13th International Teletraffic Congress, Copenhagen, June 19 - 26, 1991

Performance Analysis of Finite Capacity Polling Systems with Limited–M Service

Martin Lang, Martin Bosch

University of Stuttgart
Institute of Communications Switching and Data Technics
Seidenstraße 36, D–7000 Stuttgart 1, Federal Republic of Germany

Polling systems have a broad spectrum of applications, e.g. in modelling approaches for switching systems and Local Area Networks with media access mechanisms based on tokens. This contribution presents an approximate analysis for polling systems with the service discipline 'limited–M', whereby the realistic assumption of finite buffer capacity is taken into account. The analysis is an extension to the method presented by Tran-Gia and Raith in [15] where a finite capacity polling system with 'limited-1' service discipline has been investigated. Our analysis is based on an embedded Markov chain in conjunction with a two-moment approximation of server vacation and cycle times. The validation of the approximation is done by computer simulations.

1 Introduction

Local Area Networks (LANs) with token passing protocols are frequently modelled by polling systems with cyclic service. Most of these modelling approaches consider queues with infinite capacities. In the literature multiqueue systems served by a single server have been subject of numerous investigations [1] – [9], [11, 12, 15]. Various polling strategies like cyclic or priority service and different types of service disciplines, e.g. exhaustive, gated, or limited service have been considered. Most of these investigations assume queues with infinite capacities [1] – [7], [9, 12], whereas in practical systems the buffer space is limited.

In [9] an approximation technique for cyclic queues with non–exhaustive service and general switchover time has been developed by Kühn. He introduced conditional cycle times depending on whether the considered queue is empty or not. The queue length has been assumed to be infinite.

The analysis presented by Tran-Gia and Raith in[15] is based on these conditional cycle times. However, some modifications have been necessary to take the blocking effects and the finite queue capacities into account. The analysis uses the technique of the embedded Markov chain. Each queue is observed just prior to its polling instant. The state probabilities of an arbitrary queue immediately before its polling instant have been calculated from the state probabilities of the same queue at the instant just prior to the poll of the preceding cycle. The cycle times have been approximated by their first two moments. For the calculation of the stable state probabilities of all queues an iterative algorithm has been used.

Lee presented in [13] a vacation model and its analysis for finite queue capacity and exhaustive service discipline. This analysis has been extended in [14] to the limited service discipline. The vacation model can be applied to analyze a polling system. In this context the vacation time represents the interval between the instant the server has finished the service of a station and its arrival at the same station in the next cycle. To use this method, the problem to determine the distribution function of the vacation time remains to be solved.

In this paper we present a method to solve this problem. It is based on the algorithm presented in [15]. The resulting theory allows the analysis of a finite capacity polling system with 'limited–M' service. Symmetrical as well as nonsymmetrical load conditions are allowed. Again general switchover times are taken into account. In Section 2 we describe the model of the polling system. The analysis of this model is presented in Section 3. In Section 4 we show some numerical results validated by computer simulations. Finally, we make some concluding remarks.

2 Modelling

The basic model of a polling system is depicted in Figure 1. It consists of z stations which are served in cyclic

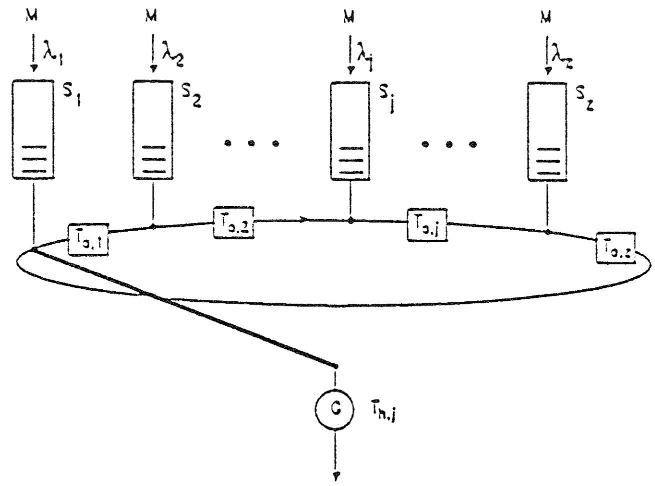

Figure 1: Basic Model of the Polling System

order by a single server with a generally distributed service time $T_{h,j}$ for each station. Every station consists of a queue with S_j buffer places. In each cycle every queue is served until it is either empty or a maximum number of M packets is reached. Then the server polls the succeeding queue. Due to these assumptions the considered service discipline is of the E–limited type (limited from the exhaustive point of view) [3, 5]. A queue individual, generally distributed switchover time $T_{o,j}$ can be selected. The interval from the end of the last service phase of a queue until the beginning of the first service phase of the same queue in the next cycle corresponds to the vacation time $T_{v,j}$ for that queue in a related vacation server model. The arrival processes are assumed to be Poissonian with queue–specific rates λ_j.

3 Performance Analysis

In the following section a numerical algorithm for an approximate analysis of the finite queue polling system is presented. The algorithm is based on the approaches presented in [9, 14, 15].

3.1 Markov Chain State Probabilities

The polling system depicted in Figure 1 can be analyzed by an embedded Markov chain. Each queue is observed just prior to its polling instant, i.e. immediately before the end of its vacation time, and at the instants immediately after each service completion. In Figure 2 the phase model for an arbitrarily chosen station j is shown. The embedded points are marked. At every embedded point except the last, i.e. after the M-th service phase, the server has to decide whether another packet can be served at this queue or not. Since the service discipline is of the E–limited type a new service phase is then started

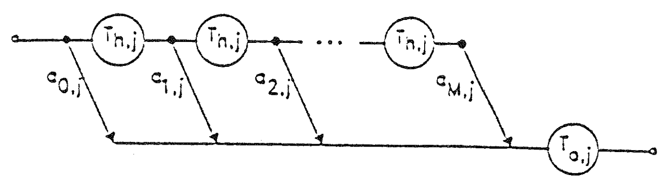

Figure 2: Phase Model for a Station j

if at least one packet is waiting for service. If the queue is empty the service of this station is completed and the server polls the next station.

Let the n-th polling cycle of station j be the epoch between the n-th and (n+1)-st poll of this station. Then $p_{k,m,j}^{(n)}$ denotes the joint probability that k packets are waiting for service at station j and m packets have already been served at this station in the n-th polling cycle. For example $p_{1,0,j}^{(n)}$ denotes the probability that one packet is waiting for service at station j just prior the n-th poll and $p_{0,3,j}^{(n)}$ denotes the probability that the queue is empty immediately after the end of the third service phase. For ease of reading the subscript j indicating the observed station is omitted in the following, i.e. the notations $p_{k,m}^{(n)}$ and S will be used instead of $p_{k,m,j}^{(n)}$ and S_j, respectively. These state probabilities can now be calculated recursively. Analogously to [14] we obtain:

$$
\begin{aligned}
p_{k,m}^{(n)} &= \sum_{i=1}^{k+1} g_{k-i+1} p_{i,m-1}^{(n)} && k = 0..S-2 \\
& && m = 1..M; \\
p_{S-1,m}^{(n)} &= \sum_{i=1}^{S} g^c_{S-i} p_{i,m-1}^{(n)} && m = 1..M \\
p_{S,m}^{(n)} &= 0 && m = 1..M \\
p_{k,0}^{(n+1)} &= \sum_{m=0}^{M-1} h_k p_{0,m}^{(n)} + \sum_{i=0}^{k} h_{k-i} p_{i,M}^{(n)} && k = 0..S-1 \\
p_{S,0}^{(n+1)} &= \sum_{m=0}^{M-1} h^c_S p_{0,m}^{(n)} + \sum_{i=0}^{S-1} h^c_{S-i} p_{i,M}^{(n)}.
\end{aligned}
\tag{1}
$$

In Equation (1) g_i and h_i denote the probability that i packets arrive during a service time T_h or vacation time T_v, respectively. $g_i^c = \sum_{k=i}^{\infty} g_k$ denotes the probability that at least i packets arrive during a service time. h_i^c is defined in the same way.

To calculate these arrival probabilities the technique presented in [15] can be used. Therefore, the probability distribution function of the service time T_h and the vacation time T_v are approximated by their first two moments.

Kühn proposed in [10] two different approximations for the probability distribution function of a random variable T with mean $\overline{T}$ and coefficient of variation c, depending on the value of c.

In case of hypoexponential process types ($0 \leq c \leq 1$), a series of a deterministic (D) and an exponential (M) phase is chosen. Its probability distribution function F is given by

$$F(t) = \begin{cases} 0 & 0 \leq t \leq t_1 \\ 1 - e^{-(t-t_1)/t_2} & t \geq t_1 \end{cases} \tag{2}$$

where

$$t_1 = \overline{T}(1-c) \tag{3}$$

and

$$t_2 = \overline{T}c. \tag{4}$$

In case of hyperexponential process types ($c \geq 1$), an alternative of two exponential phases is chosen. Its probability distribution function is given by

$$1 - q\,e^{-t/t_1} - (1-q)\,e^{-t/t_2} \tag{5}$$

where

$$t_{1,2} = \overline{T}/(1 \pm \sqrt{\frac{c^2-1}{c^2+1}}) \tag{6}$$

and

$$q = \overline{T}/2t_1. \tag{7}$$

3.2 Calculation of the Vacation Time

Let the random variable $T_{s,j}$ be the time between the beginning of the first and the end of the last service phase of station j. It will be referred to as station time in the following. Further, let $\Phi_{s,j}$, $\Phi_{h,j}$, and $\Phi_{o,j}$ be the Laplace-Stieltjes transforms (LST) of the station time, service time, and switchover time of station j, respectively. The station index j will be omitted again in the following formulas. Under the assumption of independence between the event that exactly i packets are served by the server in one cycle at this station and the service time T_h, the LST Φ_s is given by

$$\Phi_s = \sum_{i=0}^{M} d_i \Phi_h^i \tag{8}$$

where d_i denotes the probability that exactly i packets are served:

$$d_i = \begin{cases} a_0 & i = 0 \\ a_i \prod_{k=0}^{i-1}(1-a_k) & i = 1..M \end{cases} \tag{9}$$

and

$$a_i = \begin{cases} p_{0,i} \left[\sum_{k=0}^{S} p_{k,i} \right]^{-1} & i = 0..M-1 \\ 1 & i = M. \end{cases} \tag{10}$$

In the preceding equations a_i denote the branching probabilities of the phase model in Figure 2. Note that Equation (8) becomes exact if the service time is constant. Further, it should be mentioned that the superscript $^{(n)}$ for the polling cycle has been omitted for ease of reading.

Under the assumption of independence between $T_{s,j}$ and $T_{o,j}$ for $j = 1..z$, the LST of the vacation time is given as follows:

$$\Phi_{v,j} = \prod_{k=1}^{z} \Phi_{o,k} \prod_{\substack{k=1 \\ k \neq j}}^{z} \Phi_{s,k}. \tag{11}$$

From the LST the first two moments can be easily obtained by differentiating.

3.3 Arbitrary Time State Probabilities

To calculate the state probabilities at arbitrary time instants we can use the results of [14]. In that paper equations for the joint probabilities for a distinct queue length and the server being in the m-th service phase or the vacation phase of a station are presented. We don't want to repeat the complete derivation of these equations here but only the most important results. The interested reader is referred to [13] and [14] for derivations.

Let $\pi_{k,m,j}$ be the joint probability that the queue at station j contains k packets at an arbitrary time instant and that this time instant is in the m-th service phase ($m = 1..M$) or the vacation phase ($m = 0$). In the following we will again omit the station index j. According to [14] the $\pi_{k,m}$ can be calculated from:

$$\begin{aligned}
\pi_{0,m} &= 0 && m=1..M \\
\pi_{k,m} &= \frac{\sigma}{\lambda}\left[\sum_{j=1}^{k} p_{j,m-1} - \sum_{j=0}^{k-1} p_{j,m}\right] && k=1..S-1;\ m=1..M \\
\pi_{S,m} &= \frac{\sigma}{\lambda}\Bigg[\lambda E[T_h]\sum_{j=1}^{S} p_{j,m-1} - \sum_{j=1}^{S}(S-j)p_{j,m-1} + \\
&\qquad + \sum_{j=0}^{S-1}(S-j-1)p_{j,m}\Bigg] && m=1..M \\
\pi_{k,0} &= \frac{\sigma}{\lambda}\left[\sum_{m=0}^{M-1} p_{0,m} + \sum_{j=0}^{k} p_{j,M} - \sum_{j=0}^{k} p_{j,0}\right] && k=0..S-1 \\
\pi_{S,0} &= \frac{\sigma}{\lambda}\left[\lambda E[T_v]\sum_{j=0}^{S} p_{j,0} - \sum_{j=0}^{S-1} j p_{j,M} + \sum_{j=0}^{S} j p_{j,0}\right].
\end{aligned} \tag{12}$$

In Equation (12) the $p_{k,m}$ denote the steady state probabilities of the embedded Markov chain. Note that, for $m \neq 1$, the $p_{S,m-1}$-terms are zero and therefore don't contribute to the $\pi_{S,m}$. σ denotes the frequency of embedded points. Thus, the inverse of σ is the average interval between consecutive embedded points. It can be expressed by

$$\sigma^{-1} = bE[T_v] + (1-b)E[T_h] \tag{13}$$

where b denotes the probability that an embedded point is the point just prior to the polling instant of the considered station. It is given by:

$$b = \sum_{k=0}^{S} p_{k,0}. \tag{14}$$

Using the theorem of the total probability, the state probabilities at arbitrary time instants π_k are given by:

$$\pi_k = \sum_{m=0}^{M} \pi_{k,m}. \tag{15}$$

After some algebraic manipulations we obtain the final result:

$$\pi_k = \begin{cases} \frac{\sigma}{\lambda} \sum_{m=1}^{M} p_{k,m} & k = 0..S-1 \\ 1 - \frac{(1-b)\sigma}{\lambda} & k = S. \end{cases} \tag{16}$$

With these arbitrary time state probabilities we can easily derive some characteristic system measures as the blocking probability, the mean queue length, and the mean waiting time.

3.4 Numerical Algorithm

Using the expressions derived in the previous subsections, an iterative algorithm similar to that presented in [15] can be developed. The main steps are:

1. Initialize all state probabilities and station times for all queues
2. REPEAT {iteration cycle}

    FOR all queues DO
    BEGIN
    (a) Calculate the vacation time
    (b) Calculate the probabilities for arrivals during all service and vacation times
    (c) Calculate the state probabilities
    (d) Calculate the branching probabilities
    (e) Calculate the new station time
    END;

    UNTIL (convergence criterion Δ is fulfilled);
3. Calculate the state probabilities for arbitrary time instants
4. Calculate the system characteristics

The convergence criterion Δ can, e.g., be defined as

$$\Delta = \sum_{j=1}^{z} \Delta_j \tag{17}$$

where

$$\Delta_j = \left| \sum_{k=0}^{S_j} \frac{p_{k,0,j}^{(n+1)} - p_{k,0,j}^{(n)}}{p_{k,0,j}^{(n+1)}} \right|. \tag{18}$$

4 Numerical Results

In Figures 3 and 4 some typical results obtained by the presented analysis are compared with computer simulations. A symmetrical system with 10 stations and 8 buffer places at each station has been analyzed. The service time has been assumed to be a constant $100\mu s$ and the switchover time a constant $1\mu s$ for all stations. The system has been symmetrically loaded.

Figure 3 shows the loss probability versus the offered load per station. The maximum number of services per cycle M has been varied between 1 and 40 without observing significant changes in the obtained results. This is due to the fact that the switchover time has been assumed to be very small compared to the service time. Therefore, the overhead per cycle is also small. The situation changes if the switchover time becomes nearly equal to the service time and thus the overhead per cycle is in the range of the transmission time of packets. The depicted curve represents the loss probability for $1 \le M \le 40$.

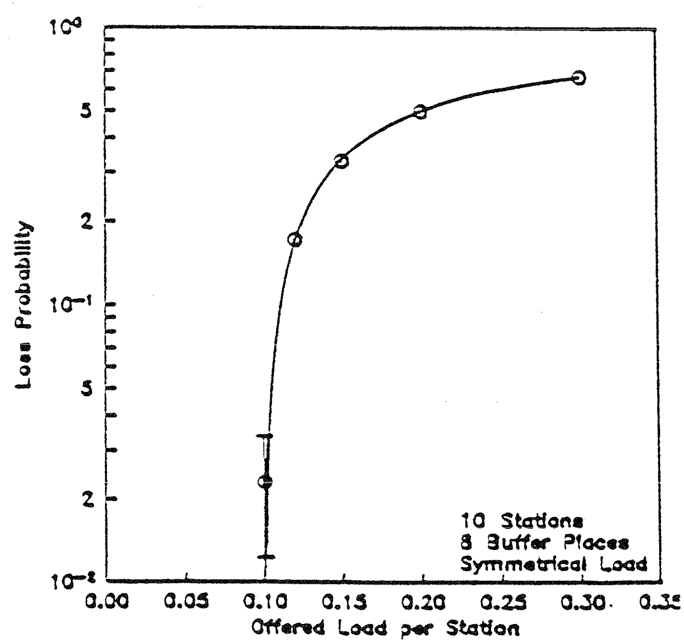

Figure 3: Loss Probability versus Offered Load

In Figure 4 the mean queue length versus the offered load per station is depicted. As a parameter the maximum number of services per cycle M is used. The curves

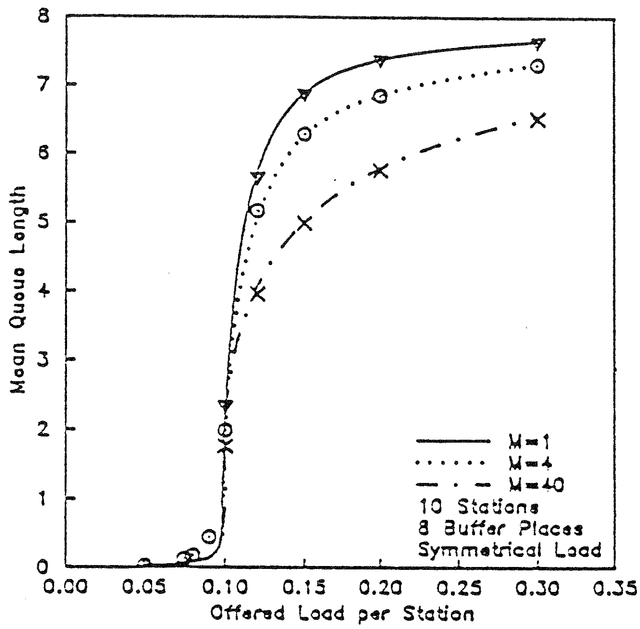

Figure 4: Mean Queue Length versus Offered Load

show the typical behavior. If the system is slightly loaded only a few buffer places are occupied but if the total system becomes overloaded the mean queue length increases rapidly. For lower load conditions the three curves are identical. Under high load a low number M results in a higher number of occupied buffer places. This is due to an increased overhead since less packets can be served per cycle. This effect will be increased if the overhead per cycle becomes higher.

These results correspond qualitatively to the well-known behaviour of polling systems [15]. As we can see, both figures demonstrate an excellent accuracy of the presented method.

5 Conclusions

In merging two known analytical methods, we have developed an approximate algorithm to calculate characteristic results for finite capacity polling systems with 'limited-M' service. The analysis is based on an iterative calculation of the number of services at each station and the vacation times. To obtain a distribution function for these random variables a two-moment approximation has been used. The comparison with computer simulations has shown an excellent accuracy.

Acknowledgements

The authors would like to thank Hans Kröner for many helpful discussions and Thomas Renger for the implementation of the simulation program.

References

[1] R.B. Cooper and G. Murray: Queues Served in Cyclic Order, The Bell System Technical Journal, Vol. 48, 1969, pp. 675–689.

[2] M. Eisenberg: Queues with Periodic Service and Changeover Times, Operations Research, Vol. 20, 1972, pp. 440–451.

[3] D. Everitt: Approximations for Asymmetric Token Rings with a Limited Service Discipline, British Telecom Tech. J., Vol. 6, No. 3, 1988, pp. 46–51.

[4] S.W. Fuhrmann: Symmetric Queues Served in Cyclic Order, Operations Research Letters, Vol. 4, No. 3, 1985, pp. 139–144.

[5] S.W. Fuhrmann and Y.T. Wang: Analysis of Cyclic Service Systems with Limited Service: Bounds and Approximations, Performance Evaluation, Vol. 9, No. 1, 1988, pp. 35–54.

[6] O. Hashida: Analysis of Multiqueue, Rev. Electr. Commun. Lab., Vol. 20, 1972, pp. 189–199.

[7] J.F. Hayes and A. Jalali-Nadushan: Numerical Solution to Limited Service Polling Models, Computer Communications, Vol. 9, No. 4, 1986, pp. 171–176.

[8] G. Kimura and Y. Takahashi: Traffic Analysis for Token Ring Systems with Limited Service, Electronics and Communications in Japan, Part 1, Vol. 71, No. 8, 1988, pp. 80–90.

[9] P. Kühn: Multiqueue Systems with Nonexhaustive Cyclic Service, The Bell System Technical Journal, Vol. 58, 1979, pp. 671–699.

[10] P. Kühn: Approximate Analysis of General Queueing Networks by Decomposition, IEEE Trans. on Communications, Vol. 27, 1979, pp. 113–126.

[11] D.R. Manfield: Analysis of a Polling System with Priorities, Proceedings of the Globecom'83, San Diego, 1983, Paper 43.4.

[12] H. Takagi: Analysis of Polling Systems, The MIT Press, Cambridge Massachusetts, 1986.

[13] T. Lee: M/G/1/N Queue with Vacation Time and Exhaustive Service Discipline, Operations Research, Vol. 32, 1984, pp. 774–784.

[14] T. Lee: M/G/1/N Queue with Vacation Time and Limited Service Discipline, Performance Evaluation 9, 1988/89, pp. 181–190.

[15] P. Tran-Gia, T. Raith: Multiqueue Systems with Finite Capacity and Nonexhaustive Cyclic Service, Int. Seminar on Computer Networking and Performance Evaluation, Tokyo, 1985, pp. 213–225.