_m(t)\, f^r_{C'}(t)\, dt \qquad (3.15a)$$

$$b''^*_m = \int_0^\infty a_m(t)\, f^r_{C''}(t)\, dt \ . \qquad (3.15b)$$

Considering both types of conditional cycle times and combining the above results the arbitrary time state probabilities can be given (c.f. [15])

$$P^*_k = \frac{\bar{T}'_C}{\bar{T}_C} P_0 b'^*_k + \frac{\bar{T}''_C}{\bar{T}_C} \sum_{i=1}^{k+1} P_i b''^*_{k-i+1} \quad , \ k=0,1,...,S_j-1$$

$$P^*_{S_j} = \frac{\bar{T}'_C}{\bar{T}_C} P_0 \sum_{i=S_j}^{\infty} b'^*_i + \frac{\bar{T}''_C}{\bar{T}_C} \sum_{i=1}^{S_j} P_i \sum_{m=S_j-i+1}^{\infty} b''^*_m \ . \qquad (3.16)$$

3.1.4 System Characteristics

With the arbitrary time state probabilities given from eqn. (3.16) and taking into account the Poisson arrival process offered to the observed queue j, the blocking probability for messages in queue j can be determined :

$$B_j = P^*_{S_j} \tag{3.17}$$

The mean delay in queue j, referred to transmitted messages, is found from Little's law as

$$T_{wj} = \frac{L_j}{\lambda_j(1-B_j)} , \tag{3.18}$$

where L_j is the mean length of queue j

$$L_j = \sum_{i=1}^{S_j} i\, P^*_i . \tag{3.19}$$

It should be noted here that the well-known formula for the mean cycle time [8] is obtained in a modified form for the case of finite queue capacity

$$T_C = \frac{\sum_{j=1}^{g} \bar{T}_{uj}}{1 - \sum_{j=1}^{g} \lambda_j \bar{T}_{hj}(1-B_j)} . \tag{3.20}$$

3.2 The Case of Markovian Server and Constant Switchover Time

In this subsection we devote attention to symmetrical polling systems with Markovian service time and constant switchover time, i.e.

$$\lambda_j = \lambda \; ; \quad T_{hj} = T_h = \frac{1}{\mu} \; ; \quad \phi_{hj}(s) = \phi_h(s) = \frac{\mu}{s+\mu} \quad ; \; j = 1,\ldots,g.$$

Under these assumptions, the conditional cycle times can be given as follows (c.f. (3.10,11))

$$\phi_{C'}(s) = e^{-st_0}(P_0 + (1-P_0)\,\phi_h(s))^{g-1}$$

$$\phi_{C''}(s) = e^{-st_0}\,\phi_h(s)(P_0 + (1-P_0)\,\phi_h(s))^{g-1} , \qquad \ldots(3.21)$$

where $t_0 = g\bar{T}_u$.

It can clearly be seen that the cycle time probability density functions consist of terms corresponding to Erlangian density functions, which are deferred according to the total system switchover time (t_0). Thus, the conditional cycle time density functions can be given in the time domain as

$$f_{C'}(t) = \sum_{i=0}^{g-1} \binom{g-1}{i} P_0^{g-1-i} (1-P_0)^i f_{E,i}(t-t_0)$$

$$f_{C''}(t) = \sum_{i=0}^{g-1} \binom{g-1}{i} P_0^{g-1-i} (1-P_0)^i f_{E,i+1}(t-t_0) , \qquad \ldots(3.22)$$

where $f_{E,i}(t)$ denotes the Erlangian probability density function of i-th order. With eqn. (3.22) the arrival probabilities given in eqn. (3.8) can be calculated in a straightforward manner as

$$b'_m = P_0^{g-1} a_m(t_0) + \sum_{i=1}^{g-1} \binom{g-1}{i} P_0^{g-1-i} (1-P_0)^i \sum_{j=0}^{m} \binom{j+i-1}{j} q^j (1-q)^i a_{m-j}(t_0)$$

$$b''_m = \sum_{i=0}^{g-1} \binom{g-1}{i} P_0^{g-1-i} (1-P_0)^i \sum_{j=0}^{m} \binom{j+i}{j} q^j (1-q)^{i+1} a_{m-j}(t_0) ,$$

where $q = \rho/(1 + \rho)$ with $\rho = \lambda/\mu$ . ...(3.23)

Based on these arrival probabilities, the Markov chain state probabilities can be calculated according to eqn. (3.9). In order to calculate the arbitrary time state probabilities of a queue, the arrival probabilities have to be calculated. The density functions of the backward recurrence conditional cycle times are

$$f^r_{C'} = \frac{1}{\bar{T}_{C'}}(1-F_{C'}(t)) = \frac{1}{\bar{T}_{C'}} \int_{\tau=t}^{\infty} \sum_{i=0}^{g-1} \binom{g-1}{i} P_0^{g-1-i} (1-P_0)^i f_{E,i}(\tau-t_0)\, d\tau$$

$$f^r_{C''} = \frac{1}{\bar{T}_{C''}}(1-F_{C''}(t)) = \frac{1}{\bar{T}_{C''}} \int_{\tau=t}^{\infty} \sum_{i=0}^{g-1} \binom{g-1}{i} P_0^{g-1-i} (1-P_0)^i f_{E,i+1}(\tau-t_0)\, d\tau .$$

...(3.24)

Using eqn.(3.24), the probability of m arrivals during the backward recurrence conditional cycle times as stated in eqn. (3.15) can be derived

$$b'^*_m = \frac{1}{\lambda\bar{T}_{C'}} \{ P_0^{g-1} \sum_{j=m+1}^{\infty} a_j(t_0) + \sum_{i=1}^{g-1} \binom{g-1}{i} P_0^{g-1-i} (1-P_0)^i$$
$$\cdot (\sum_{j=0}^{m} a_{m-j}(t_0) \sum_{k=0}^{i-1} \binom{j+k}{j} (1-q)^k q^{j+1} + \sum_{j=m+1}^{\infty} a_j(t_0)) \}$$

$$b''^*_m = \frac{1}{\lambda\bar{T}_{C''}} \{ \sum_{i=0}^{g-1} \binom{g-1}{i} P_0^{g-1-i} (1-P_0)^i$$
$$\cdot (\sum_{j=0}^{m} a_{m-j}(t_0) \sum_{k=0}^{i} \binom{j+k}{j} (1-q)^k q^{j+1} + \sum_{j=m+1}^{\infty} a_j(t_0)) \} .$$

...(3.25)

Thus, the arbitrary time state probabilities of a considered queue can be calculated according to eqn. (3.16), in order to obtain system characteristics.

3.3 The Case of General Service Time

In this subsection, the service time distribution is assumed to be generally distributed. The calculation of the conditional cycle times as well as of the state probabilities are of higher complexity (c.f. [15]).

Since the expressions of the conditional cycle times (3.11a,b) are given in the Laplace-Stieltjes domain, a Laplace inversion procedure should have been utilized during each iteration cycle, in order to enable the calculation of eqns. (3.7) or (3.8). However, for reasons of computing efforts, the two-moment approximation technique, as proposed in [14], will be used in this subsection. This will briefly be described in the following.

According to the mentioned moment matching method [14], a given random variable T with the expected value $\bar{T}$ and the coefficient of variation c is approximately described by means of the following substitute distribution function F(t) :

i. $0 \leq c \leq 1$

$$F(t) = \begin{cases} 0 & 0 \leq t \leq t_1 \\ 1 - e^{-(t-t_1)/t_2} & t > t_1 \end{cases} \tag{3.26a}$$

where $t_1 = \bar{T}(1-c)$ and $t_2 = \bar{T}c$

ii. $c \geq 1$

$$F(t) = 1 - p\, e^{-t/t_1} - (1-p)\, e^{-t/t_2} \tag{3.26b}$$

where

$$t_{1,2} = \bar{T}\left(1 \pm \sqrt{\frac{c^2-1}{c^2+1}}\right)^{-1}$$

and $p = \bar{T}/2t_1$, $p\, t_1 = (1-p)\, t_2$.

The method is applied to the conditional cycle times T_C' and T_C'', which are derived in (3.11a,b) in conjunction with their means and variances in (3.12). We obtain subsequently the arrival probabilities as follows, where $\{b_m\}$ are given for $\{b_m'\}$ and $\{b_m''\}$

i. $0 \leq c \leq 1$

$$b_m = \int_0^{\infty} a_m(t)\, f_C(t)\, dt = \frac{(\lambda t_2)^m}{(1+\lambda t_2)^{m+1}}\, e^{-\lambda t_1} \sum_{k=0}^{m} \frac{t_1}{t_2} (\lambda t_2 + 1)^k / k! \tag{3.27a}$$

ii. $c \geq 1$

$$b_m = p\, \frac{(\lambda t_1)^m}{(1+\lambda t_1)^{m+1}} + (1-p)\, \frac{(\lambda t_2)^m}{(1+\lambda t_2)^{m+1}} \,. \tag{3.27b}$$

Based on the arrival probabilities (3.27a,b) and the state equations (3.9) the Markov chain state probabilities are calculated. Analogously, the arrival probabilities during the backward recurrence conditional cycle times given by eqns. (3.15a,b) can be explicitly written as follows:

i. $0 \leq c \leq 1$

$$b_m^* = \frac{1}{\lambda(t_1+t_2)} \left(1 - \sum_{k=0}^{m} \frac{(\lambda t_1)^k}{k!}\, e^{-\lambda t_1}\right) + \frac{t_2}{t_1+t_2}\, b_m \tag{3.28a}$$

where b_m corresponds to eqn.(3.27a)

ii. $c \geq 1$

$$b_m^* = \frac{(\lambda t_1^m)}{2(1+\lambda t_1)^{m+1}} + \frac{(\lambda t_2^m)}{2\,(1+\lambda t_2)^{m+1}} \tag{3.28b}$$

3.4 Calculation Algorithm for Markov Chain State Probabilities

In this subsection a numerical algorithm is given to calculate the Markov chain state probabilities. The algorithm, which utilizes the derived expressions for the Markov chain state probabilities and the conditional cycle time [15], is an alternating iteration scheme with the following main steps :

Step 1: Depending on the actual Markov chain state probabilities, the conditional cycle time probability density function is obtained according to the expressions given in subsection 3.2 or 3.3, corresponding to the type of the service time distribution function

Step 2: Calculation of the arrival probability vectors

Step 3: Based on the arrival probability vectors the Markov chain state probabilities of all queues are determined in a cyclic manner according to eqn. (3.9); the obtained values will be used in the next iteration cycle. In the case of a symmetrical system, the calculation must only be done for one queue.

The convergence criteria for the iteration are defined from

$$\Sigma \Delta_j = \sum_{j=1}^{g} \Delta_j < \varepsilon \text{ with } \Delta_j = \left| \sum_{k=1}^{S_j} kP_{k,j}^{(n)} - \sum_{k=1}^{S_j} kP_{k,j}^{(n-1)} \right| \qquad (3.29)$$

If the convergence condition of the iteration is fulfilled, the arbitrary time state probabilities according to subsection 3.1 are calculated and, subsequently, the performance measures required.

4. APPROXIMATION ACCURACY

In this section, numerical results will be presented and discussed for the two cases of symmetrically and nonsymmetrically loaded polling systems, in order to illustrate the accuracy of the algorithm presented in subsection 3. For both systems the time variables are standardized by $T_{hj} = 1$, $j=1,2,\ldots,g$, and the switchover time is chosen to be constant. The coefficient of variation of the service time is denoted by $c_h = c[T_{hj}]$, $j=1,2,\ldots,g$.

In order to validate the approximation, computer simulations are provided. The simulation results are depicted with their 95 percent confidence intervals.

4.1 Systems with Symmetrical Load Conditions

In this subsection, two symmetrically loaded systems with g=4 and g=20 are taken into account. Figs. 2 and 3 depict the mean and the coefficient of variation of the cycle time as function of the offered traffic intensity

$$\rho_0 = \sum_{j=1}^{g} \lambda_j \bar{T}_{hj} \qquad (4.1)$$

for different values of the number g of queues, the mean switchover time and the service time coefficient of variation. Fig. 2 exhibits the effect that servers with higher variance lead to shorter cycle times. This effect can be explained considering the higher blocking probability by large c_h.

It can also be seen in Fig. 2 that the mean cycle time approaches the limiting value (the maximal cycle), given by the sum of switchover times and average service times for all queues. As expected, for the two limiting cases of the mean cycle time, which correspond to lower traffic intensities (the empty cycle) and overloads (the maximal cycle), the cycle time coefficient of variation is very small (c.f. Fig. 3). This effect is caused by the very high (or low) probability for a queue to be empty at the considered low (or high) traffic intensity. As depicted in Fig. 3, there exists a maximum value for the cycle time coefficient of variation. This maximum increases by increasing service time variation or decreasing mean switchover time.

The mean waiting time and the blocking probability for messages are shown as functions of the offered traffic intensity in Figs. 4 and 5, respectively, for different service time coefficients of variation. In Fig. 4, a crossover effect of the waiting time characteristics can be recognized. While the overall approximation accuracy is satisfying concerning the waiting time, the algorithm presented is less accurate for large values of the service time coefficient of variation.

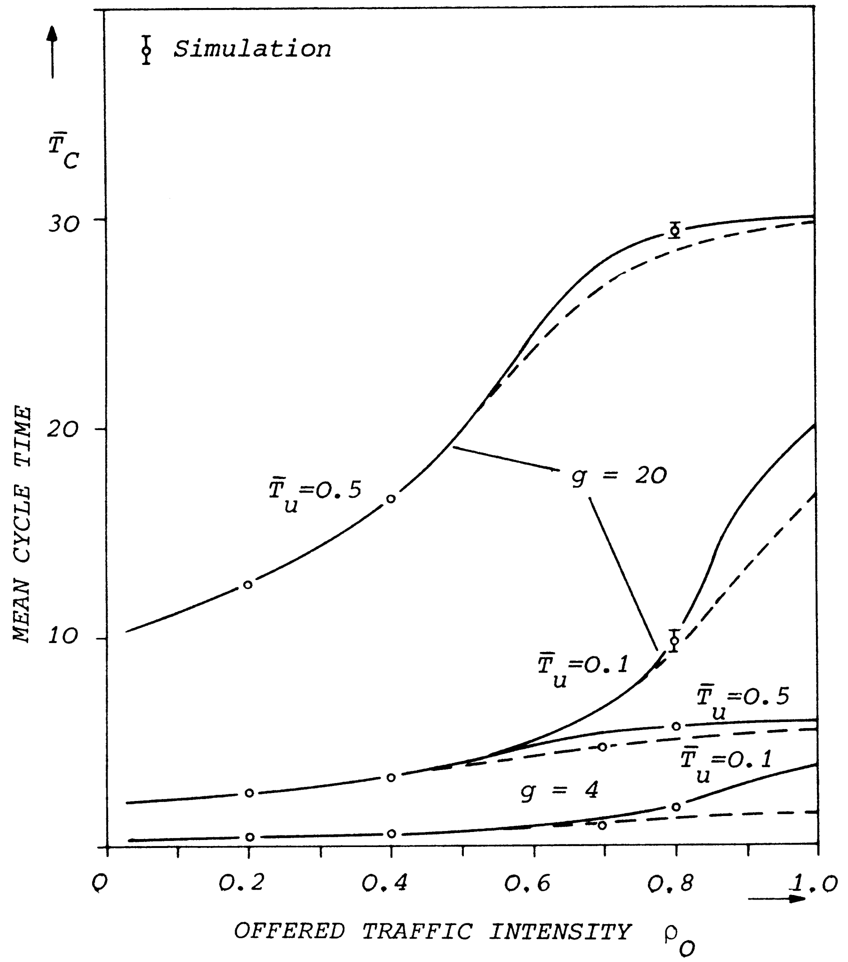

Fig. 2

Mean cycle time vs offered traffic (symmetrical load)

Parameters:
$\bar{T}_h = 1;\ c_u = 0$
$S_j = 5;\ j=1,\dots,g$
$c_h = 1$ ———
$c_h = 5$ -----

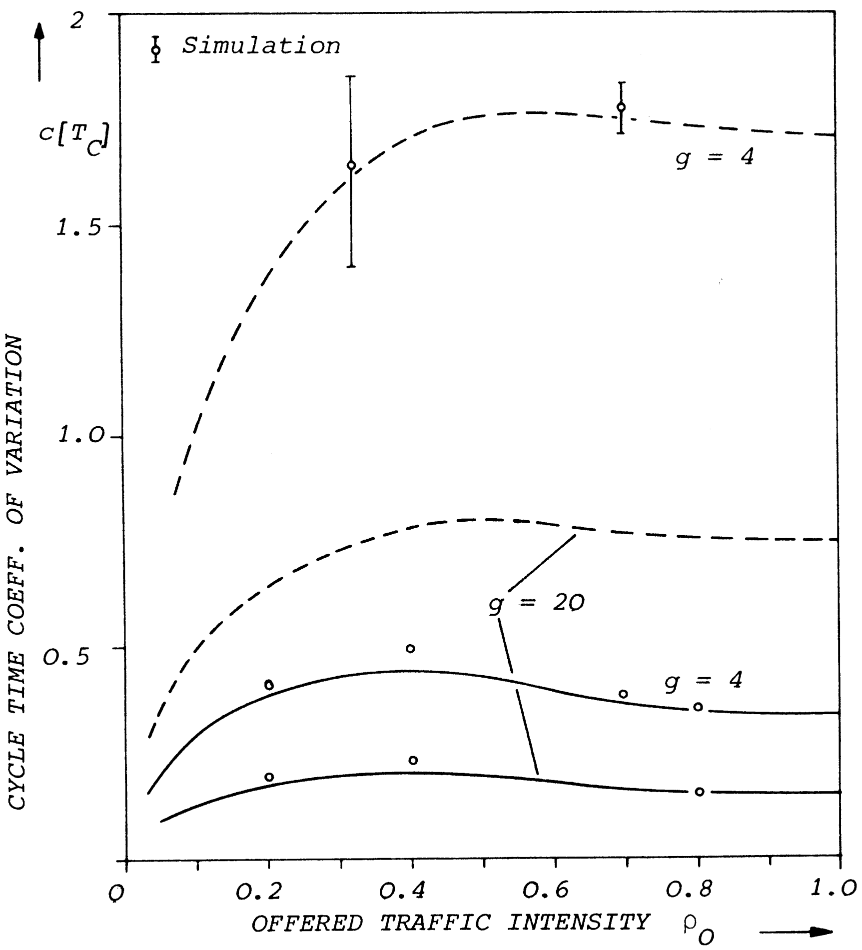

Fig. 3

Cycle time coefficient of variation vs offered traffic (symmetrical load)

Parameters:
$\bar{T}_{uj}=\bar{T}_u=0.5\cdot\bar{T}_h;\ c_u=0$
$S_j = 5;\ j=1,\dots,g$
$c_h = 1$ ———
$c_h = 5$ -----

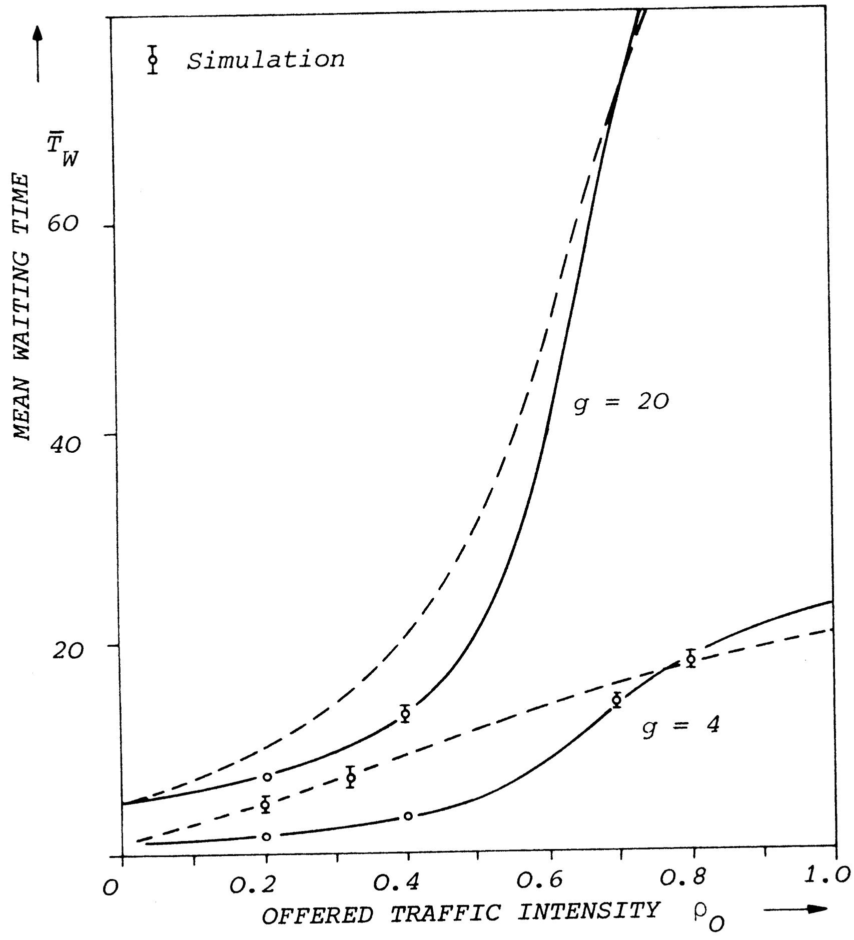

Fig. 4

Mean waiting time vs offered traffic (symmetrical load)

Parameters:
$\bar{T}_{uj} = \bar{T}_u = 0.5 \cdot \bar{T}_h$; $\bar{T}_h = 1$
$S_j = 5$; $j = 1, \ldots, g$; $c_u = 0$
$c_h = 1$ ———
$c_h = 5$ - - - - -

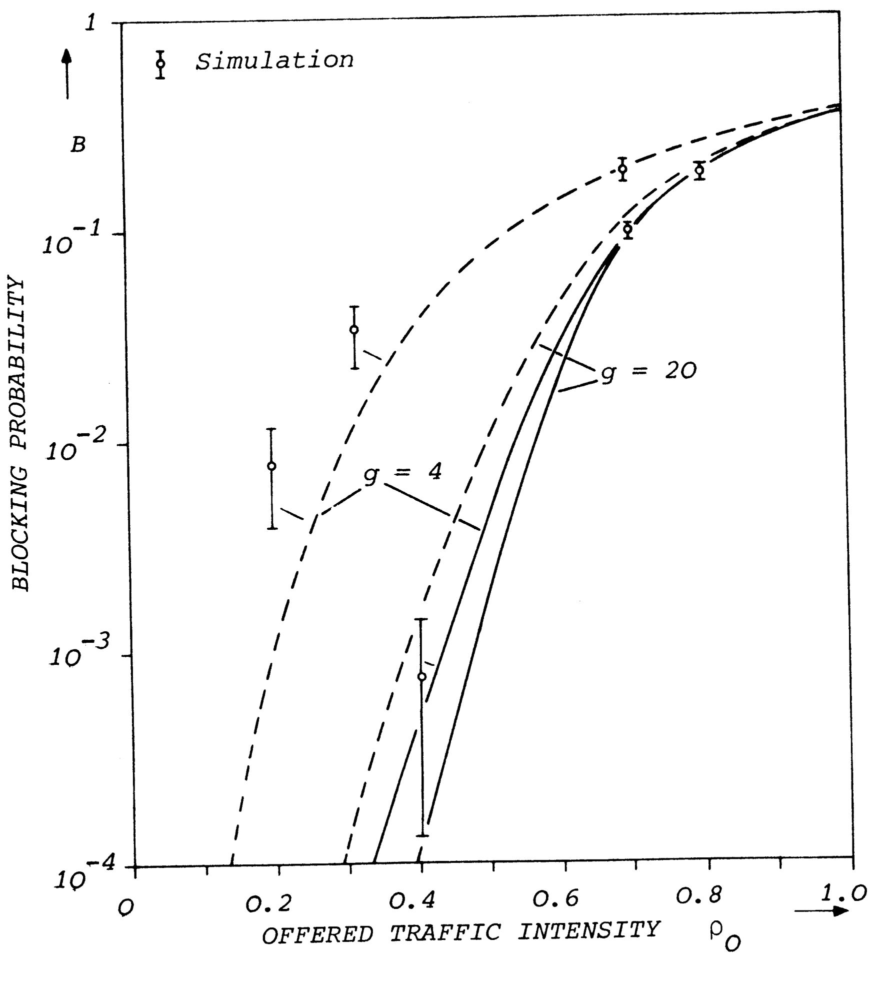

Fig. 5

Blocking probability vs offered traffic (symmetrical load)

Parameters:
$\bar{T}_{uj} = \bar{T}_u = 0.5 \cdot \bar{T}_h$; $\bar{T}_h = 1$
$S_j = 5$; $j = 1, \ldots, g$; $c_u = 0$
$c_h = 1$ ———
$c_h = 5$ - - - - -

The case of Markovian service time distribution function ($c_h=1$) can be calculated by both methods discussed in subsections 3.2 and 3.3. From the approximation accuracy point of view and for Markovian service time, there is no significant difference between results obtained using the direct approach and using the moment matching method. However, the direct method is more effective concerning the computing efforts and the number of iteration cycles required for convergence.

It can clearly be seen that the overall approximation accuracy for the given system parameters is good. However, the accuracy depends very strongly on the number of queues and the mean values of the switchover time. In general, the accuracy of the algorithm increases with increasing values of switchover time (c.f.[8]). It should be noted here that results delivered by the presented method always show the same tendencies and phenomena as they are obtained by computer simulations.

4.2 Systems with Nonsymmetrical Load Conditions

Nonsymmetrical load conditions exist in polling systems, in which overload occurs in a part of the system. This phenomenon can be observed, e.g., in systems with distributed control (switching systems, local area networks,etc.), in the case of a dramatic induction of an overload situation arising in a particular subsystem throughout the whole system.

The example in this subsection considers a multiqueue system with g=4, in which the first two queues are overloaded with the traffic intensity $\rho_1 = \rho_2 = \lambda_1 \bar{T}_{h1} = \lambda_2 \bar{T}_{h2}$. Symmetrical conditions are assumed for the remainder queues.

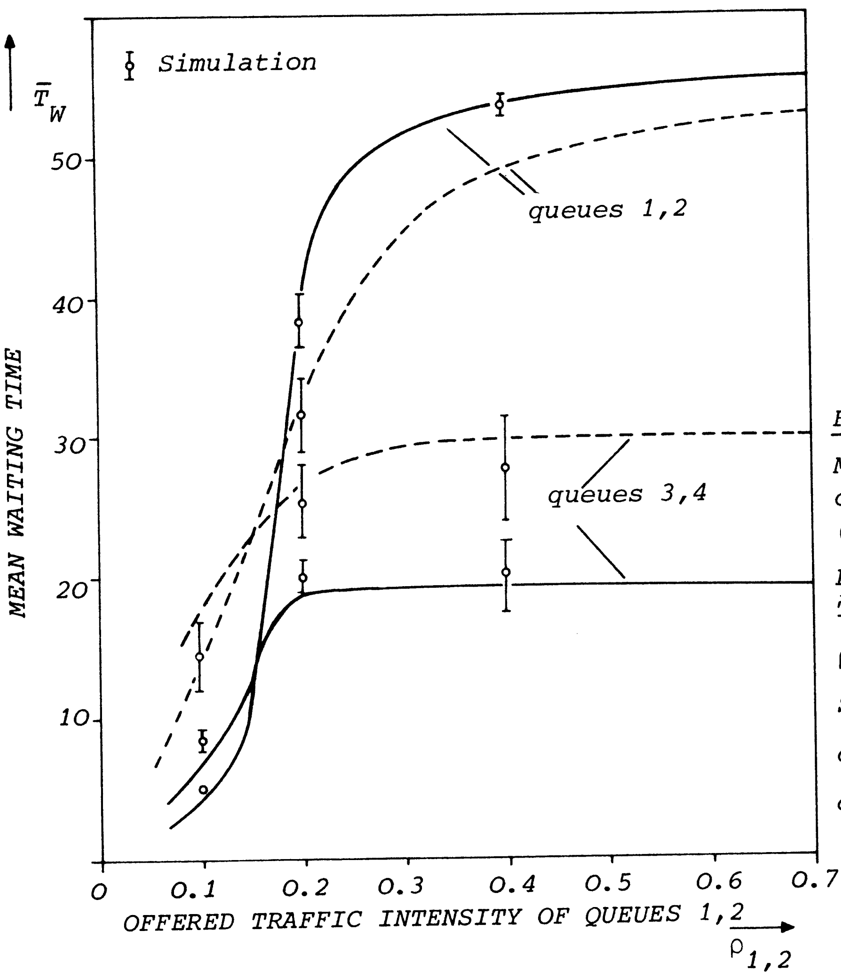

Fig. 6

Mean waiting time vs offered traffic of queues 1,2 (nonsymmetrical load)

Parameters:

$\bar{T}_{uj} = \bar{T}_u = 0.5 \cdot \bar{T}_h$; $\bar{T}_h = 1$

$\rho_{3,4} = 0.15$; $c_u = 0$;

$S_j = 10$; $j=1,\ldots,g$; $g = 4$

$c_h = 1$ ————

$c_h = 5$ - - - - -

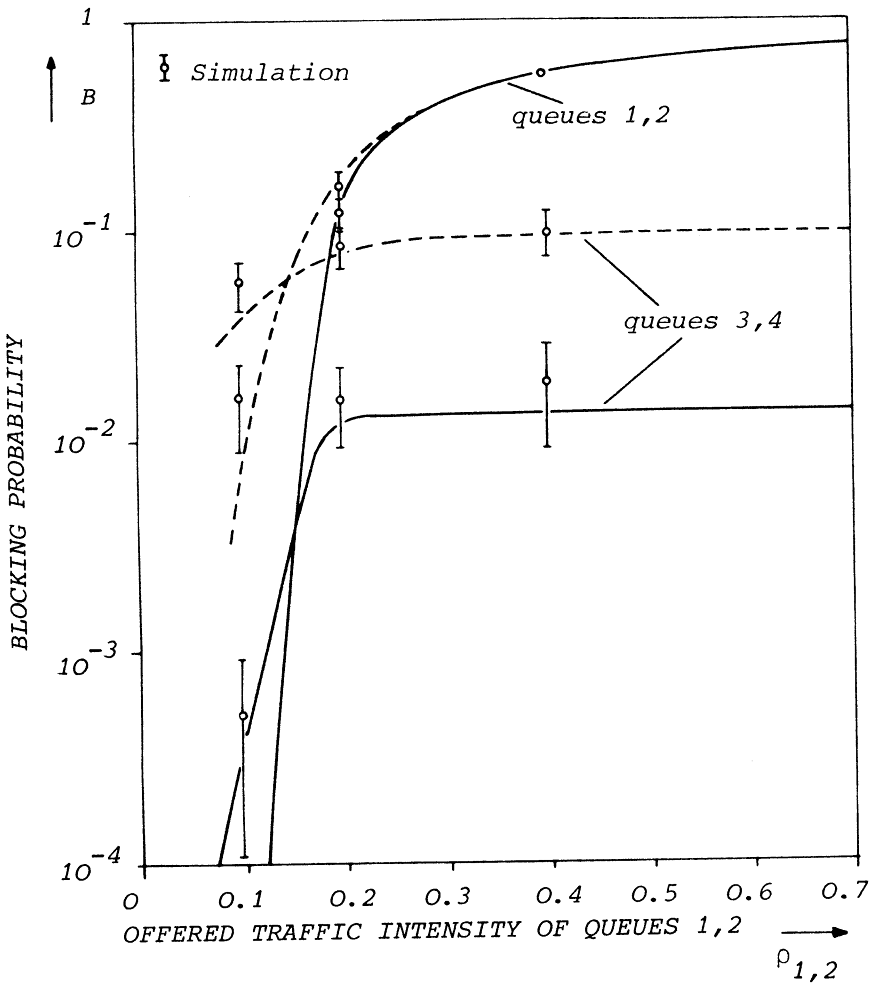

Fig. 7

Blocking probability vs offered traffic of queues 1,2 (nonsymmetrical load)

Parameters:
$\bar{T}_{uj} = \bar{T}_u = 0.5 \cdot \bar{T}_h$; $\bar{T}_h = 1$
$\rho_{3,4} = 0.15$; $c_u = 0$;
$S_j = 10$; $j=1,\dots,g$; $g=4$
$c_h = 1$ ———
$c_h = 5$ - - - - -

Figs. 6 and 7 show the influence of the overload in queues 1 and 2 on the delay and blocking characteristics of queues 3 and 4, for different values of service time coefficient of variation. It can be clearly seen that the blocking probability and the mean delay in queues 3 and 4 increase rapidly with increasing traffic intensity until a certain level ($\rho_{1,2} = 0.2$). Above this level, according to the finite queue capacity and the blocking effects in the system, the influence of the local overload (queues 1 and 2) to the normally loaded system part (queues 3 and 4) is limited. For the system parameters discussed, the approximation accuracy is higher for smaller service time coefficients of variation.

5. CONCLUSIONS

In this paper, we presented an approximation technique for polling systems, where the realistic assumption of finite capacity has been considered. The numerical algorithm provided can be applied to performance investigations of a class of computer and communication systems, such as local area networks with token-ring protocols or stored program controlled switching systems with distributed control structures. The cycle time distribution function, which is required for the iteration scheme in the analysis, was given in terms of Erlangian distribution functions in the simple case of Markovian server or approximated using a two moment matching method for generally distributed service times. Results like mean cycle time, blocking probability, etc. for symmetrical as well as nonsymmetrical load conditions are discussed. The accuracy and convergence of the presented algorithm is good over a wide range of system parameters.

ACKNOWLEDGEMENT

The authors would like to express their thanks to Prof. P.J. Kuehn for reviewing the manuscript and to Mr. M. Weixler for his programming efforts.

REFERENCES

[1] R.B. Cooper and G. Murray, "Queues served in Cyclic Order", Bell Syst. Tech. J. 48 (1969) 675-689.

[2] R.B. Cooper, "Queues served in Cyclic Order: Waiting Times", Bell Syst. Tech. J. 49 (1970) 399-413.

[3] M. Eisenberg, "Two Queues with Changeover Times" , Oper. Res. 19(1971) 386-401.

[4] M. Eisenberg, "Queues with Periodic Service and Changeover Times", Oper. Res. 20 (1972) 440-451.

[5] O. Hashida, "Analysis of Multiqueue", Rev. El. Commun. Lab. 20(1972) 189-199.

[6] O. Hashida, K. Ohara, "Line Accomodation Capacity of a Communication Control Unit", Rev. El. Commun. Lab. 20 (1972) 231-239.

[7] M. A. Leibowitz, "An Approximate Method for Treating a Class of Multiqueue Problems", IBM J. Res. Develop. 5 (1961) 204-209.

[8] P.J. Kuehn, "Multiqueue Systems with Nonexhaustive Cyclic Service", Bell Syst. Tech. J. 58 (1979) 671-699.

[9] W. Bux, "Local-Area Subnetworks : A Performance Comparison", IEEE Trans. Comm. COM-29 (1981) 1465-1473.

[10] O. J. Boxma, "Two Symmetric Queues with Alternating Service and Switching Times", Proc. Performance '84, Paris, pp. 409-431.

[11] H. Takagi, L. Kleinrock, "Analysis of Polling Systems", Japan Science Inst. Research Report, 1985.

[12] R.J.T. Morris, Y.T. Wang, "Some Results for Multiqueue Systems with Multiple Cyclic Servers", Proc. 2nd Int. Symp.on the Performance of Comp. Comm. Systems, Zurich 1984, 245- 258.

[13] T. Raith, "Performance Analysis of Multibus Interconnection Networks in Distributed Systems", Proc. 11th Internat. Teletraffic Congress, Sept. 1985, Kyoto, Japan, paper 4.2A-5.

[14] P.J. Kuehn, "Approximate Analysis of General Queueing Networks by Decomposition", IEEE Trans. Comm. COM-27(1979) 113-126.

[15] P. Tran-Gia, T. Raith, "Multiqueue Systems with Finite Capacity and Non-exhaustive Cyclic Service", Proc. Internat. Seminar on Computer Networking and Performance Evaluation, Sept. 1985, Tokyo, Japan.