

# Optimal design of a finite-buffer polling network with mixed service discipline and general service order sequence

L.-C. Hwang  
C.-J. Chang

*Indexing terms:* Polling network, Queueing theory

**Abstract:** Polling schemes have been widely used in LANs, MANs, and distributed systems. The authors design a polling network (system) in which each station is finite-buffered and is served according to a mixed service discipline and a general service order sequence. They first analyse the network by means of an imbedded Markov chain approach to obtain probability generating functions (PGFs) for the number of customers in stations at four observation points: customer-service beginning, customer-service ending, server's arrival, and server's departure. Two performance measures of mean waiting time and blocking probability are then derived. Finally, the authors search for an optimal pattern of the mixed service discipline and service order sequence for a polling network via a genetic algorithm. The results show that a polling network with an optimal pattern of mixed service discipline and service order sequence offers a great improvement in performance over a polling network with a unique service discipline and cyclic service order sequence; a polling network has different optimal patterns for different traffic intensities; and a near-optimal pattern could exist for all traffic loads.

## 1 Introduction

Polling schemes have been widely used in local area networks (LANs), metropolitan area networks (MANs), and distributed communication systems. In the past, these networks or systems were generally designed using a unique service discipline and a cyclic service order sequence for all stations. As customer demand for multimedia services increases, however, the above networks can be improved by increasing the transmission speed or redesigning the medium access control (MAC) protocol to support multiple classes of services. In addition, the network should be designed in such a way that each station is assigned an appropriate service discipline and a suitable polling number and polling order in the sequence.

© IEE, 1995

Paper 15371 (E7, C3), first received 18th October 1993 and in revised form 19th July 1994

The authors are with the Department of Communication Engineering and Center for Telecommunications Research, National Chiao Tung University, Hsinchu, Taiwan 300, Republic of China

*IEE Proc.-Commun., Vol. 142, No. 1, February 1995*

Many sorts of service disciplines and service order sequences have been proposed for use in polling networks. In Reference 1, Takagi presented a very good survey of these techniques. Previous discussions of service disciplines for cyclic service order sequence can be found in References 2 to 6. Ferguson and Aminetazh [2] studied exhaustive and gated service disciplines; Fuhrmann and Wang [3] studied a limited discipline; Tran-Gia and Raith [4] and Tran-Gia [5] studied a nonexhaustive service discipline; and Takagi [6] studied exhaustive, gated, and limited service disciplines. Previous research on service disciplines for general service order sequence can be found in References 7 to 11. Eisenberg [7] studied exhaustive and gated service disciplines; Baker and Rubin [8] studied an exhaustive service discipline; Choudhury [9] studied a gated service discipline; and Chang and Hwang [10] studied a gated-limited service discipline. Boxma *et al.* [11] have studied the optimal service order sequence for a polling network with infinite buffer.

In this paper, we design a finite-buffered polling network with a mixed service discipline and general service order sequence. The mixed service discipline means that any stage (a turn in the service order sequence) in the network can be served using the limited, exhaustive, or gated service discipline. We analyse the polling network to obtain the mean waiting time and the blocking probability by means of a Markov chain approach and operators representing various service disciplines. We then use the genetic algorithm to search for an optimal pattern of the mixed service discipline and service order sequence for a polling network.

## 2 Analysis

The polling network is assumed to contain  $R$  stations and  $P$  stages in the service order sequence. We use  $r$  and  $i$  to represent the indexes of the station and stage, respectively, and use the notation  $r_i$  to represent the underlying station of stage  $i$ . Note that in the following the index of the stage uses modulo- $P$  arithmetic and is equal to  $P$  if the remainder is zero.

The arrival process for station  $r$  is assumed to be an independent Poisson process with rate  $\lambda_r$ ,  $1 \leq r \leq R$ ; the service discipline for stage  $i$  is assumed to be limited

This research is supported by the National Science Council, Taiwan, Republic of China, under contract number: NSC 81-0404-E-009-528.

(exhaustive-limited, gated-limited, or nonexhaustive), exhaustive, or gated. If stage  $i$  adopts exhaustive-limited (E-limited) service discipline, the server will serve  $k_i$  customers or serve until the stage empty, which happens first, at stage  $i$ . If stage  $i$  adopts gated-limited (G-limited) service discipline, the server will serve  $k_i$  customers or all the customers in the stage seen at the server's arrival epoch, which happens first, at stage  $i$ .  $k_i$  is a positive integer number. If stage  $i$  adopts nonexhaustive service discipline, the server will serve one customer at stage  $i$  if stage  $i$  is not empty at the server's arrival epoch. If stage  $i$  adopts exhaustive service discipline, the server will serve stage  $i$  until it is empty. If stage  $i$  adopts gated service discipline, the server will serve all the customers in stage  $i$  seen at the server's arrival epoch. For the following analysis, we first define the notations below.

- $S_r$  = the service time of a customer in station  $r$  with mean  $s_r$
- $\hat{S}_r(z)$  = the PGF of the number of customers arriving at station  $r$  during  $S_r$
- $B_r$  = the buffer size for station  $r$
- $L_r$  = the mean system size of station  $r$
- $\eta_r$  = the effective arrival rate of station  $r$
- $w_r$  = the mean waiting time of the customer served at station  $r$
- $q_r(n)$  = the probability that there are  $n$  customers in station  $r$  observed at the arrival point of a customer in station  $r$ ,  $0 \leq n \leq B_r$
- $\pi_r(n)$  = the probability that there are  $n$  customers in station  $r$  observed at the departure point of a customer from station  $r$ ,  $0 \leq n \leq B_r - 1$
- $c$  = the mean whole cycle time
- $U_i$  = the walking time for the server to switch over from stage  $i$  to stage  $i + 1$  with mean  $u_i$
- $\beta_i$  = the mean number of customers served in stage  $i$  during a whole cycle
- $e_i(n)(b_i(n))$  = the probability that there are  $n$  customers in station  $r_i$  when an  $i$  customer ends (begins) service, where  $i$ -customer is the customer served at stage  $i$
- $\hat{e}_i(z)(\hat{b}_i(z))$  = the PGF of the random variable with probability mass function  $e_i(n)(b_i(n))$
- $d_i(n)(a_i(n))$  = the probability that there are  $n$  customers in station  $r_i$  when the server departs from (arrives at) stage  $i$
- $\hat{d}_i(z)(\hat{a}_i(z))$  = the PGF of the random variable with probability mass function  $d_i(n)(a_i(n))$ .

Similar to Reference 7, the relation among  $\hat{a}_i(z)$ ,  $\hat{b}_i(z)$ ,  $\hat{d}_i(z)$ , and  $\hat{e}_i(z)$  can be obtained by

$$\beta_i \hat{b}_i(z) + \hat{d}_i(z) = \beta_i \hat{e}_i(z) + \hat{a}_i(z) \quad (1)$$

Note that  $\hat{a}_i(1)$ ,  $\hat{b}_i(1)$ ,  $\hat{d}_i(1)$ , and  $\hat{e}_i(1)$  are all equal to 1. Moreover, the number of customers in station  $r_i$  when an  $i$ -customer (i.e. customer served at stage  $i$ ) ends service is equal to the number of customers in station  $r_i$  when the  $i$ -customer begins service plus the number of customers arriving at station  $r_i$  during the service time of this  $i$ -customer (the summation cannot be greater than the finite buffer size  $B_r$ ) minus one (the served  $i$ -customer). This results in

$$\hat{e}_i(z) = \mathcal{B}_r[\hat{b}_i(z)\hat{S}_r(z)]z^{-1} \quad (2)$$

where  $\mathcal{B}_r$  is an operator on the PGF that adds the coefficients of the high-order  $z$  terms whose powers are greater

than  $B_r$ , to the coefficient of  $z^{B_r}$  and then truncates the high-order  $z$  terms. Mathematically, the operator  $\mathcal{B}_r$  on PGF  $\hat{f}(z) = \sum_{n=0}^{\infty} f(n)z^n$  can be expressed as

$$\begin{aligned} \mathcal{B}_r[\hat{f}(z)] &= \mathcal{B}_r\left[\sum_{n=0}^{\infty} f(n)z^n\right] \\ &= \sum_{n=0}^{B_r-1} f(n)z^n + \sum_{n=B_r}^{\infty} f(n)z^{B_r} \end{aligned} \quad (3)$$

Substituting eqn. 2 into eqn. 1 yields

$$\beta_i \hat{b}_i(z) - \mathcal{B}_r[\beta_i \hat{b}_i(z)\hat{S}_r(z)]z^{-1} = \hat{a}_i(z) - \hat{d}_i(z) \quad (4)$$

If  $\hat{a}_i(z)$  and  $\hat{d}_i(z)$  are found,  $\beta_i \hat{b}_i(z)$  can be obtained by comparing the coefficients of the right-hand side and left-hand side of eqn. 4. Then  $b_i(z)$  is obtained by normalising  $\beta_i \hat{b}_i(z)$  with  $\beta_i$  which is equal to  $\beta_i \hat{b}_i(1)$  and  $\hat{e}_i(z)$  is obtained by eqn. 2 for all  $i$ . Consequently, the mean whole cycle time,  $c$ , can be obtained by

$$c = \sum_{i=1}^P (\beta_i s_r + u_i) \quad (5)$$

and the effective arrival rate for station  $r$ ,  $\eta_r$ , is obtained by

$$\eta_r = \frac{1}{c} \sum_{i|r_i=r} \beta_i \quad 1 \leq r \leq R \quad (6)$$

Using  $\eta_r$  and  $\lambda_r$  yields blocking probability for station  $r$ ,  $q_r(B_r)$ , as

$$q_r(B_r) = 1 - \frac{\eta_r}{\lambda_r} \quad 1 \leq r \leq R \quad (7)$$

Furthermore,  $\pi_r(n)$ , for all  $r$ , can be obtained by

$$\pi_r(n) = \frac{1}{\eta_r c} \sum_{i|r_i=r} \beta_i e_i(n) \quad 0 \leq n \leq B_r - 1 \quad (8)$$

Using  $q_r(B_r)$  and  $\pi_r(n)$ , we obtain  $q_r(n)$ , for all  $r$ , by [12, Section 5.1.8]

$$q_r(n) = [1 - q_r(B_r)]\pi_r(n) \quad 0 \leq n \leq B_r - 1 \quad (9)$$

Then, owing to Poisson arrivals see time averages (PASTA) property [13, p. 294],  $L_r$  is obtained by

$$L_r = \sum_{n=1}^{B_r} n q_r(n) \quad 1 \leq r \leq R \quad (10)$$

Finally, from Little's law, the mean waiting time,  $w_r$ , can be obtained by

$$w_r = \frac{L_r}{\eta_r} - s_r \quad 1 \leq r \leq R \quad (11)$$

To obtain  $\hat{a}_i(z)$  and  $\hat{d}_i(z)$ , we observe all stations in the network and define the following notation

- $\hat{\mathbf{a}}_i(\mathbf{Z})$  = the PGF of the number of customers in all independent stations when the server arrives at stage  $i$ ,  $\hat{\mathbf{a}}_i(\mathbf{Z}) \equiv \sum_{\mathbf{a}^i} \Pr(\mathbf{a}^i) \mathbf{Z}^{\mathbf{a}^i}$ , where  $\mathbf{a}^i = (a_1^i, a_2^i, \dots, a_R^i)$ ,  $\mathbf{Z} = (z_1, z_2, \dots, z_R)$ , and  $\mathbf{Z}^{\mathbf{a}^i} \triangleq \prod_{r=1}^R z_r^{a_r^i}$ ; similarly
- $\hat{\mathbf{d}}_i(\mathbf{Z})$  = The PGF of the number of customers in all independent stations when the server departs from stage  $i$
- $\hat{\mathbf{U}}_i(\mathbf{Z})$  = The PGF of the number of customers arriving at all independent stations during  $U_i$
- $\hat{\mathbf{S}}_i(\mathbf{Z})$  = the PGF of the number of customers arriving at all independent stations during  $S_r$

The relation between  $\hat{\mathbf{a}}_i(\mathbf{Z})$  and  $\hat{\mathbf{d}}_i(\mathbf{Z})$  depends on the service discipline assigned by the network for stage  $i$ . We

here use the operator  $\mathcal{S}_i$  to express their relationship as follows

$$\hat{a}_i(\mathbf{Z}) = \mathcal{S}_i[\hat{d}_i(\mathbf{Z})] \quad (12)$$

We discuss the operator  $\mathcal{S}_i$  for various service disciplines in the Appendix.

The number of customers in station  $r$  when the server arrives at stage  $i + 1$  is equal to the number of customers in station  $r$  when the server departs from stage  $i$  plus the number of customers entering station  $r$  during walking time  $U_i$ ,  $1 \leq r \leq R$ . Therefore,  $\hat{a}_{i+1}(\mathbf{Z})$  can be given by

$$\hat{a}_{i+1}(\mathbf{Z}) = \mathcal{L}[\hat{d}_i(\mathbf{Z})\hat{U}_i(\mathbf{Z})] \quad (13)$$

where  $\mathcal{L}$  is a linear operator on  $\mathbf{Z}^{B_r}$  that transforms  $z_r^{n_r}$  to  $z_r^{B_r}$  if  $n_r \geq B_r$  and leaves  $z_r^{n_r}$  unchanged if  $n_r < B_r$  for all  $r$ . Substituting  $\hat{d}_i(\mathbf{Z})$  of eqn. 12 into eqn. 13, we have

$$\hat{a}_{i+1}(\mathbf{Z}) = \mathcal{L}[\mathcal{S}_i[\hat{a}_i(\mathbf{Z})]\hat{U}_i(\mathbf{Z})]$$

For simplicity, we define

$$\hat{a}_{i+1}(\mathbf{Z}) = \mathcal{L}[\mathcal{S}_i[\hat{a}_i(\mathbf{Z})]\hat{U}_i(\mathbf{Z})] \triangleq \mathcal{D}_i[\hat{a}_i(\mathbf{Z})] \quad (14)$$

Then, for a polling network with the general service order sequence, we have the following recursive relation

$$\hat{a}_{i+p}(\mathbf{Z}) = \mathcal{D}_{i+p-1}[\mathcal{D}_{i+p-2}[\cdots[\mathcal{D}_{i+1}[\mathcal{D}_i[\hat{a}_i(\mathbf{Z})]]]\cdots]] \quad (15)$$

Since  $\hat{a}_{i+p}(\mathbf{Z}) = \hat{a}_i(\mathbf{Z})$ , we can solve the  $\prod_{r=1}^R (B_r + 1)$  unknown coefficients of  $\hat{a}_i(\mathbf{Z})$  by using an iterative algorithm. The problem of the existence of such a huge number of unknowns seems inherent in the exact analysis of finite-capacity multiple-queue systems [5, p. 386]. Our approach makes the number of unknowns about  $R$  times less than that in Section 3.1 of Reference 5. Furthermore, the iterative algorithm uses  $\prod_{r=1}^R (B_r + 1)$  times less memory space than the matrix solving method used in Reference 15. If  $\hat{a}_i(\mathbf{Z})$  is found for a certain  $i$ , for example,  $i = 1$ ,  $\hat{d}_i(\mathbf{Z})$  can be solved via eqn. 12 and  $\hat{a}_i(z)$  and  $\hat{d}_i(z)$  can be obtained by  $\hat{a}_i(\mathbf{Z}_{r_i})$  and  $\hat{d}_i(\mathbf{Z}_{r_i})$ .  $\mathbf{Z}_{r_i}$  is  $(1, 1, \dots, z, 1, \dots, 1)$ , obtained by replacing all elements of  $\mathbf{Z}$  by 1 except for  $z_{r_i}$ , which is replaced by  $z$ . Finally,  $\hat{a}_i(\mathbf{Z})$ ,  $\hat{d}_i(\mathbf{Z})$ ,  $\hat{a}_i(z)$ , and  $\hat{d}_i(z)$  for  $1 \leq i \leq P$  can be found recursively.

### 3 Optimal pattern design of the mixed service discipline and service order sequence

First, we describe the numerical procedures used to obtain the blocking probability and the mean waiting time of a customer served at station  $r$ ,  $r = 1, 2, \dots, R$ .

*Numerical Algorithm:*

*Step 1:* [Initialisation]

$$i = 1$$

$$\text{Give an initial } \hat{a}_i(\mathbf{Z}) \left( \equiv \sum_{\mathbf{a}^i} \text{Pr}(\mathbf{a}^i) \mathbf{Z}^{\mathbf{a}^i} \right)$$

*Step 2:* [Find  $\hat{a}_i(\mathbf{Z})$  by an iterative algorithm]

$$\hat{a}_i^*(\mathbf{Z}) \left( \equiv \sum_{\mathbf{a}^i} \text{Pr}^*(\mathbf{a}^i) \mathbf{Z}^{\mathbf{a}^i} \right) = \hat{a}_i(\mathbf{Z})$$

$$\hat{a}_{i+p}(\mathbf{Z}) = \mathcal{D}_{i+p-1} \times [\mathcal{D}_{i+p-2} [\cdots [\mathcal{D}_{i+1} [\mathcal{D}_i[\hat{a}_i(\mathbf{Z})]]] \cdots]] \quad (\text{from eqn. 15})$$

Find a weighting factor  $\omega$  according to the convergence situation in Seelen's algorithm [16]

$$\hat{a}_i(\mathbf{Z}) = \hat{a}_{i+p}(\mathbf{Z}) + \omega(\hat{a}_{i+p}(\mathbf{Z}) - \hat{a}_i^*(\mathbf{Z}))$$

$$\text{IF } \frac{|\text{Pr}(\mathbf{a}^i) - \text{Pr}^*(\mathbf{a}^i)|}{\text{Pr}(\mathbf{a}^i)} > 10^{-3} \text{ for any } \mathbf{a}^i$$

GO TO Step 2

END IF

*Step 3:* [Obtain the PGFs]

DO  $i = 1, P$

$$\hat{d}_i(\mathbf{Z}) = \mathcal{S}_i[\hat{a}_i(\mathbf{Z})] \text{ (from eqn. 12)}$$

$$\hat{a}_i(z) = \hat{a}_i(\mathbf{Z}_{r_i}) \text{ and } \hat{d}_i(z) = \hat{d}_i(\mathbf{Z}_{r_i})$$

Get  $\hat{b}_i(z)$  and  $\beta_i$  from eqn. 4

$$\hat{e}_i(z) = \mathcal{B}_{r_i}[\hat{b}_i(z)\hat{S}_{r_i}(z)]z^{-1} \text{ (from eqn. 2)}$$

IF  $i \neq P$

$$\hat{a}_{i+1}(\mathbf{Z}) = \mathcal{L}[\hat{d}_i(\mathbf{Z})\hat{U}_i(\mathbf{Z})] \text{ (from eqn. 13)}$$

END IF

END DO.

*Step 4:* [Obtain the performance measures]

$$c \text{ (the mean whole cycle time)} = \sum_{i=1}^P (\beta_i s_{r_i} + u_i)$$

DO  $r = 1, R$

$$\eta_r \text{ (the effective arrival rates)} = \frac{1}{c} \sum_{(i|r_i=r)} \beta_i$$

$$q_r(B_r) \text{ (the blocking probability)} = 1 - \frac{\eta_r}{\lambda_r}$$

DO  $n = 0, B_r - 1$

$q_r(n)$  (the buffer occupancy probability)

$$= \frac{1}{\lambda_r c} \sum_{(i|r_i=r)} \beta_i e_i(n)$$

END DO

$$L_r \text{ (the mean queue length)} = \sum_{n=1}^{B_r} n q_r(n)$$

$$w_r \text{ (the mean waiting times)} = \frac{L_r}{\eta_r} - s_r$$

END DO

When the service order sequence of the network and the service disciplines for all stages are assigned,  $\mathcal{S}_i$  and  $\mathcal{D}_i$  for all  $i$  are determined and then the blocking probability and the mean waiting time can be found by the above numerical algorithm. Within the numerical algorithm, we utilise Seelen's algorithm to speed up the convergence of the iteration [16]. The reader can refer to Reference 16 for details.

Next, we search for an optimal pattern of the mixed service discipline and service order sequence for the network using a generic algorithm. Genetic algorithms, which combine the survival of the fittest with the innovative flair of human search [14], are a form of the random search method which is suitable for discontinuous and multimodal problems. We describe this algorithm in the following.

Genetic Algorithm:

Step 1: [Initialisation]

$n = 0$

Heuristically choose a population of  $m$  candidates

$\{x_1^n, x_2^n, \dots, x_m^n\}$

$x_{op}$  (optimal candidate) =  $x_1^0$

Step 2: [Evaluate fitness]

DO  $i = 1, m$

Find fitness  $f(x_i^n)$  ( $f(\cdot)$  is a defined objective function)

IF  $f(x_i^n) > f(x_{op})$

$x_{op} = x_i^n$

END IF

END DO

Step 3: [Check the termination criterion]

IF the termination criterion is satisfied

GO TO Step 5

END IF

Step 4: [Produce the next generation]

Create  $\{\hat{x}_1^{n+1}, \hat{x}_2^{n+1}, \dots, \hat{x}_m^{n+1}\}$  from  $\{x_1^n, x_2^n, \dots, x_m^n\}$  according to the weights of fitnesses  $f(x_i^n)$ ,  $i = 1, 2, \dots, m$

Generate  $\{x_1^{n+1}, x_2^{n+1}, \dots, x_m^{n+1}\}$  by crossover and mutation [14]

$n = n + 1$

GO TO Step 2

Step 5: [End]

PRINT the optimal candidate  $x_{op}$  and its fitness  $f(x_{op})$

In this paper, we heuristically define the objective function for a given candidate  $x_i^n$  in the  $n$ th generation, denoted by  $f(x_i^n)$ , as

$$f(x_i^n) = \sum_{r=1}^R \frac{\rho_r}{\rho} \frac{1 - q_r(B_r)}{1 + \frac{w_r}{s_r}} \quad (16)$$

We define this objective function to find a system with smaller blocking probability and smaller mean waiting time. The explicit parameters of the objective function  $f(x_i^n)$  are the mean waiting times and the blocking probabilities, which are performance measures for a candidate  $x_i^n$  representing a pattern of the mixed service discipline and service order sequence (the implicit parameters). Note that the service discipline of each stage and the service order sequence are coded into a binary string of genes. Here, we utilise the GAUCSD 1.4 developed at the University of California, San Diego [17], and adopt a predetermined number of searched candidates as our termination criterion. The selection of the mutation rate, the population size, etc. are suggested in this package.

### 3.1 Numerical examples and discussion

In this Section we use an example to illustrate the design of an optimal pattern of the mixed service discipline and service order sequence for a finite-buffered polling network. We assume that there are five stations in the network, and the arrival rates are 2 : 1 : 1 : 1 : 1 for stations 1–5, respectively. The buffer size is assumed to be

the same for every station and is equal to 3 in this example; the service time  $S_r$  is exponentially distributed for station  $r$ ,  $1 \leq r \leq R$ ; the mean service time  $s_r$  is equal to 1 for every station  $r$ ; and the walking time  $U_i$  is deterministic and is equal to 0.1 for every stage  $i$ .

If the network is designed to fulfil the requirement of being fair to all customers, the traffic intensities for all stations should be made as equal as possible. In this example, there are four light-load stations and one high-load station, so we consider four types of service order sequences that poll the high-load station from one to four times. The four sequences are as follows: Order I: {1, 2, 3, 4, 5}; Order II: {1, 2, 3, 1, 4, 5}; Order III: {1, 2, 1, 3, 1, 4, 5}; and Order IV: {1, 2, 1, 3, 1, 4, 1, 5}. The service discipline of each stage is assumed to be E-limited with limitation number 3, G-limited with limitation number 3, nonexhaustive, exhaustive, or gated. We use  $\bar{E}$ ,  $\bar{G}$ , N, E, and G to denote E-limited, G-limited, nonexhaustive, exhaustive, and gated service disciplines, respectively. We shall refer to the aggregation of the capital letter of the service discipline of every stage as the pattern of the mixed service discipline. For example, ENGENE means E, N, G, N, and E for stages 1, 2, 3, 4, and 5, respectively.

There are a total of 87040 cases for an enumerative search. GAUCSD suggests only 210 cases to be searched; the efficiency is about 99.76%. The optimal pattern of the mixed service discipline and service order sequence is shown in Table 1. The optimal pattern is EEEEEEEE

Table 1: Optimal patterns of the mixed service discipline and service order sequence

Traffic intensity	Optimal pattern of service discipline and service order
0.2	EEEEEEEE, Order IV
0.5	EEEEEEEE, Order IV
1.0	ENEEENEN, Order IV
1.5	ENENENEN, Order IV
2.0	ENENENEN, Order IV
3.0	ENENENEN, Order IV

and Order IV for traffic intensities below 1.0, ENEEENEN and Order IV for traffic intensities around 1.0, and ENENENEN and Order IV for traffic intensities above 1.0. We further plot the fitness of the above three optimal patterns in Fig. 1; as can be seen from the figure, the differences between the optimal patterns are not significant for traffic intensities below 1.0. We could suggest that the near-optimal pattern of ENENENEN and Order IV be used for all traffic intensities in this example. The near-optimal pattern is here defined as the pattern which can support system performance near to those supported by the optimal patterns at all traffic intensities. A network adopting the optimal pattern of ENENENEN and Order IV has fitness 0.1362 at buffer size 3 and system traffic intensity 1.5. A network adopting only the nonexhaustive service discipline for all stages and cyclic service order sequence has fitness 0.0554 at buffer size 3 and system traffic intensity 1.5. The optimal pattern offers about 146% improvement over the other pattern in this example.

## 4 Concluding remarks

This paper considers an optimal design problem for a finite-buffered polling network with mixed service discipline and general service order sequence. The analytical approach is by way of imbedded Markov chains and several operators defined to facilitate the analysis. We use

a genetic algorithm to find the optimal pattern of the mixed service discipline and service order sequence. The results show that the performance of a polling network

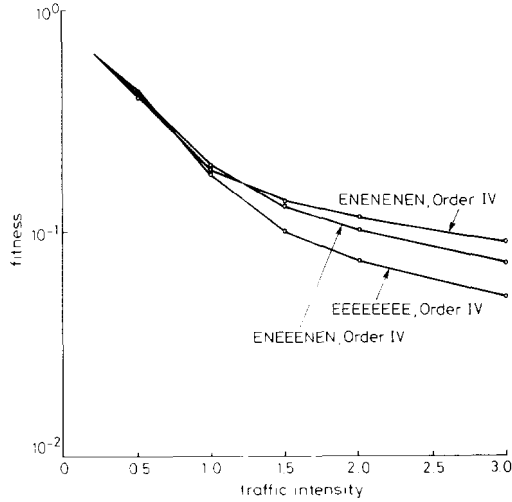


Fig. 1 Fitness of various optimal patterns for different traffic intensities

Buffer size = 3  
mean service time = 1  
walking time = 0.1  
 $\lambda_1 : \lambda_2 : \lambda_3 : \lambda_4 : \lambda_5 = 2 : 1 : 1 : 1 : 1$   
o simulation

can be greatly improved if an optimal pattern of service discipline and service order sequence is adopted; a polling network has different optimal patterns for different traffic intensities; and there could exist an near-optimal pattern for all traffic loads.

## 5 References

- 1 TAKAGI, H.: 'Queueing analysis of polling models: an update', in 'Stochastic analysis of computer and communication systems' (Elsevier Science Publishers B. V., North Holland, 1990), pp. 267-318
- 2 FERGUSON, M.J., and AMINETZAH, Y.J.: 'Exact result for non-symmetric token ring system', *IEEE Trans.*, 1985, **COM-33**, (3), pp. 223-231
- 3 FUHRMANN, S.W., and WANGY, Y.T.: 'Analysis of cyclic service systems with limited service: bound and approximations', *Perform. Eval.*, 1988, **9**, (1), pp. 35-54
- 4 TRAN-GIA, P., and RAITH, T.: 'Performance analysis of finite capacity polling systems with nonexhaustive service', *Perform. Eval.*, 1988, **9**, (1), pp. 1-16
- 5 TRAN-GIA, P.: 'Analysis of polling systems with general input process and finite capacity', *IEEE Trans. Commun.*, 1992, **40**, (2), pp. 337-344
- 6 TAKAGI, H.: 'Analysis of finite-capacity polling systems', *Adv. Appl. Probab.*, 1991, **23**, pp. 373-387
- 7 EISENBERG, M.: 'Queues with periodic service and changeover time', *Oper. Res.*, 1972, **20**, pp. 440-451
- 8 BAKER, J.E., and RUBIN, I.: 'Polling with a general-service order table', *IEEE Trans.*, 1987, **COM-35**, (3), pp. 283-288
- 9 CHOUDHURY, G.L.: 'Polling with a general service order table: gated service', *Proc. IEEE INFOCOM '90*, pp. 268-276
- 10 CHANG, C.J., and HWANG, L.C.: 'An asymmetrical polling system with general service order sequence and gated-limited service discipline', *Proc. IEEE Int. Symp. on Information theory*, 1991, Budapest, Hungary, p. 158
- 11 BOXMA, O.J., LEVY, H., and WESTSTRATE, J.A.: 'Efficient visit frequencies for polling tables: minimization of waiting cost', *Queueing Systems*, 1991, **9**, pp. 133-162
- 12 GROSS, D., and HARRIS, C.M.: 'Fundamentals of queueing theory' (John Wiley & Sons, 1985, second edition)

- 13 WOLFF, R.W.: 'Stochastic modeling and the theory of queues' (Prentice-Hall, 1989)
- 14 GOLDBERG, G.E.: 'Genetic algorithms in search, optimization & machine learning' (Addison-Wesley, 1989)
- 15 JUNG, W.Y., and UN, C.K.: 'Analysis of a finite-buffer polling system with exhaustive service based on virtual buffering', To appear in *IEEE Trans. Comm.*
- 16 SEELEN, L.P.: 'An algorithm for Ph/Ph/c queue', *Eur. J. Oper. Res.*, 1986, **23**, pp. 118-127
- 17 SCHRAUDOLPH, N.N., and GREFENSTETTE, J.J.: 'A user's guide to GAUCSD 1.4', Tech. Rep., CSE DEP., UC San Diego, La Jolla, CA 92093-0114

## 6 Appendix: The operator $\mathcal{S}_i$ for various service disciplines

### 6.1 Limited service discipline

We here discuss E-limited, G-limited, and nonexhaustive service disciplines. If the E-limited service discipline is assigned for stage  $i$ , then after completing service of a customer in stage  $i$ , the server will continue to serve the next customer if the number of customers already served is less than  $k_i$  and there are still customers in station  $r_i$ ; otherwise it will stop serving at stage  $i$ . According to this service discipline, we define an operator  $\mathcal{S}_i$  on  $\mathbf{Z}^{a^i}$  that changes  $z_r^{a_r}$  to  $\mathcal{S}_i(z_r^{a_r})z_r^{a_r-1}$  for  $r = r_i$  and  $a_{r_i}^i \geq 1$ , because the PGF of the number of customers arriving at all stations during the service time of an  $i$ -customer is  $\mathcal{S}_i(\mathbf{Z})$ , and that leaves  $z_r^{a_r}$  unchanged otherwise. Because the buffer at each station is finite and at most  $B_r(B_{r_i} - 1)$  customers can be in station  $r$ ,  $r \neq r_i$  ( $r = r_i$ ), when an  $i$ -customer ends service, we also define an operator  $\mathcal{X}_i$  on  $\mathbf{Z}^{a^i}$  that transforms  $z_r^{a_r}$  to  $z_r^{B_r}$  if  $a_r^i \geq B_r$ , and leaves  $z_r^{a_r}$  unchanged if  $a_r^i < B_r$ , for all  $r$  except  $r_i$ , and transforms  $z_{r_i}^{a_{r_i}^i}$  to  $z_{r_i}^{B_{r_i}-1}$  if  $a_{r_i}^i \geq B_{r_i}$  and leaves  $z_{r_i}^{a_{r_i}^i}$  unaltered if  $a_{r_i}^i < B_{r_i}$ . Expressed mathematically,  $\mathcal{S}_i$  and  $\mathcal{X}_i$  are given by

$$\mathcal{S}_i[\mathbf{Z}^{a^i}] = \begin{cases} z_1^{a_1^i} \cdots z_{r_i-1}^{a_{r_i-1}^i} \mathcal{S}_i(z_{r_i}^{a_{r_i}^i}) z_{r_i}^{a_{r_i}^i-1} z_{r_i+1}^{a_{r_i+1}^i} \cdots z_R^{a_R^i} & \text{if } a_{r_i}^i \geq 1 \\ \mathbf{Z}^{a^i} & \text{if } a_{r_i}^i < 1 \end{cases} \quad (17)$$

and

$$\begin{aligned} \mathcal{X}_i[\mathbf{Z}^{a^i}] &= z_1^{\min(a_1^i, B_1)} \cdots z_{r_i-1}^{\min(a_{r_i-1}^i, B_{r_i-1})} \\ &\quad \times z_{r_i}^{\min(a_{r_i}^i, B_{r_i}-1)} z_{r_i+1}^{\min(a_{r_i+1}^i, B_{r_i+1})} \cdots z_R^{\min(a_R^i, B_R)} \\ &= \left( \prod_{\substack{r=1 \\ r \neq r_i}}^R z_r^{\min(a_r^i, B_r)} \right) z_{r_i}^{\min(a_{r_i}^i, B_{r_i}-1)} \end{aligned} \quad (18)$$

For convenience, we denote by  $\mathcal{F}_i$  a new operator that combines the two operators  $\mathcal{S}_i$  and  $\mathcal{X}_i$ .  $\mathcal{F}_i$  is given by

$$\mathcal{F}_i[\cdot] = \mathcal{X}_i[\mathcal{S}_i[\cdot]] \quad (19)$$

Note that the two operators  $\mathcal{S}_i$  and  $\mathcal{X}_i$  are linear but not commutative. We also define  $\mathcal{F}_i^{k_i}$  as

$$\mathcal{F}_i^{k_i}[\cdot] = \mathcal{F}_i^{k_i-1}[\mathcal{F}_i[\cdot]] \quad \text{for } k_i \geq 2 \quad (20)$$

Then for the E-limited service discipline, the operator  $\mathcal{S}_i$  in eqn. 12 is  $\mathcal{F}_i^{k_i}$ .

If stage  $i$  is served with the G-limited service discipline, the server will serve  $\min(a_{r_i}^i, k_i)$  customers for station  $r_i$  when the server attends stage  $i$ . Then, the relation between  $\hat{a}_i(\mathbf{Z})$  and  $\hat{d}_i(\mathbf{Z})$  in eqn. 12 is expressed as

$$\hat{d}_i(\mathbf{Z}) = \sum_{a_1^i=0}^{B_1} \cdots \sum_{a_R^i=0}^{B_R} \Pr\{a^i\} \mathcal{F}_i^{\min(a_{r_i}^i, k_i)}[\mathbf{Z}^{a^i}] \quad (21)$$

The nonexhaustive service discipline is a special case of the G-limited or E-limited service discipline with  $k_i = 1$ . If stage  $i$  adopts the nonexhaustive service discipline, the operator  $\mathcal{S}_i$  in eqn. 12 is  $\mathcal{F}_i$ .

### 6.2 Exhaustive service discipline

This is a special case of the E-limited service discipline with  $k_i = \infty$ . If stage  $i$  adopts this service discipline, the operator  $\mathcal{S}_i$  in eqn. 12 is  $\mathcal{F}_{r_i}^\infty$ . There will be no customers in station  $r_i$  when the server departs from stage  $i$ . The infinite procedure will terminate when  $\hat{d}_i(0)$  is equal to 1.

### 6.3 Gated service discipline

This is a special case of the G-limited service discipline with  $k_i = \infty$ , then the relation between  $\hat{a}_i(\mathbf{Z})$  and  $\hat{d}_i(\mathbf{Z})$  in eqn. 12 is expressed as

$$\hat{d}_i(\mathbf{Z}) = \sum_{a_1=0}^{B_1} \cdots \sum_{a_R=0}^{B_R} \Pr \{ \mathbf{a}^i \} \mathcal{F}_{r_i}^{a_i} [ \mathbf{Z}^{\mathbf{a}^i} ] \quad (22)$$