Performance Analysis of Finite Capacity Polling Systems with Nonexhaustive Service

Phuoc Tran-Gia *

Institute of Computer Science, University of Würzburg, Am Hubland, D-8700 Würzburg, Fed. Rep. Germany

Thomas Raith *

Research and Technology Division, Daimler-Benz AG, Stuttgart, Fed. Rep. Germany

In performance investigations of token passing local area networks, communication subsystems in switching systems with distributed control, etc., the class of polling models, i.e., multiqueue systems with cyclic service, is often employed. This paper presents an approximate analysis method for this class of models, 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 exponential 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 validation of the approximation is done 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.
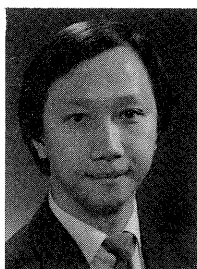
Phuoc Tran-Gia received the M.S. degree (Dipl.-Ing.) from the Stuttgart University, in 1977 and the Ph.D. degree (Dr.-Ing.) from the University of Siegen in 1982, both in Electrical Engineering. In 1977, he joined Standard Elektrik Lorenz (ITT), Stuttgart, where he worked in software development of digital switching systems. From 1979 until 1982 he worked as an assistant professor in the Department of Communications of the University of Siegen. From 1983 to 1986, Dr. Tran-Gia was head of a research group at the Institute of Communications Switching and Data Technics, Stuttgart University, Fed. Rep. Germany. His research activities were in the field of environment simulations for communication systems and applied queueing theory, especially with focus on discrete-time methods. In 1986, he joined the IBM Zürich Research Laboratory where he worked on the architecture and performance evaluation of computer communication systems. Since July 1988 he has been Professor for Computer Science (Distributed Systems) at the University of Würzburg, Fed. Rep. Germany.
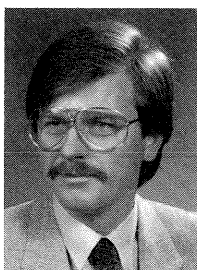
Thomas Raith received the M.Sc. degree (Dipl.-Ing.) and the Ph.D. degree (Dr.-Ing.) in Electrical Engineering from the University of Stuttgart, Fed. Rep. Germany, in 1981 and 1987, respectively. From 1982 to 1986 he was with the Institute of Communications Switching and Data Technics at the University of Stuttgart, working in the field of performance investigation of computer communication systems. In 1987 he joined the Research and Technology Division of Daimler-Benz AG in Stuttgart. He is currently involved in performance and system design of in-vehicle controller area networks for real-time applications.

Dr. Raith is member of the ITG (German Communication Society) and GI (German Informatics Society).

* This work was done while the authors were with the Institute of Communications Switching and Data Technics, University of Stuttgart, Fed. Rep. Germany.

1. Introduction

Polling models are often used in performance investigations of communication and computer systems, for example local area networks operating with token passing protocols or switching systems with distributed control structures. Most of the investigations, which employed polling models with varying degree of complexity, consider queueing systems having infinite capacities. Several approximation techniques for the system analysis are proposed.

Single server polling systems have been the subject of numerous studies in the literature [1–9,11,12,19,20]. A number of modelling approaches considering various polling mechanisms like cyclic or priority order [15] and several service disciplines, for example, 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.

Multiqueue systems with cyclic polling strategy, symmetrical load conditions, constant switchover time and gating service were approximately analyzed by Leibowitz [12]. In [3,4], 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 [5]. The approach of Cooper and Murray in [4] has been generalized by Eisenberg [6] and Hashida [7] to nonzero switchover time. An approximation technique for cyclic queues with nonexhaustive service and general switchover time has been developed by Kuehn [9]. Morris and Wang [16] and Raith [17] provided analytical approaches to deal with polling systems with multiple servers. Raith and Tran-Gia [18] considered the feedback effect of receiver blocking. An exact solution for a system with two queues and nonzero switchover time was presented by Boxma [1]. A survey on polling system analysis, where various system classes are considered, was provided by Takagi and Kleinrock [19]. In the context of finite capacity systems, Mack et al. [13,14] investigated the well-known 'repair man model', which corresponds to the case of a finite number of sources.

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 in transform domains by applying generating functions or Laplace transforms, do not lead to closed form expressions or effective algorithms for performance measures of interest.

In the following, an approximate analysis method for polling systems with finite queue capacity and nonexhaustive cyclic service (or, more precisely, limited service by one) is presented, whereby a numerical algorithm is developed (cf. [21,22]). According to the type of the service time distribution function (exponential or general), two numerical schemes for the cycle time calculation in conjunction with an iteration method are derived. The accuracy of the approximation method will be illustrated by means of numerical results, which have been obtained for a wide range of system parameters.

2. Model description

The basic model of a multiqueue system is illustrated in Fig. 1. The model consists of a number g of finite capacity queues. The queues are served nonexhaustively by a single server according to a limited service scheme. The service is organized in a cyclic order by a single server. The service time distribution function can be individually chosen for each queue. Also, individual arrival rates and sizes of the buffer capacity are considered. After the 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 and to scan the succeeding queue, is assumed to have a queue-dependent general distribution function. At the scanning epoch, i.e., at the end of the corresponding switchover time, the server will process one message
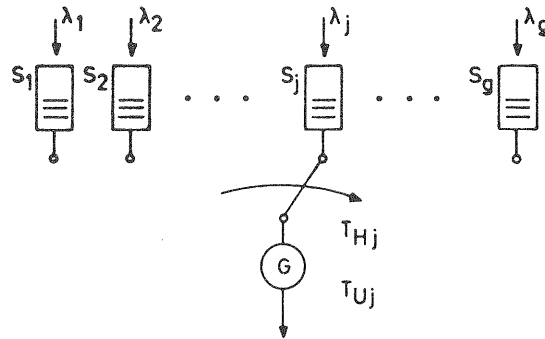
Fig. 1. The basic queueing model.

in the queue (limited service by one), if there is at least one message waiting for service. If the queue is empty, the observed inter-scan period consists of just the switchover time. The arrival processes are assumed to be Poisson with queue-individual rates, according to which the system can be modelled to be symmetrically or nonsymmetrically loaded.

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

g : number of queues in the system,
λ_j : arrival rate of messages offered to queue j,
$T_{H,j}$: r.v. for the service time of messages in queue j,
$T_{U,j}$: r.v. for the switchover time corresponding to queue j,
S_j : capacity of queue j.

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 [8,9], 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. Throughout this section, the approximate analysis is based on the assumption of independency between cycle times and cycle time segments. This assumption will be stated in Section 3.1.2.

In Section 3.1, general equations for the state analysis and system characteristics will be derived while, in Section 3.2, the special case of exponential server and, subsequently, the case of generally distributed service time will be investigated. Finally, Section 3.3 gives an outline of the calculation algorithm for the Markov chain state probabilities and performance measures of interest.

For a random variable T_X, the following notation will be used:

$F_X(t)$: probability distribution function,
$f_X(t)$: probability density function,
$\Phi_X(s)$: Laplace–Stieltjes transform of $F_X(t)$,
$E[T_X]$: mean value of T_X,
$\mathrm{Var}[T_X]$: variance of T_X,
c_X : coefficient of variation of T_X.

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 nth 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 nth scanning epoch. We define the Markov chain state probabilities

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

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

$$P_{k,j} = \Pr\{ X(0^-) = k \} = \lim_{n \to \infty} P_{k,j}^{(n)}, \quad k = 0, 1, \ldots, S_j. \tag{2}$$

For ease of reading, subscript j indicating the observed queue will be suppressed, for example, the notation P_k will be used instead of $P_{k,j}$. For the capacity of the queue j observed, the symbols S or S_j will be used interchangeably.

In order to calculate the transition probabilities of the Markov chain

$$p_{ik} = \Pr\{ X^{(n+1)}(0^-) = k \mid X^{(n)}(0^-) = i \}, \tag{3}$$

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

$$P_k^{(n)}(t) = \Pr\{ X^{(n)}(t) = k \}, \quad k = 0, 1, \ldots, S, \tag{4}$$

can be obtained as follows:

$$P_k^{(n)}(t) = P_0^{(n)} a_k(t) + \sum_{i=0}^{k} P_{i+1}^{(n)} a_{k-i}(t), \quad k = 0, 1, \ldots, S-1,$$

$$P_S^{(n)}(t) = P_0^{(n)} \sum_{i=S}^{\infty} a_i(t) + \sum_{i=0}^{S-1} P_{i+1}^{(n)} \sum_{m=S-i}^{\infty} a_m(t), \tag{5}$$

where

$$a_m(t) = \left[(\lambda_j t)^m / m! \right] e^{-\lambda_j t}. \tag{6}$$

Using the consideration of conditional cycle time [9], where the following r.v.'s are defined with respect to the observed queue j:

T_C : r.v. for the cycle time,

$T_{C'}$: 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 a nonempty 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

$$P_k^{(n+1)} = P_0^{(n)} \int_0^{\infty} a_k(t) f_{C'}(t) \, \mathrm{d}t + \sum_{i=1}^{k+1} P_i^{(n)} \int_0^{\infty} a_{k-i+1}(t) f_{C''}(t) \, \mathrm{d}t, \quad k = 0, 1, \ldots, S-1,$$

$$P_S^{(n+1)} = P_0^{(n)} \sum_{i=S}^{\infty} \int_0^{\infty} a_i(t) f_{C'}(t) \, \mathrm{d}t + \sum_{i=1}^{S} P_i^{(n)} \sum_{m=S-i+1}^{\infty} \int_0^{\infty} a_m(t) f_{C''}(t) \, \mathrm{d}t. \tag{7}$$

Defining the arrival probabilities, i.e., the probabilities for m arrivals during a conditional cycle of type $T_{C'}$ or $T_{C''}$,

$$b_m' = \int_0^{\infty} a_m(t) f_{C'}(t) \, \mathrm{d}t, \qquad b_m'' = \int_0^{\infty} a_m(t) f_{C''}(t) \, \mathrm{d}t, \tag{8}$$

we obtain from (7) the following set of Markov chain state equations:

$$P_k^{(n+1)} = P_0^{(n)}b_k' + \sum_{i=1}^{k+1} P_i^{(n)}b_{k-i+1}'', \quad k = 0, 1, \ldots, S-1,$$

$$P_S^{(n+1)} = P_0^{(n)} \sum_{i=S}^{\infty} b_i' + \sum_{i=1}^{S} P_i^{(n)} \sum_{m=S-i+1}^{\infty} b_m''.$$

(9)

Equation (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 (8), which is the subject of Section 3.2.

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)

$$\Phi_{E,j}(s) = \Phi_{U,j}(s) \cdot \left((1 - P_{0,j})\Phi_{H,j}(s) + P_{0,j}\right).$$

(10)

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

$$\Phi_{C',j}(s) = \Phi_{U,j}(s) \cdot \prod_{k=1,k\neq j}^{g} \Phi_{E,k}(s),$$

(11a)

$$\Phi_{C'',j}(s) = \Phi_{U,j}(s) \cdot \Phi_{H,j}(s) \cdot \sum_{k=1,k\neq j}^{g} \Phi_{E,k}(s).$$

(11b)

3.1.3. Arbitrary time state probabilities

In order to calculate system characteristics, for example blocking probability for messages or mean waiting time in the queue, it is useful to obtain first the arbitrary time state probabilities. Again, subscript j indicating the observed queue is suppressed in this section.

Define $\{P_k^*, k = 0, 1, \ldots, S\}$ to be the arbitrary time state probabilities, i.e., the distribution of the number X_j of messages in the considered queue j at an arbitrary observation instant. According to the two types of conditional cycle times we define π' and π'' to be, respectively, the probability of an outside observer to see a cycle of type $T_{C'}$ or $T_{C''}$. It can clearly be seen that

$$\pi' = P_0 \cdot E[T_{C'}]/E[T_C],$$

(12a)

where

$$E[T_C] = P_0 \cdot E[T_{C'}] + (1 - P_0) \cdot E[T_{C''}]$$

(12b)

and

$$\pi'' = 1 - \pi'.$$

(12c)

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

$$f_{C'}^{r}(t) = \frac{1 - F_{C'}(t)}{E[T_{C'}]}, \qquad f_{C''}^{r}(t) = \frac{1 - F_{C''}(t)}{E[T_{C''}]}.$$

(13)

The arrival probabilities during the backward recurrence time $T_{C'}^r$ and $T_{C''}^r$ can be given as

$$b_m'^* = \int_0^{\infty} a_m(t)f_{C'}^{r}(t)\,\mathrm{d}t, \qquad b_m''^* = \int_0^{\infty} a_m(t)f_{C''}^{r}(t)\,\mathrm{d}t.$$

(14)

Considering both types of conditional cycle times and combining the above results, the arbitrary time state probabilities can be written as follows:

$$P_k^* = \pi' b_k'^* + \pi'' \sum_{i=1}^{k+1} \frac{P_i}{1 - P_0} b_{k-i+1}''^*, \quad k = 0, 1, \ldots, S - 1,$$

$$P_S^* = \pi' \sum_{i=S}^{\infty} b_i'^* + \pi'' \sum_{i=1}^{S} \frac{P_i}{1 - P_0} \sum_{m=S-i+1}^{\infty} b_m''^*. \tag{15}$$

Finally, we obtain from (12) and (15)

$$P_k^* = \frac{E[T_{C'}]}{E[T_C]} P_0 b_k'^* + \frac{E[T_{C''}]}{E[T_C]} \sum_{i=1}^{k+1} P_i b_{k-i+1}''^*, \quad k = 0, \ldots, S - 1,$$

$$P_S^* = \frac{E[T_{C'}]}{E[T_C]} P_0 \sum_{i=S}^{\infty} b_i'^* + \frac{E[T_{C''}]}{E[T_C]} \sum_{i=1}^{S} P_i \sum_{m=S-i+1}^{\infty} b_m''^*. \tag{16}$$

3.1.4. System characteristics

With the arbitrary time state probabilities given from equations (16) and taking into account the Poisson arrivals offered to the observed queue j see time averages, the blocking probability for messages in queue j can be determined as follows:

$$B_j = P_S^*. \tag{17}$$

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

$$E[T_{W,j}] = E[X_j] / \{\lambda_j \cdot (1 - B_j)\}, \tag{18}$$

where

$$E[X_j] = \sum_{i=1}^{S} i \cdot P_i^*. \tag{19}$$

It should be noted here that the well-known formula for the mean cycle time [9] is obtained in a modified form for the case of finite queue capacity,

$$E[T_C] = \left[\sum_{j=1}^{g} E[T_{U,j}] \right] \Big/ \left[1 - \sum_{j=1}^{g} \lambda_j \cdot E[T_{H,j}] \cdot (1 - B_j) \right]. \tag{20}$$

Equation (20) can briefly be derived as follows, using mean values consideration (cf. [9]). Given the system in steady state, the average number of messages arriving at queue j during a cycle time (seen from an arbitrary queue) is $\lambda_j \cdot E[T_C]$. Using the property of Poisson arrivals see time average, the mean number of messages being accepted at queue j during a cycle is thus $\lambda_j \cdot E[T_C] \cdot (1 - B_j)$.

Since the system is stable, on average, the number of all accepted messages in the whole system during a cycle

$$\sum_{j=1}^{g} \lambda_j \cdot E[T_C] \cdot (1 - B_j)$$

must be served during a cycle time. Hence, the mean cycle time can be given as

$$E[T_C] = \sum_{j=1}^{g} E[T_{U,j}] + \sum_{j=1}^{g} E[T_C] \cdot \lambda_j \cdot (1 - B_j) \cdot E[T_{H,j}],$$

which yields equation (20).

3.2. Calculation of cycle time and arrival probabilities

In the following we shall start with the case of a symmetrical system with exponential service times. In this special case, efficient and simple expressions for the cycle time analysis can be obtained, without using the two-moment matching approximation for distribution functions. This approximation is used for general service times as will be described in Section 3.2.2.

3.2.1. The case of exponential server and constant switchover time

In this section we devote our attention to symmetrical polling systems with exponential service time and constant switchover time, i.e.,

$$\lambda_j = \lambda, \quad E[T_{H,j}] = E[T_H] = \frac{1}{\mu}, \quad \Phi_{H,j}(s) = \Phi_H(s) = \frac{\mu}{s + \mu}, \qquad j = 1, 2, \ldots, g.$$

Under these assumptions, the conditional cycle times can be given as follows (cf. equations (10) and (11)):

$$\begin{aligned}
\Phi_{C'}(s) &= e^{-st_0} \left(P_0 + (1 - P_0) \Phi_H(s) \right)^{g-1}, \\
\Phi_{C''}(s) &= e^{-st_0} \Phi_H(s) \left(P_0 + (1 - P_0) \Phi_H(s) \right)^{g-1},
\end{aligned} \tag{21}$$

where $t_0 = gE[T_U]$.

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

$$\begin{aligned}
f_{C'}(t) &= \sum_{i=0}^{g-1} \binom{g-1}{i} P_0^{g-1-i} (1 - P_0)^i e_i(t - t_0), \\
f_{C''}(t) &= \sum_{i=0}^{g-1} \binom{g-1}{i} P_0^{g-1-i} (1 - P_0)^i e_{i+1}(t - t_0),
\end{aligned} \tag{22}$$

where $e_i(t)$ denotes the Erlangian probability density function of ith order. With equation (22), the arrival probabilities given in equation (8) can be calculated in a straightforward manner as

$$\begin{aligned}
b_m' &= P_0^{g-1} a_m(t_0) + \sum_{i=1}^{g-1} \binom{g-1}{i} P_0^{g-1-i} (1 - P_0)^i \sum_{j=0}^{m} \binom{j+i-1}{j} q^j (1 - q)^i a_{m-j}(t_0), \\
b_m'' &= \sum_{i=0}^{g-1} \binom{g-1}{i} P_0^{g-1-i} (1 - P_0)^i \sum_{j=0}^{m} \binom{j+i}{j} q^j (1 - q)^{i+1} a_{m-j}(t_0),
\end{aligned} \tag{23}$$

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

Based on these arrival probabilities, the Markov chain state probabilities can be calculated according to equation (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 (cf. equations (13))

$$\begin{aligned}
f_{C'}^r(t) &= \frac{1}{E[T_{C'}]} \int_{\tau=t}^{\infty} \sum_{i=0}^{g-1} \binom{g-1}{i} P_0^{g-1-i} (1 - P_0)^i e_i(\tau - t_0) \, d\tau, \\
f_{C''}^r(t) &= \frac{1}{E[T_{C''}]} \int_{\tau=t}^{\infty} \sum_{i=0}^{g-1} \binom{g-1}{i} P_0^{g-1-i} (1 - P_0)^i e_{i+1}(\tau - t_0) \, d\tau.
\end{aligned} \tag{24}$$

Using equations (24), the probability of m arrivals during the backward recurrence conditional cycle times as stated in equations (14) can be derived as

$$b'^*_m = \frac{1}{E[T_{C'}]} \left\{ P_0^{g-1} \sum_{j=m+1}^{\infty} a_j(t_0) + \sum_{i=1}^{g-1} \binom{g-1}{i} P_0^{g-1-i}(1-P_0)^i \right.$$

$$\left. \times \left(\sum_{j=0}^{m} a_{m-j}(t_0) \sum_{k=0}^{i-1} \binom{j+k}{j}(1-q)^k q^{j+1} + \sum_{j=m+1}^{\infty} a_j(t_0) \right) \right\}, \qquad (25a)$$

$$b''^*_m = \frac{1}{E[T_{C''}]} \left\{ \sum_{i=0}^{g-1} \binom{g-1}{i} P_0^{g-1-i}(1-P_0)^i \right.$$

$$\left. \times \left(\sum_{j=0}^{m} a_{m-j}(t_0) \sum_{k=0}^{i} \binom{j+k}{j}(1-q)^k q^{j+1} + \sum_{j=m+1}^{\infty} a_j(t_0) \right) \right\}. \qquad (25b)$$

Thus, the arbitrary time state probabilities of a considered queue can be calculated according to equations (16), in order to obtain system characteristics.

3.2.2. The case of general service time

In this section, 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 (cf. [21,22]).

Since the expressions of the conditional cycle times (equations (11a) and (11b)) 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 equation (7) or (8). However, for reasons of computing efforts, the two-moment approximation technique, as proposed in [10], will be used in this section. This will briefly be described in the following.

Equations (10), (11a) and (11b) yield the first two moments of the conditional cycle times, thus

$$E[T_{C',j}] = E[T_{U,j}] + \sum_{k=1,k\neq j}^{g} E[T_{E,k}],$$

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (26)$$

$$E[T_{C'',j}] = E[T_{U,j}] + E[T_{H,j}] + \sum_{k=1,k\neq j}^{g} E[T_{E,k}],$$

where

$$E[T_{E,j}] = E[T_{U,j}] + (1-P_{0,j}) \cdot E[T_{H,j}],$$

as well as

$$\text{Var}[T_{C',j}] = \text{Var}[T_{U,j}] + \sum_{k=1,k\neq j}^{g} \text{Var}[T_{E,k}],$$

$$\text{Var}[T_{C'',j}] = \text{Var}[T_{U,j}] + \text{Var}[T_{H,j}] + \sum_{k=1,k\neq j}^{g} \text{Var}[T_{E,k}], \qquad (27)$$

$$\text{Var}[T_{E,j}] = \text{Var}[T_{U,j}] + (1-P_{0,j}) \cdot \text{Var}[T_{H,j}] + P_{0,j} \cdot (1-P_{0,j}) \cdot E[T_{H,j}]^2.$$

According to the moment matching method [10] mentioned above, a given random variable T with the

expected value $E[T]$ and the coefficient of variation c is approximately described by means of the following substitute distribution function $F(t)$:

Case 1: $0 \leqslant c \leqslant 1$

$$F(t) = \begin{cases} 0 & 0 \leqslant t \leqslant t_1, \\ 1 - e^{-(t-t_1)/t_2}, & t \geqslant t_1, \end{cases} \tag{28a}$$

where $t_1 = E[T] \cdot (1 - c)$ and $t_2 = E[T] \cdot c$.

Case 2: $c \geqslant 1$

$$F(t) = 1 - p\, e^{-t/t_1} - (1 - p)\, e^{-t/t_2}, \tag{28b}$$

where

$$t_{1,2} = E[T] \cdot \left(1 \pm \sqrt{\frac{c^2 - 1}{c^2 + 1}} \right)^{-1} \quad \text{and} \quad p = E[T]/2t_1, \quad pt_1 = (1 - p)t_2.$$

The method is applied to the conditional cycle times $T_{C'}$ and $T_{C''}$, which are derived in (11a), (11b) in conjunction with their means and variances in (26) and (27). We obtain subsequently the arrival probabilities as follows, where $\{b_m\}$ are given for $\{b_m'\}$ and $\{b_m''\}$.

Case 1: $0 \leqslant c \leqslant 1$

$$b_m = \int_0^\infty a_m(t) f_C(t)\, dt = \frac{(\lambda t_2)^m}{(1 + \lambda t_2)^{m+1}}\, e^{-\lambda t_1} \sum_{k=0}^m \frac{((t_1/t_2)(\lambda t_2 + 1))^k}{k!}. \tag{29a}$$

Case 2: $c \geqslant 1$

$$b_m = p\frac{(\lambda t_1)^m}{(1 + \lambda t_1)^{m+1}} + (1 - p)\frac{(\lambda t_2)^m}{(1 + \lambda t_2)^{m+1}}. \tag{29b}$$

Based on the arrival probabilities (29a) and (29b) and the state equations (9), the Markov chain state probabilities are calculated. Analogously, the arrival probabilities during the backward recurrence conditional cycle times given by equations (14) can explicitly be written as follows.

Case 1: $0 \leqslant c \leqslant 1$

$$b_m^* = \frac{1}{\lambda(t_1 + t_2)}\left(1 - \sum_{k=0}^m \frac{(\lambda t_1)^k}{k!}\, e^{-\lambda t_1} \right) + \frac{t_2}{t_1 + t_2}b_m, \tag{30a}$$

where b_m corresponds to equation (29a).

Case 2: $c \geqslant 1$

$$b_m^* = \frac{(\lambda t_1)^m}{2(1 + \lambda t_1)^{m+1}} + \frac{(\lambda t_2)^m}{2(1 + \lambda t_2)^{m+1}}. \tag{30b}$$

3.3. Calculation algorithm for Markov chain state probabilities

In this section, a numerical algorithm is given to calculate the Markov chain state probabilities and the system characteristics. The algorithm, which utilizes the expressions derived for the Markov chain state probabilities and the conditional cycle time, is an alternating iteration scheme, which can be formulated as follows:

Algorithm

I. Initialization of Markov chain probabilities for all queues and conditional cycle times.

II. **repeat** {iteration cycle}

for all queues do

begin

   i. In the case of general service distribution, calculation of the mean and the variance for the approximate conditional cycle time $T_{C'}$ and $T_{C''}$ according to equations (26) and (27).

   ii. In the case of general service distribution, calculation of the parameters for the substitute conditional cycle times according to equations (28a) and (28b), depending on the range of the coefficient of variation.

   iii. Calculation of the arrival probabilities $\{b'_m\}$ and $\{b''_m\}$ according to equation (23) for the case of exponential server and constant switchover time or to equations (29a) and (29b) for the case of general service time distribution.

   iv. Calculation of the Markov chain state probabilities $\{P_{k,j}\}$ for the current iteration cycle according to equation (9).

   v. Update $E[T_{E,j}]$ and $\text{Var}[T_{E,j}]$ for the current iteration cycle according to equations (26) and (27).

   vi. Calculation of the current convergence indicator Δ_j defined below in equation (31).

end

until $\sum_{j=1}^{g} \Delta_j < \varepsilon$ (e.g., $\varepsilon = 10^{-6}$).

The convergence indicator Δ_j for the iteration is defined from

$$\Delta_j = \text{abs}\left[\sum_{k=1}^{S_j} k \cdot P_{k,j}^{(n)} - \sum_{k=1}^{S_j} k \cdot P_{k,j}^{(n-1)} \right]. \tag{31}$$

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

III. Determine system characteristics

   i. Calculate the mean and the variance for the backward recurrence time $T_{C'}^{r}$ and $T_{C''}^{r}$ according to equations (13).

   ii. Calculate $\{b'^*_m\}$ and $\{b''^*_m\}$ according to equations (14).

   iii. Calculate the arbitrary time state probabilities $\{P_k^*\}$ according to equation (16).

   iv. Calculate system characteristics, such as blocking probabilities, mean waiting time etc. according to equations (17)–(20).

4. Numerical results

Numerical results will be presented and discussed in the following for the two cases of symmetrically and nonsymmetrically loaded polling systems, in order to illustrate the performance behavior. The time variables are standardized by $T_{H,j} = 1$, $j = 1, 2, \ldots, g$, and the switchover time is chosen to be constant. The approximation is validated by means of computer simulations. The simulation results are depicted with their 95 percent confidence intervals, calculated using the Student-t test technique.

4.1. Symmetrical systems

In this section, two symmetric systems with $g = 8$ and $g = 32$ are taken into account. The mean waiting time and the blocking probability for messages are shown as functions of the offered traffic intensity

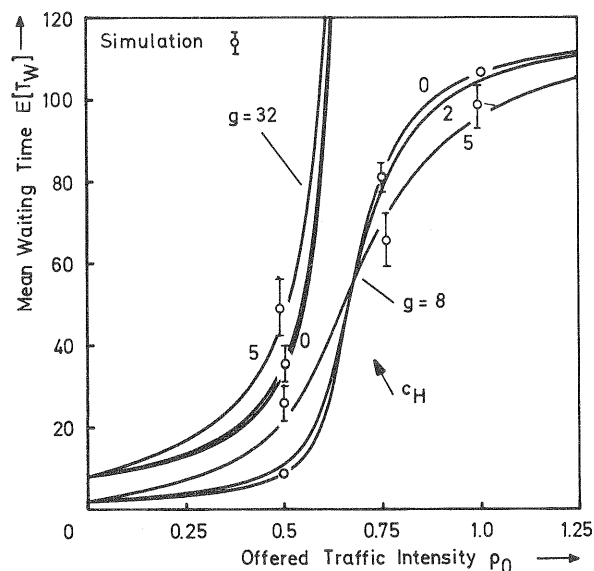$$\rho_0 = \sum_{j=1}^{g} \lambda_j \cdot E\left[T_{H,j}\right] \tag{32}$$

Fig. 2. Mean waiting time vs. offered traffic intensity. Parameters: $E[T_{U,j}] = E[T_U] = 0.5E[T_H]$, $S_j = 10$, $j = 1, 2, \ldots, g$.

in Figs. 2 and 3, respectively, for different server coefficients of variation. In Fig. 2, a crossover effect of the waiting time characteristics ($g = 8$) can be recognized, due to the finite number of waiting places. This effect can be explained using the fact that at higher load the blocking probability increases and the effective arrival rate of accepted messages decreases with higher server coefficient of variation.

Figs. 4 and 5 depict the mean and the coefficient of variation of the cycle time as function of the offered traffic intensity. Fig. 4 exhibits the effect that servers with higher variance lead to shorter mean cycle times in the range of traffic intensity (0.5, 0.75). It can also be seen in Fig. 4 that for higher traffic intensities the mean cycle time approaches a maximum value (the maximum cycle), which is given by the sum of switchover times and average service times for all queues. On the other hand, for lower traffic levels, an
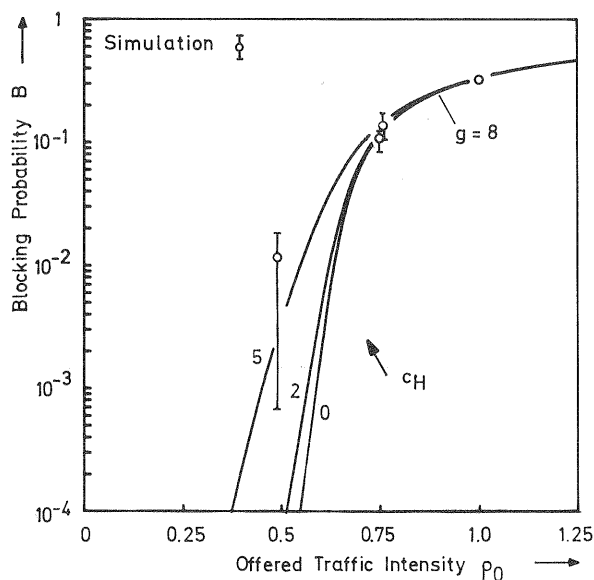

Fig. 3. Blocking probability vs. offered traffic intensity. Parameters: $E[T_{U,j}] = E[T_U] = 0.5E[T_H]$, $S_j = 10$, $j = 1, 2, \ldots, g$.
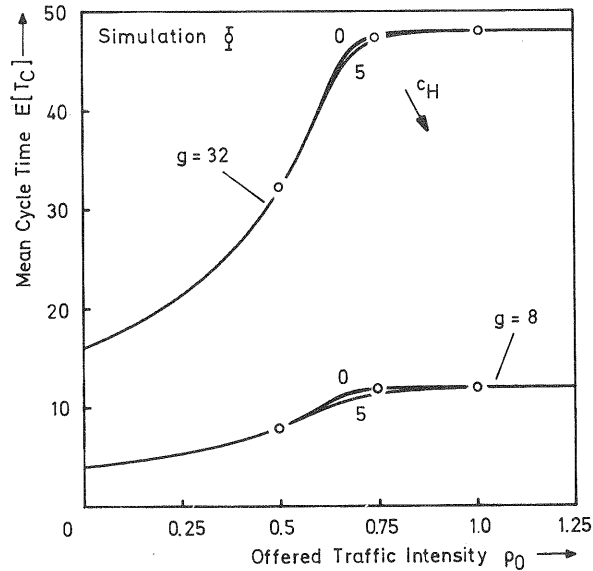
Fig. 4. Mean cycle time vs. offered traffic intensity. Parameters: $E[T_{U,j}] = E[T_U] = 0.5E[T_H]$, $S_j = 10$, $j = 1, 2, \ldots, g$.

empty cycle is obtained, corresponding to the sum of all switchover times. As expected, for disappearing message traffic, the cycle time coefficient of variation starts at zero, due to the chosen constant switchover time. Under overload conditions, where the maximum cycle takes place, the cycle time coefficient of variation approaches a limiting value, which is independent of the traffic level. As depicted in Fig. 5, there exists a maximum value for the cycle time coefficient of variation. This maximum increases by increasing service time coefficient of variation or decreasing means switchover time.

The influence of the finite queue capacity S on the mean waiting time is illustrated in Fig. 6 and on the blocking probability in Fig. 7, for different traffic intensities. Due to the finite queue capacity, the waiting time for accepted messages is limited (Fig. 6). However, for dimensioning purposes, the increasing blocking probability has to be taken into account.
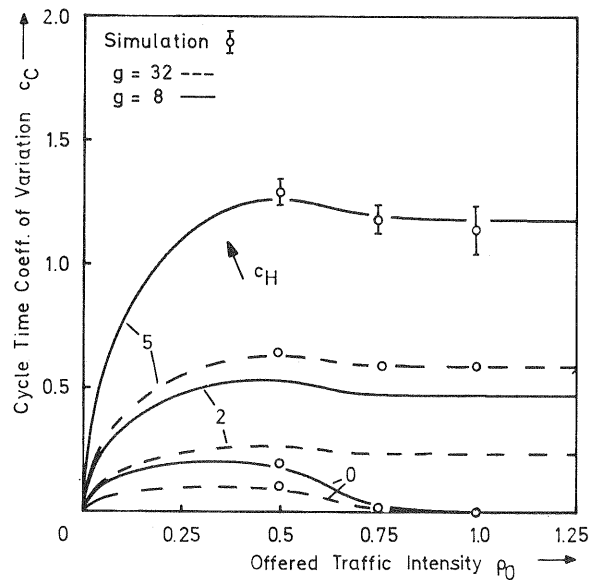
Fig. 5. Cycle time coefficient of variation vs. offered traffic intensity. Parameters: $E[T_{U,j}] = E[T_U] = 0.5E[T_H]$, $S_j = 10$, $j = 1, 2, \ldots, g$.
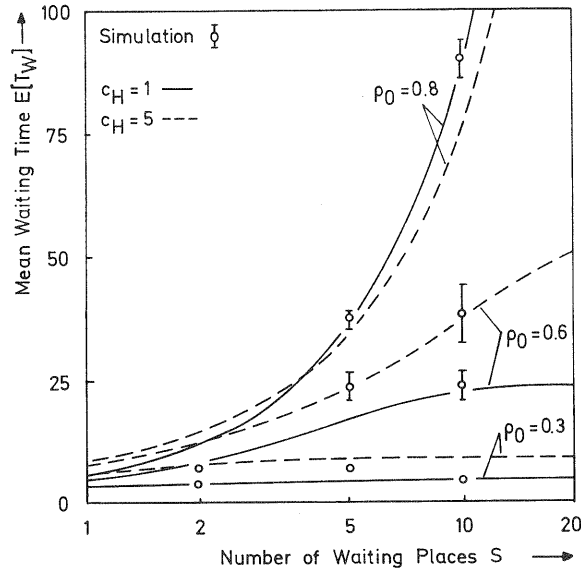
Fig. 6. Mean waiting time vs. number of waiting places. Parameters: $E[T_{U,j}] = E[T_U] = 0.5E[T_H]$, $j = 1, 2, \ldots, g$, $g = 8$.

The case of exponential service time distribution function ($C_H = 1$) can be calculated by both methods discussed in Sections 3.2.1 and 3.2.2. From the approximation accuracy point of view, 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 until convergence.

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, for example in systems with distributed control structures
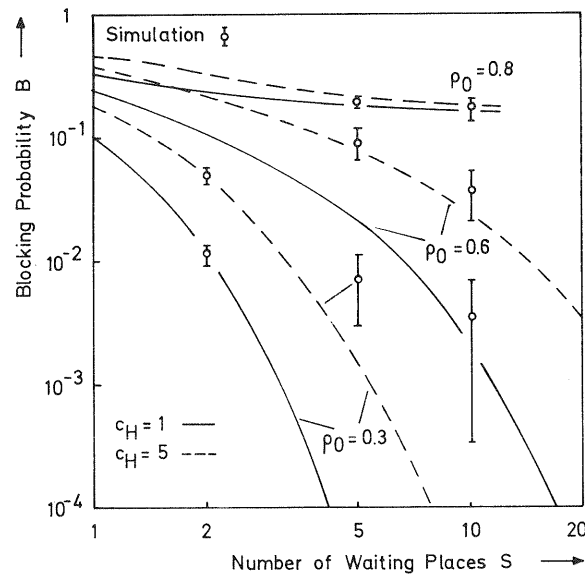

Fig. 7. Blocking probability versus number of waiting places. Parameters: $E[T_{U,j}] = E[T_U] = 0.5E[T_H]$, $j = 1, 2, \ldots, g$, $g = 8$.
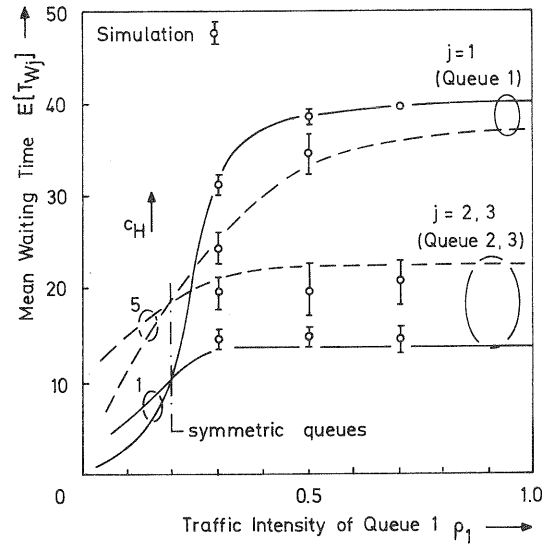
Fig. 8. Mean waiting time vs. offered traffic intensity of queue 1. Parameters: $E[T_{U,j}] = E[T_U] = 0.5E[T_H]$, $S_j = 10$, $j = 1, 2, \ldots, g$, $\rho_2 = \rho_3 = 0.2$.

(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 section considers a polling system with $g = 3$, in which the first queue is considered to be overloaded with the traffic intensity $\rho_1 = \lambda_1 \cdot E[T_{H,1}]$. Symmetrical conditions are assumed for the remainder queues.

Figs. 8 and 9 show the influence of the overload in queue 1 to the mean waiting times and blocking probabilities in queue 2 and 3, for different values of service time coefficient of variation. It can clearly be seen that the mean waiting time in queues 2 and 3 increases rapidly with increasing traffic intensity until a certain level ($\rho_1 = 0.35$). Above this level, due to the blocking effects in the system (Fig. 9), the influence of the local overload situation in queue 1 to other queues in the system is limited.
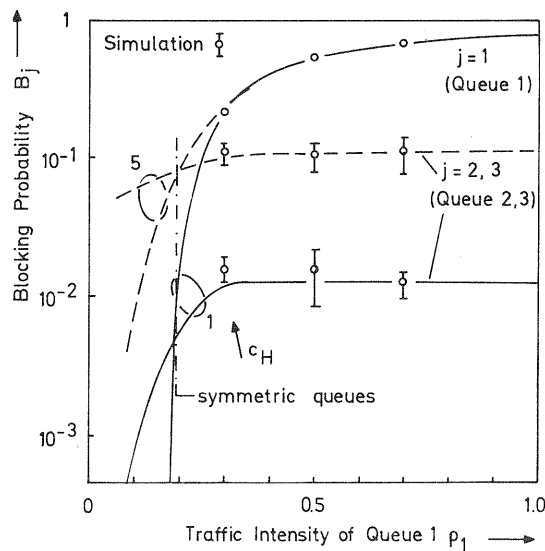

Fig. 9. Blocking probability vs. offered traffic intensity of queue 1. Parameters: $E[T_{U,j}] = E[T_U] = 0.5E[T_H]$, $S_j = 10$, $j = 1, 2, \ldots, g$, $\rho_2 = \rho_3 = 0.2$.

4.3. Approximation accuracy

The overall approximation accuracy for the given system parameters is sufficient for system engineering application. However, as expected, the accuracy strongly depends on system parameters, especially on the number of queues and the mean switchover time. In general, the following tendencies are observed: the approximation accuracy of the algorithm increases with increasing values of switchover time, increasing number of queue (cf. [9]) and smaller values of the service time coefficient of variation. The approximation is also less accurate for extreme nonsymmetrical system configurations and offered traffic intensity patterns. Concerning the essential new model component in this investigation, the assumption of finite queue capacity, it is observed that the approximation is more accurate for smaller queue capacities (e.g., $1 \leqslant S_j \leqslant 10$).

Further parameter studies show that the approximation accuracy is satisfactory when the product of the mean switchover time and the number of queues is larger than the mean service time. It should be noted here that the results delivered by the presented method always show the same tendencies and phenomena as they are obtained by computer simulations.

5. Conclusions

In this paper, an approximate analysis approach for polling systems is presented, where the realistic assumption of finite buffer capacity has been considered. The algorithm can be applied to performance investigations of a class of computer and communication systems modelled by multiqueue systems with ordinary cyclic service and switchover time, such as local area networks with token passing 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 exponential server or approximated using a two moment matching method for generally distributed service times. System characteristics of interest for symmetrical as well as nonsymmetrical load conditions are discussed. It is shown that the accuracy and the convergence behavior of the algorithm is good over a wide range of system parameters.

Acknowledgment

References

[1] O.J. Boxma, Two symmetric queues with alternating service and switching times, *Proc. Performance '84*, Paris (1984) 409–431.

[2] W. Bux and H.L. Truong, Mean-delay approximation for cyclic-service queueing systems, *Performance Evaluation* 3 (1981) 187–196.

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

[4] R.B. Cooper and G. Murray, Queues served in cyclic order, *Bell Syst. Tech. J.* 48 (1969) 675–689.

[5] M. Eisenberg, Two queues with changeover times, *Oper. Res.* 19 (1971) 3386–401.

[6] M. Eisenberg, Queues with periodic service and change-over times, *Oper. Res.* 20 (1972) 440–451.

[7] O. Hashida, Analysis of multiqueue, *Rev. El. Commun. Lab.* 20 (1972) 189–199.

[8] O. Hashida and K. Ohara, Line accomodation capacity of a communication control unit, *Rev. El. Commun. Lab.* 20 (1972) 231–239.

[9] P.J. Kuehn, Multiqueue systems with nonexhaustive cyclic service, *Bell. Syst. Tech. J.* 58 (1979) 671–699.

[10] P.J. Kuehn, Approximate analysis of general queueing networks by decomposition, *IEEE Trans. Comm.* COM-27 (1979) 113–126.

[11] P.J. Kuehn, Performance of ARQ-protocols for HDX-transmission in hierarchical polling systems, *Performance Evaluation* 1 (1981) 19–30.

[12] M.A. Leibowitz, An approximate method for treating a class of multiqueue problems, *IBM J. Res. Develop.* 5 (1961) 204–209.

[13] C. Mack, The efficiency of N machines uni-directionally patrolled by one operative when walking time is constant and repair times are variable, *J. Royal Stat. Soc., Ser. B* **19** (1) (1957) 173–178.

[14] C. Mack, T. Murphy and N.L. Webb, The efficiency of N machines uni-directionally patrolled by one operative when walking time and repair times are constants, *J. Royal. Stat. Soc., Ser. B* **19** (1) (1957) 166–172.

[15] D.R. Manfield, Analysis of a polling system with priorities, *IEEE Trans. Comm.* **COM-33** (1985) 1001–1006.

[16] R.J.T. Morris and Y.T. Wang, Some results for multi-queue systems with multiple cyclic servers, *Proc. 2nd Internat. Symp. on the Performance of Computer Communications Systems*, Zürich (1984) 245–258.

[17] T. Raith, Performance analysis of multibus interconnection networks in distributed systems, *Proc. 11th Internat. Teletraffic Congress* Kyoto, Japan (September 1985) paper 4.2A-5.

[18] T. Raith and P. Tran-Gia, Performance analysis of polling mechanisms with receiver blocking, *Proc. 8th Internat. Conf. on Computer Communications (ICCC)*, Munich (September 1986) C5-1.

[19] H. Takagi and L. Kleinrock, *Analysis of Polling Systems* (MIT Press, 1986) (Japan Science Inst. Research Rept., 1985).

[20] H. Takagi, On the analysis of a symmetric polling system with single message buffers, *Performance Evaluation* **5** (1985) 149–157.

[21] P. Tran-Gia and T. Raith, Multiqueue systems with finite capacity and nonexhaustive cyclic service, *Proc. Internat. Seminar on Computer Networking and Performance Evaluation*, Tokyo, Japan (North-Holland, Amsterdam, 1985).

[22] P. Tran-Gia and T. Raith, Approximation for finite capacity multiqueue systems, *Proc. GI / NTG Conf. on Measurement, Modelling and Evaluation of Computer Systems*, Dortmund, FRG (October 1985) 332–345.