ELSEVIER

Performance Evaluation 25 (1996) 173–191

PERFORMANCE EVALUATION An International Journal

Analysis of a general service nonexhaustive polling system using a heuristic combination method and pseudoconservation law

Chung-Ju Chang *, Lain-Chyr Hwang

Department of Communication Engineering and Center for Telecommunications Research, National Chiao Tung University, Hsinchu 30039, Taiwan, ROC

Received 1 May 1993; revised 1 November 1994

Abstract

This paper is devoted to the analysis of the mean waiting time for a polling system with general service order sequence and nonexhaustive service discipline. We obtain an expression for the mean waiting time in terms of a so-called residual cycle time and derive a pseudoconservation law (PCL) to help improve the accuracy of estimations of the mean waiting time. By multiplying the estimations of the residual cycle times for all stations by *equal* and *unequal* constants, we obtain two separate solutions for the mean waiting time. We furthermore propose a heuristic method that combines these separate mean waiting times into a final solution for the mean waiting time. Numerical examples show that our combination method generates accurate estimations for the mean waiting times in both cyclic and general cases over all traffic loads.

Keywords: Mean waiting time; General service order sequence; Nonexhaustive service discipline; Pseudoconservation law

1. Introduction

Polling systems have a wide range of applications. Takagi has made a very good survey in [1], which presents an overview of the art of polling model analysis, a comprehensive list of references, and some challenging problems.

In this paper, we investigate the mean waiting time of a polling system with *general* service order sequence and *nonexhaustive* service discipline. The nonexhaustive service discipline has been adopted in the implementation of token-ring networks because of its perceived fairness [2,3], and the general order of service sequence is frequently encountered in practice [4,5] because it is an alternative priority scheme that gives stations high priority by listing them more often in the polling sequence.

* Corresponding author. Tel.: 886-35-712121; fax: 886-35-710116.

SSDI 0166-5316(94)00053-0

Exact analysis of such a system is quite difficult; in general, only approximate analytical methods for estimating the mean waiting time for a *cyclic* service nonexhaustive polling system have been presented in the past. Srinivasan [2] proposed an approximate technique called Myopic Analysis of Cyclic Non-Exhaustive Service System (MACNESS) that appears to be very effective in obtaining the mean waiting times, in comparison with methods reported earlier. Boxma, Groenendijk and Weststrate [5] derived a pseudoconservation law (PCL) for systems in which stations were polled according to a general service order sequence and were served in a mixed service discipline. They, however, imposed a restriction that stations with nonexhaustive service discipline are served only once during a cycle.

The pseudoconservation law provides an exact expression for a weighted sum of the mean waiting time of every station in terms of known parameters and can be used to improve the effectiveness of the approximation for the mean waiting time. Here, we extend the concept of stochastic decomposition of the workload to general service nonexhaustive polling systems and derive a pseudoconservation law for such systems. Subsequently, we obtain an expression for the mean waiting time, in which there is a term called the residual cycle time that plays an important role in determining the accuracy of the estimation of the mean waiting time. We multiply the residual cycle time by an *equal* constant for all stations and apply the PCL to improve the approximation of the residual cycle time and then the mean waiting time. Such an approximation method can improve accuracy for *light* traffic loads, nevertheless, it cannot maintain the same level of accuracy for heavy traffic loads. Hence, as in [3], we also present another method in which we multiply each residual cycle time by an *unequal* constant to obtain accurate estimations for *heavy* traffic loads. Furthermore, to improve the estimation accuracy for *all* traffic loads, we heuristically combine the two solutions for the mean waiting times obtained by assuming equal and unequal weights on the residual cycle times into a final solution for the mean waiting time. Numerical examples shown below illustrate that our combination method generates accurate estimations of the mean waiting times for both cyclic and general cases.

This paper is organized as follows. In Section 2, the system model is described and the pseudoconservation law for the system is derived. In Section 3, the analytical approach for the mean waiting time is presented. In Section 4, several numerical examples are presented and discussed. We use the results of these examples to show the correctness of the derived PCL. Finally, concluding remarks are made in Section 5.

2. The pseudoconservation law

Consider a polling system that adopts a general order of service sequence and nonexhaustive service discipline. The polling system is assumed to have R stations and P stages, where the stage denotes the turn in the polling sequence [4]. We assume that customers arrive at station "r" according to an independent Poisson process with arrival rate λ_r, $1 \leq r \leq R$. We let the service time of a customer in station r, denoted by H_r, be generally distributed with mean h_r and the second moment $h_r^{(2)}$. An independent switchover time U_i between stage i and stage $i+1$ is also assumed to be generally distributed, $1 \leq i \leq P$. Let u_i and $u_i^{(2)}$ denote the mean and the second moment of U_i, respectively. In the following, all the indices corresponding to stages are in modulo-P arithmetic, where the result equals P if the remainder is zero.

The pseudoconservation law (PCL) for general service nonexhaustive polling systems can be obtained by

$$\sum_{i=1}^{P}(1-\pi_i^0)\left[\frac{h_{R_i}}{c_0}-\rho_{R_i}\sum_{j=i}^{a_i-1}\frac{u_j}{u}\right]\tilde{w}_i$$

$$=\frac{\rho}{2(1-\rho)}\sum_{r=1}^{R}\lambda_r h_r^{(2)}+\frac{\rho}{2u}\sum_{i=1}^{P}u_i^{(2)}+\sum_{i=1}^{P}(1-\pi_i^0)h_{R_i}\rho_{R_i}\sum_{j=i}^{a_i-1}\frac{u_j}{u}$$

$$+\sum_{i=1}^{P}\rho_{R_i}\sum_{j=i}^{a_i-1}\frac{u_j}{u}\sum_{k=i+1}^{j}\left[u_{k-1}+(1-\pi_k^0)h_{R_k}\right], \quad (1)$$

where R_i is the underlying station of stage i; a_i is the first stage after stage i to correspond to the same underlying station R_i; π_i^n, $n \geq 0$, is the probability of the number of customers in station R_i at a scan instant of stage i; $\tilde{w}_i$ is the mean waiting time of a customer served in stage i; ρ is the total load of the overall system, which is equal to the sum of the load of station r (denoted by ρ_r, $\rho_r = \lambda_r h_r$), $1 \leq r \leq R$; u is the total mean switchover time in an entire service sequence, which is equal to $\sum_{i=1}^{P} u_i$; c_0 is the mean whole cycle time, which is equal to $u/1-\rho$; $\sum$ is the sum, defined in [5, (2.8)], given by

$$\sum_{i=m}^{n} x_i = \begin{cases} \sum_{i=m}^{n} x_i & \text{if } m \leq n, \\ \sum_{i=m}^{P} x_i + \sum_{i=1}^{n} x_i & \text{if } m \geq n+1; \end{cases}$$

and $\sum$ defined similarly, but modified to make it also suitable for the case of $a_i = i+1$, is given by

$$\sum_{i=m}^{n} x_i = \begin{cases} 0 & \text{if } m = n+1, \\ \sum_{i=m}^{n} x_i & \text{otherwise.} \end{cases}$$

The detailed derivation of the PCL is given in Appendix A.

Our goal is to estimate the mean waiting time of a customer served in station r, denoted by w_r, but it is $\tilde{w}_i$ of stage i which is involved in PCL of (1). In order to relate the PCL with w_r, $1 \leq r \leq R$, we obtain $\tilde{w}_i$ in terms of w_{R_i} by

$$\tilde{w}_i = \frac{\lambda_{R_i} c_0 s_i}{(1-\pi_i^0)\sum_{\{k \mid R_k = R_i\}} s_k} \cdot \left(w_{R_i} + \frac{1}{\lambda_{R_i}}\right) - \frac{1}{\lambda_{R_i}}, \quad (2)$$

where s_i is the mean number of customers at station R_i when the server scans stage i. The derivation of (2) is presented in Appendix B and the expressions for π_i^0 and s_i (for stage i) are derived in Appendix C.

3. The mean waiting time

The mean waiting time of an arbitrarily selected customer served at station r consists of two components. The first is the mean residual cycle time measured from the labeled customer's arriving epoch to the next server's scan instant at station r; we denote the mean residual cycle time by $\check{c}_r$. The second is the mean time measured from the next scan instant of the server at station r to the epoch of the labeled customer's service-beginning; this mean time is equal to $q_r c_r^{+r}$ for nonexhaustive service discipline, where q_r $(= \lambda_r w_r)$ is the mean queue length in station r and c_r^{+r} is the mean conditional cycle time of station r (the cycle time of station r is the time between two successive scans at station r) if a customer is served at station r.

However, the cycle (vacation) time including (following) the service on the head of line (HOL) customer is different from the cycle (vacation) time including (following) the service on the other customers [2], because a labeled customer arriving during either service time or switchover time on the previous cycle of the HOL customer is likely to have a larger "interrupted" service time or switchover time (interrupted by the labeled customer). This is the so-called biasing effect [6, pp. 67, 68]. The larger interrupted service time or switchover time causes the cycle time that including the HOL customer to be larger than c_r^{+r}. We here replace the above $q_r c_r^{+r}$ by $q_r c_r^{+r} + d_r$ in our system, where d_r is the mean value used to make up this difference. Therefore, the mean waiting time of a customer in station r, w_r is equal to $(\check{c}_r + q_r c_r^{+r} + d_r)$. And using Little's formula, we have

$$(1 - \lambda_r c_r^{+r}) w_r = \check{c}_r + d_r. \tag{3}$$

Usually, $\check{c}_r$ is greater than d_r and so $\check{c}_r$ plays an important role in the accuracy of the estimation of w_r.

To make it easier for the readers to follow our analytical approach for finding d_r, $\check{c}_r$ and c_r^{+r}, we shall first introduce our notational conventions. We generally denote by $N_{x,y}^z$ the various kinds of (conditional) cycle times ($N = c$) or vacation times ($N = v$). The subscript "x" is used to represent the corresponding station or stage for the (conditional) cycle or vacation time; the subscript "y" is used to denote whether the server is in service at, on vacation from, or in switchover time from, a certain station or stage y when a labeled customer arrives at station r; and the superscript "z" is used to represent whether there is a customer served at a certain station or stage z during or before the conditional cycle or vacation time. When y is equal to $+r$, $-r$, $i+$ or $i-$, the server is in service at station r, on vacation from station r, in service at stage i or in switchover time from stage i, respectively, when the labeled customer arrives. When z is equal to $+r$, $i+$ or $i-$, there is a service at station r, a service at stage i or no service at stage i, respectively, during or before the conditional cycle or vacation time. Note that r is used to denote the index of the station, while i is used to denote the index of the stage; "+" or "−" is put before the index of station r and is put after the index of stage i; and y is dependent on the labeled customer's arrival epoch, but z does not correlate with the labeled customer. y or z will be omitted if the condition of y or z is not necessary, and x is equal to 0 when the notation represents a whole cycle. For example, the mean whole cycle time is denoted by c_0. Furthermore, we use "$\hat{}$", "$\check{}$" or "$\tilde{}$" above the notation "N" to stand for the (conditional) cycle or vacation time including or after the service on the HOL customer, the (conditional) "residual" cycle or vacation time, or the (conditional) cycle or vacation time corresponding to a "stage", respectively.

We derive d_r first. d_r is equal to the difference between $\hat{v}_{r,-r}$ and v_r^{+r} if the server is on vacation from station r and station r is not empty when the labeled customer arrives; it is equal to zero otherwise. $\hat{v}_{r,-r}$ is the mean conditional vacation time of station r after the HOL customer if the server is on vacation from station r when the labeled customer arrives and v_r^{+r} is the mean conditional vacation time of station r if there is a server at station r before this vacation. The probability that the server is on vacation from station r and station r is not empty is $1 - \xi_r - \rho_r$, where ξ_r is the probability that station r is empty and ρ_r is the load of station r. Thus, we have

$$d_r = (1 - \xi_r - \rho_r)(\hat{v}_{r,-r} - v_r^{+r}). \tag{4}$$

ξ_r in (4) can be approximated (similar to [2,7]) by

$$\xi_r \cong \frac{1 - \lambda_r c_r^{+r}}{1 - \lambda_r (v_r^{+r} - \check{v}_{r,-r})}. \tag{5}$$

$\check{v}_{r,-r}$ in (5) is the mean conditional residual vacation time of station r if the server is on vacation from station r when the labeled customer arrives.

In a cyclic system, [2,8] made an assumption that $\check{c}_r$ is thought of as equal for all r. However, Fuhrmann and Wang [3] reported that as the system becomes heavily loaded, there is a tendency for queues with relatively small values of c_r^{+r} to have larger values of $\check{c}_r$. In a general system, we will consider $\check{c}_r$ to be proportional to $\tilde{c}_i^{(2)}/2\tilde{c}_i$ if the server is on the pseudocycle time of stage i and $R_i = r$, where $\tilde{c}_i$ ($\tilde{c}_i^{(2)}$) is the first moment (the second moment) of the pseudocycle time of stage i. Here, the pseudocycle time of stage i is the time between the server's scan instants at stage i and stage a_i. The probability that the server is on the pseudocycle time of stage i is $\tilde{c}_i/c_0$. Thus $\check{c}_r$ can be expressed approximately as

$$\check{c}_r \cong K_r \sum_{\{i \mid R_i=r\}} \frac{\tilde{c}_i}{c_0} \cdot \frac{\tilde{c}_i^{(2)}}{2\tilde{c}_i}, \tag{6}$$

where

$$\tilde{c}_i^{(2)} \cong \sum_{k=i}^{a_i-1} \left\{(1-\pi_k^0)h_{R_k}^{(2)} + u_i^{(2)} - [(1-\pi_k^0)h_{R_k}]^2 - u_k^2\right\} + \left\{\sum_{k=i}^{a_i-1}[(1-\pi_k^0)h_{R_k} + u_k]\right\}^2, \tag{7}$$

and K_r is a variable that can be regarded as *equal* or *unequal* for all r. If K_r is regarded as equal for all r, we denote K_r by K_e. Then the mean waiting time of a customer in station r in this case, denoted by w_r^E instead of w_r, can be obtained by

$$w_r^E \cong \frac{1}{1-\lambda_r c_r^{+r}} \left[K_e \sum_{\{i \mid R_i=r\}} \frac{\tilde{c}_i^{(2)}}{2c_0} + (1-\xi_r-\rho_r)(\hat{v}_{r,-r} - v_r^{+r}) \right]. \tag{8}$$

We can use the PCL to obtain the constant K_e and then the mean waiting time w_r^E. On the other hand, if K_r is regarded as unequal for different stations r, we let $K_r = K_u/\Psi_r$, where K_u is a constant and Ψ_r varies for different r. As in [3] for the cyclic case, we could let Ψ_r be $m_r c_r^{+r}$, where m_r is the total number of polls of station r in a polling sequence and c_r^{+r} is the mean conditional cycle time of station r if there is a service at station r; instead, however, we conducted a substantial number of experiments to determine a better Ψ_r. Our experiments showed that Ψ_r is better chosen as $m_r\check{c}_r'$ ($\Psi_r = m_r\check{c}_r'$), where $\check{c}_r'$ is the mean residual cycle time obtained via an alternative approach as follows.

When the labeled customer arrives at station r, $\check{c}_r'$ is equal to the mean conditional residual vacation time $\check{v}_{r,-r}$ if the server is on vacation from station r, and $\check{c}_r'$ is equal to $(h_r^{(2)}/2h_r) + \hat{v}_{r,+r}$ if the server is serving the HOL customer in station r, where $h_r^{(2)}/2h_r$ is the residual service time of the served customer and $\hat{v}_{r,+r}$ is the mean conditional vacation time of station r after the HOL customer if the server is in service at station r when the labeled customer arrives. The probability that the server is on vacation from station r is $1-\rho_r$, while the probability that the server is in service at station r is ρ_r. As a result, $\check{c}_r'$ is equal to $(1-\rho_r)\check{v}_{r,-r} + \rho_r((h_r^{(2)}/2h_r) + \hat{v}_{r,+r})$. This idea is intuitively correct, because the mean residual cycle time of a single-poll station is approximately twice the mean residual cycle time of a two-poll station. In other words, the mean residual cycle time multiplied by the number of polling times should be approximately equal for all stations. Therefore, if the estimated $m_r\check{c}_r'$ is larger (smaller) for a certain r than for others, then the $\check{c}_r$ of that station r in (6) is overestimated (underestimated) and K_r should be made smaller (larger) to compensate for the overestimation (underestimation) by letting it be a constant K_{u} divided by $m_r\check{c}_r'$. Therefore, $\check{c}_r$ in (6) can be expressed as

$$\check{c}_r \cong \frac{K_{\mathrm{u}}}{2c_0}\left\{m_r\left[(1-\rho_r)\check{v}_{r,-r}+\rho_r\left(\frac{h_r^{(2)}}{2h_r}+\check{v}_{r,+r}\right)\right]\right\}^{-1}\cdot\sum_{\{i\,|\,R_i=r\}}\tilde{c}_i^{(2)}. \tag{9}$$

Therefore, in this case the mean waiting time of a customer in station r, denoted by w_r^{U} instead of w_r, is given by

$$w_r^{\mathrm{U}} \cong \frac{1}{1-\lambda_r c_r^{+r}}\left\{\frac{K_{\mathrm{u}}}{2c_0}\left\{m_r\left[(1-\rho_r)\check{v}_{r,-r}+\rho_r\left(\frac{h_r^{(2)}}{2h_r}+\hat{v}_{r,+r}\right)\right]\right\}^{-1}\cdot\sum_{\{i|R_i=r\}}\tilde{c}_i^{(2)}\right.$$
$$\left.+\left[(1-\xi_r-\rho_r)(\hat{v}_{r,-r}-v_r^{+r})\right]\right\}. \tag{10}$$

We can use the PCL in a similar manner to obtain the constant K_{u} and then the estimated mean waiting time w_r^{U}.

Furthermore, we heuristically combine the mean waiting times w_r^{E} and w_r^{U} by letting

$$w_r^{\mathrm{C}} = (1-\rho)w_r^{\mathrm{E}} + \rho w_r^{\mathrm{U}}, \tag{11}$$

where w_r^{C} denotes the mean waiting time of station r in this combination method. We add weights of $(1-\rho)$ and ρ on w_r^{E} and w_r^{U}, respectively, because the estimation of w_r^{E} is more accurate under light traffic conditions, while the estimation of w_r^{U} is more accurate under heavy traffic conditions. The unknown parameters c_r^{+r}, v_r^{+r}, $\check{v}_{r,-r}$, $\check{v}_{r,+r}$ and $\hat{v}_{r,-r}$ in the above derivation can be obtained in Appendix D.

4. Numerical examples and discussion

To clarify our analytical approach, we first state the numerical algorithm for the analysis below.

Step 1 [Set initial values of π_i^0]

- Set initial values of π_i^0 by letting

$$\pi_i^0 = 1 - \frac{\lambda_r c_0}{m_r} \quad \forall i \text{ such that } R_i = r,\quad 1 \le r \le R.$$

Step 2 [Find new values of π_i^0]

- Obtain $A_{r_{j+1}}^{r_j+}(z)$ and $A_{r_{j+1}}^{r_j-}(z)$ (defined in Appendix C and r_j denotes the stage corresponding to the jth poll of station r in the polling sequence) in (C.4).
- Generate a set of m_r simultaneous equations derived from (C.2) and (C.3) for a given r, $1 \le r \le R$.
- Solve the m_r simultaneous equations to find new values of π_i^0, denoted by $\hat{\pi}_i^0$, and normalize the solution by multiplying by a constant so as to match

$$m_r - \lambda_r c_0 = \sum_{\{i\,|\,Ri=r\}} \hat{\pi}_i^0, \quad 1 \le r \le R.$$

- Let $\Delta_i = |\pi_i^0 - \hat{\pi}_i^0|$

  $\pi_i^0 = \hat{\pi}_i^0$

  IF ($\Delta_i <$ a pre-determined threshold for all i) THEN

  Terminate the iteration.

  ELSE

  Repeat Step 2.

Step 3 [Find mean queue lengths at scan instants]
- Obtain the $\pi_{r_j}^n$, $n \geq 1$, $1 \leq j \leq m_r$, $1 \leq r \leq R$, from (C.2) recursively.
- Obtain the mean queue length at scan instant of stage i, $s_i = \sum_{n=1}^{\infty} n \cdot \pi_i^n$ for $1 \leq i \leq P$.

Step 4 [Find mean conditional cycle times and vacation times]
- Find means of conditional cycle times and conditional vacation times in Appendix D. (Note that these mean conditional cycle times in (D.3), (D.6) and (D.9) are evaluated by an iterative method similar to that used in (C.6).)

Step 5 [Obtain the mean waiting times w_r^{E}, w_r^{U} and w_r^{C}]
- Substitute $w_{R_i}^{\mathrm{E}}$ of (8) ($w_{R_i}^{\mathrm{U}}$ of (10)), π_i^0 and s_k for all k such that R_k is equal to R_i into (2) to express $\tilde{w}_i$ in terms of K_{e} (K_{u}).
- Use (1) of PCL to find K_{e} and K_{u}, and then obtain w_r^{E} and w_r^{U} for all r.
- Obtain the mean waiting time of station r, w_r^{C} from (11).

To assess the accuracy of our approximate analysis, we here show examples of cyclic systems, introduced in [2,3], and general systems; we conduct a system simulation and define an estimation error as

$$\text{estimation error} = \sum_{r=1}^{R} \frac{\rho_r |(w_r - w_r^{\mathrm{S}})|}{\rho w_r^{\mathrm{S}}} \times 100\%,$$

where w_r^{S} is the simulation result for the mean waiting time of station r and w_r is our analytical result for the mean waiting time of station r. In Tables 1–6, the simulation results of the mean waiting times w_r^{S} are indicated with 95% confidence intervals [10, Section 6.3]; "Srinivasan", "B&M" and "F&W" denote the results of the approximation methods from Srinivasan [2], Boxma and Meister [8], and Fuhrmann and Wang [3], respectively; and "C&H" denotes the approximation results from our analytical method.

Our first example (Example 1) is an asymmetric cyclic service nonexhaustive polling system with the same arrival rates. The parameters of the system are assumed to be $R = P = 16$; $\lambda_r = 1/16$ for all r; all service time distributions are exponential with $h_1 = h_7$, $h_r = h_1/3$ for any $r \neq 1, 7$; and all switchover times are equal to 0.05. Table 1 shows that, under light and medium loads, Srinivasan, B&M and w_r^{E} have very accurate results; w_r^{C} is good; but w_r^{U} is somewhat worse. Under heavy load, w_r^{C} and w_r^{U} have very accurate results; next, in order, are Srinivasan, w_r^{E}, F&W and B&M.

Our next example (Example 2) is an asymmetric cyclic service nonexhaustive polling system with the same service rates. The parameters of the system are assumed to be $R = P = 16$; $\lambda_r = 0.16$, $r = 1, 2, 3, 4$; $\lambda_r = 0.03$, $r = 5, 6, \ldots, 16$; all service times are exponentially distributed with identical mean h_r for all r; and all switchover times are equal to 0.05. Table 2 shows the results of the mean waiting times of stations. Under light load, all the methods are very accurate except w_r^{U}. Under medium load, Srinivasan and w_r^{C} are very accurate; w_r^{E} and w_r^{U} are good; but B&M is somewhat worse. Under heavy load, F&W is the best; next, in order, are Srinivasan, w_r^{U} and w_r^{C}, but there is only a slight difference between these three methods;

Table 1
The mean waiting times for an asymmetric cyclic system with the same arrival rates (Example 1) (see also [2, Table 7] and [3, Case 3]). (Simulation result w_r^S is indicated with 95% confidence interval)

Method	Station	Load		
		0.3	0.5	0.8
w_r^S	1	0.833 ± 0.007	1.733 ± 0.014	8.960 ± 0.047
	2–6	0.796 ± 0.006	1.600 ± 0.012	8.004 ± 0.041
	7	0.825 ± 0.006	1.718 ± 0.014	8.950 ± 0.047
	8–16	0.796 ± 0.006	1.593 ± 0.012	7.993 ± 0.041
w_r^E, w_r^U	1	0.830, 0.781	1.729, 1.567	9.526, 8.870
	2–6	0.798, 0.819	1.596, 1.664	7.774, 8.038
	7	0.830, 0.870	1.729, 1.570	9.518, 8.937
	8–16	0.798, 0.819	1.596, 1.664	7.774, 8.040
@estimation error(%)		0.321, 3.777	0.278, 5.735	3.851, 0.543
w_r^C	1	0.815	1.648	9.001
	2–6	0.804	1.630	7.985
	7	0.815	1.649	9.053
	8–16	0.804	1.630	7.987
@estimation error(%)		1.210	2.852	0.334
Srinivasan	1	0.835	1.752	9.349
	2–6	0.796	1.586	7.900
	7	0.835	1.752	9.340
	8–16	0.796	1.586	7.850
@estimation error(%)		0.218	0.878	2.153
B&M (F&W)	1	0.831	1.742	10.060 (7.950)
	2–6	0.797	1.590	7.540 (8.480)
	7	0.831	1.742	10.060 (7.950)
	8–16	0.797	1.590	7.540 (8.480)
@estimation error(%)		0.233	0.528	7.702 (7.595)

w_r^E is worse; and B&M is the worst. From these two examples of cyclic systems, we can conclude that our results are as good as Srinivasan's and most of the estimation errors are within 5%.

We now turn to examples of polling systems with general order of service sequence. To assess the validity of our analysis, we here illustrate three general polling systems with different general orders of service sequence, service time distributions and switchover time distributions. The first example (Example 3) has system parameters $R = 5$, $P = 10$; service order sequence: {1 2 3 1 4 1 5 3 1 4}; λ_r: (0.1 0.02 0.05 0.02 0.01), $r = 1, 2, \ldots, 5$; all service times are exponentially distributed with identical means h_r for all r; and all switchover times are equal to 0.1. Table 3 lists the mean waiting times of w_r^S, w_r^E, w_r^U and w_r^C. w_r^E is more accurate than w_r^U under light and medium loads, w_r^U is more accurate than w_r^E under heavy load and w_r^C is more effective than w_r^E and w_r^U under all traffic loads.

The next example (Example 4) is the same as Example 3 except that the service time distribution is hyperexponential with identical means h_r for all r and the switchover time is exponential with mean 0.1; the hyperexponential distribution consists of three exponential distributions with mean $= 0.5h_r$, h_r and $1.5h_r$ in equal probability. The results of w_r^S, w_r^E, w_r^U and w_r^C are listed in Table 4. As in Table 3, w_r^C is more effective than w_r^E and w_r^U under all traffic loads. The higher accuracy of w_r^C compared with w_r^E and w_r^U

Table 2
The mean waiting times for an asymmetric cyclic system with the same service rates (Example 2) (see also [2, Table 12] and [3, Case 4]). (Simulation result w_r^S is indicated with 95% confidence interval)

Method	Station	Load		
		0.3	0.5	0.8
w_r^S	1–4	0.902 ± 0.004	1.926 ± 0.009	18.321 ± 0.055
	5–16	0.716 ± 0.006	1.262 ± 0.011	3.570 ± 0.019
w_r^E, w_r^U	1–4	0.899, 0.908	1.907, 1.951	17.618, 17.853
	5–16	0.717, 0.704	1.276, 1.214	4.173, 4.002
@estimation error(%)		0.263, 1.029	1.031, 2.200	8.536, 5.991
w_r^C	1–4	0.902	1.929	17.806
	5–16	0.713	1.245	4.036
@estimation error(%)		0.151	0.585	6.498
Srinivasan	1–4	0.901	1.922	17.901
	5–16	0.714	1.255	3.967
@estimation error(%)		0.172	0.333	5.471
B&M (F&W)	1–4	0.897	1.884	16.870 (18.500)
	5–16	0.720	1.307	3.140 (3.530)
@estimation error(%)		0.556	2.679	9.405 (1.029)

Table 3
The mean waiting times for an asymmetric general system with exponential service time and deterministic switchover time (Example 3). (Simulation result w_r^S is indicated with 95% confidence interval)

Mean waiting time	Station	Load		
		0.3	0.5	0.8
w_r^S	1	0.859 ± 0.009	2.763 ± 0.020	18.835 ± 0.072
	2	1.545 ± 0.027	4.380 ± 0.069	24.803 ± 0.208
	3	1.096 ± 0.014	3.408 ± 0.035	23.217 ± 0.127
	4	1.013 ± 0.020	2.747 ± 0.043	10.581 ± 0.082
	5	1.453 ± 0.035	3.937 ± 0.084	16.517 ± 0.175
w_r^E, w_r^U	1	0.896, 0.682	3.075, 2.412	20.158, 17.762
	2	1.446, 1.973	3.757, 5.255	22.984, 27.509
	3	1.056, 1.156	3.165, 3.573	20.034, 22.445
	4	0.977, 1.026	2.733, 2.897	13.376, 13.587
	5	1.392, 1.863	3.461, 4.663	17.411, 19.403
@estimation error(%)		4.263, 15.988	9.498, 11.036	10.586, 8.485
w_r^C	1	0.832	2.743	18.241
	2	1.604	4.506	26.604
	3	3.086	3.369	21.963
	4	0.992	2.815	13.545
	5	1.534	4.062	19.005
@estimation error(%)		2.700	1.345	7.207

Table 4
The mean waiting times for an asymmetric general system with hyperexponential service time and exponential switchover time (Example 4). (Simulation result w_r^S is indicated with 95% confidence interval)

Mean waiting time	Station	Load		
		0.3	0.5	0.8
w_r^S	1	0.997 ± 0.010	3.233 ± 0.025	21.945 ± 0.084
	2	1.707 ± 0.032	4.934 ± 0.079	27.860 ± 0.233
	3	1.263 ± 0.017	4.016 ± 0.043	26.837 ± 0.150
	4	1.141 ± 0.023	3.164 ± 0.050	11.628 ± 0.090
	5	1.685 ± 0.044	4.514 ± 0.100	18.089 ± 0.194
w_r^E, w_r^U	1	1.064, 0.791	3.597, 2.784	23.503, 20.429
	2	1.604, 2.274	4.219, 6.037	26.151, 31.584
	3	1.215, 1.341	3.642, 4.137	23.153, 25.962
	4	1.124, 1.194	3.142, 3.365	15.315, 16.279
	5	1.543, 2.148	3.878, 5.362	19.496, 22.689
@estimation error(%)		5.483, 17.052	10.171, 11.516	11.155, 10.876
w_r^C	1	0.982	3.191	21.044
	2	1.805	5.128	30.497
	3	1.253	3.890	25.400
	4	1.145	3.253	16.086
	5	1.725	4.620	22.050
@estimation error(%)		1.654	2.237	9.267

is because w_r^E and w_r^U compensate for each other—one is overestimated and the other is underestimated. From the two examples, we find the estimation errors of w_r^C are within 5% under light and medium loads and within 10% under heavy load.

We finally examine another polling system with more stations (Example 5), for which $R = 8$, $P = 15$; service order sequence: {1 2 3 4 5 1 6 7 1 8 2 4 1 5 7}; λ_r: (0.1 0.03 0.01 0.03 0.02 0.01 0.04 0.01), $r = 1, 2, \ldots, 8$; all service times are deterministic and equal to 1; and all switchover times are exponentially distributed with identical means 0.1. Table 5 lists the mean waiting times of w_r^S, w_r^E, w_r^U and w_r^C. w_r^E is more accurate than w_r^U under light and medium loads, w_r^U is better than w_r^E under heavy load and w_r^C is more effective than w_r^E and w_r^U under all traffic loads.

Additionally, Table 6 shows the simulation results of the mean waiting times of stages and the RHS and LHS of the PCL for Example 3. The table verifies the validity of the PCL for the general order of service sequence. As a matter of fact, all other examples are also checked, but the results are not shown here.

5. Concluding remarks

In this paper, we have presented a new method for analyzing the mean waiting time for a nonexhaustive polling system with general order of service sequence. We first derived a PCL to help improve the precision of estimations of the mean waiting time and obtained the mean waiting time in terms of a residual cycle time. We multiplied the estimated mean residual cycle time of each station by an equal or unequal constant and used the PCL to find the multiplication constants and then the corresponding solutions for the mean waiting times. The method to derive the mean waiting time obtained by multiplying by an equal constant

Table 5
The mean waiting times for an asymmetric general system with deterministic service time and exponential switchover time (Example 5). (Simulation result w_r^{S} is indicated with 95% confidence interval)

Mean waiting time	Station	Load		
		0.3	0.5	0.8
w_r^{S}	1	0.621 ± 0.003	1.513 ± 0.007	9.709 ± 0.030
	2	0.916 ± 0.008	1.958 ± 0.016	8.746 ± 0.044
	3	1.483 ± 0.018	2.930 ± 0.037	12.110 ± 0.096
	4	0.893 ± 0.007	1.943 ± 0.016	8.702 ± 0.044
	5	0.896 ± 0.009	1.888 ± 0.018	7.327 ± 0.042
	6	1.482 ± 0.018	2.925 ± 0.037	12.205 ± 0.098
	7	0.914 ± 0.007	2.034 ± 0.015	10.929 ± 0.052
	8	1.469 ± 0.018	2.924 ± 0.037	12.238 ± 0.099
w_r^{E}, w_r^{U}	1	0.643, 0.549	1.617, 1.400	9.423, 9.203
	2	0.901, 0.909	1.909, 1.945	9.086, 8.983
	3	1.464, 1.684	2.830, 3.281	12.570, 12.909
	4	0.886, 0.917	1.890, 1.970	9.049, 9.165
	5	0.891, 0.881	1.852, 1.823	8.048, 7.472
	6	1.464, 1.691	2.829, 3.301	12.566, 12.997
	7	0.908, 0.944	1.979, 2.085	10.337, 10.865
	8	1.464, 1.686	2.829, 3.287	12.569, 12.920
@estimation error(%)		1.938, 7.433	4.343, 5.414	4.154, 4.042
w_r^{C}	1	0.615	1.509	9.248
	2	0.903	1.927	9.004
	3	1.530	3.055	12.841
	4	0.895	1.930	9.141
	5	0.888	1.838	7.587
	6	1.532	3.065	12.911
	7	0.919	2.032	10.759
	8	1.530	3.058	12.850
@estimation error(%)		1.218	1.173	4.065

Table 6
Simulation results of the mean waiting times of stages and the LHS and RHS of pseudoconservation law (Example 3)

Load	Stage	Mean waiting time of stage					PCL	
							LHS	RHS
0.3	1–5	0.855	1.545	1.102	0.878	1.044	0.300	0.298
	6–10	0.901	1.453	1.089	0.811	0.981		
0.5	1–5	2.857	4.380	3.423	2.691	2.820	1.499	1.506
	6–10	2.956	3.937	3.393	2.622	2.671		
0.8	1–5	19.771	24.803	23.139	17.956	10.662	13.763	13.885
	6–10	19.709	16.517	23.298	18.227	10.495		

is similar to that of Srinivasan, and the approach to derive the mean waiting time obtained by multiplying by an unequal constant is modified from the method of Fuhrmann and Wang. We further combined the two mean waiting times into a final solution. Five numerical examples were presented. We found that the estimations of the mean waiting time obtained by multiplying by an equal constant are generally good for light and medium loads; the estimations of the mean waiting time obtained by multiplying by unequal constants are generally good for heavy loads; and the estimations of the mean waiting time obtained by our combination method are effective for all traffic loads. The estimation errors of our combination method are in general within 5% for both cyclic and general cases and have almost the same accuracy as those of Srinivasan in the cyclic case, but Srinivasan's approach cannot be used directly in the general order of service approximations.

Acknowledgements

This study was supported by National Science Council, Taiwan, Republic of China, under contract number NSC 83-0404-E009-002.

Appendix A. The derivation of a pseudoconservation law for a general service nonexhaustive system

We first define ϕ_r^i as the mean amount of work in station r as the server departs from stage i and $t_i = (1-\pi_i^0)h_{R_i}$ as the mean visit time at stage i, for $1 \le r \le R$ and $1 \le i \le P$. Here, the work in station r at a time epoch denotes the total service time of all customers in station r. If the time epoch is just on the service time of station r, the work includes the residual service time.

We start from the following equation, derived from [11, (3.4)–(3.6)]:

$$\sum_{r=1}^{R}\rho_r w_r = \frac{\rho}{2(1-\rho)}\sum_{r=1}^{R}\lambda_r h_r^{(2)} + \frac{\rho}{2\cdot u}\sum_{i=1}^{P}u_i^{(2)} + \sum_{j=1}^{P}\frac{u_j}{u}\sum_{r=1}^{R}\phi_r^j. \tag{A.1}$$

Although [11] studied the cyclic service case, [11, (3.4)–(3.6)] are still suitable for our system because they do not involve the service order. $\sum_{j=1}^{P} u_j \sum_{r=1}^{R} \phi_r^j$ in the last term of (A.1) can be rewritten as

$$\begin{aligned}\sum_{j=1}^{P}u_j\sum_{r=1}^{R}\phi_r^j &= \sum_{r=1}^{R}\sum_{j=1}^{P}u_j\phi_r^j = \sum_{r=1}^{R}\sum_{i=1}^{m_r}\sum_{j=r_i}^{r_{i+1}-1}u_j\phi_r^j \\ &= \sum_{r=1}^{R}\sum_{i=1}^{m_r}\sum_{j=r_i}^{r_{i+1}-1}u_j\left\{\phi_r^{r_i} + \rho_r\sum_{k=r_i+1}^{j}(u_{k-1}+t_k)\right\},\end{aligned} \tag{A.2}$$

where we decompose ϕ_r^j, j from r_i to $r_{i+1}-1$ into two parts: the mean work left in station r at the server's departure from stage r_i, $\phi_r^{r_i}$ and the mean work arriving at station r during the time interval between the server's departures from stage r_i and stage j, $\rho_r\sum_{k=r_i+1}^{j}[u_{k-1}+t_k]$. If we replace r_i by i so as to make $\sum_{r=1}^{R}\sum_{i=1}^{m_r}$ equivalent to $\sum_{i=1}^{P}$ and replace t_k by $(1-\pi_k^0)h_{R_k}$, (A.2) becomes

$$\sum_{j=1}^{P} u_j \sum_{r=1}^{R} \phi_r^j = \sum_{i=1}^{P} \sum_{j=1}^{a_i-1} u_j \left\{ \phi_{R_i}^i + \rho_{R_i} \sum_{k=i+1}^{j} [u_{k-1} + (1-\pi_k^0) h_{R_k}] \right\}. \tag{A.3}$$

With respect to the nonexhaustive service discipline, similar to [11, (3.13), (3.14)], $\phi_{R_i}^i$ in (A.3) is $(1 - \pi_i^0)\lambda_{R_i}(\tilde{w}_i + h_{R_i})h_{R_i}$ and we know that $\sum_{r=1}^{R} \rho_r w_r = \sum_{i=1}^{P} \varphi_i \tilde{w}_i$, where φ_i is the load of stage i. Consequently, the PCL for general service nonexhaustive polling systems is obtained by

$$\begin{aligned} &\sum_{i=1}^{P} (1-\pi_i^0) \left[\frac{h_{R_i}}{c_0} - \rho_{R_i} \sum_{j=i}^{a_i-1} \frac{u_j}{u} \right] \tilde{w}_i \\ &= \frac{\rho}{2(1-\rho)} \sum_{r=1}^{R} \lambda_r h_r^{(2)} + \frac{\rho}{2 \cdot u} \sum_{i=1}^{P} u_i^{(2)} + \sum_{i=1}^{P} (1-\pi_i^0) h_{R_i} \rho_{R_i} \sum_{j=i}^{a_i-1} \frac{u_j}{u} \\ &\quad + \sum_{i=1}^{P} \rho_{R_i} \sum_{j=i}^{a_i-1} \frac{u_j}{u} \sum_{k=i+1}^{j} \left[u_{k-1} + (1-\pi_k^0) h_{R_k} \right]. \end{aligned} \tag{A.4}$$

(A.4) can be simplified to become [11, (3.22)] for a system with a cyclic service order sequence and nonexhaustive service discipline and can be extended to a system with a general service order sequence and mixed service disciplines (exhaustive, gated and nonexhaustive), where stages corresponding to the same station can have different service disciplines. we have verified the validity of (A.4) by simulations in Table 6.

Appendix B. The derivation of $\tilde{w}_i$ in terms of w_{R_i}

Denote the number of customers waiting at station R_i when the server scans stage i by S_i with mean s_i, $1 \le i \le P$, the probability of $\{S_i = n\}$ by π_i^n, $n \ge 0$, and the probability generating function (pgf) of S_i by $S_i(z)$, where $S_i(z) \equiv \sum_{n=0}^{\infty} \pi_i^n z^n$. The number of customers left in station R_i as a customer departs from that station at stage i is equal to the number of customers arriving at station R_i during the waiting time of the customer served at stage i and during the customer's service time H_{R_i}; it is also equal to the number of customers in station R_i as the service begins minus one plus those customers arriving at station R_i during the service time of the customer. Consequently, the pgf of the number of customers arriving at station R_i during the waiting time of the customer served at stage i, denoted by $\tilde{W}_i(z)$, is equal to $[(S_i(z) - \pi_i^0)/(1 - \pi_i^0)] \cdot z^{-1}$. Because the arrival process is a Poisson process, the Laplace Stieltjes transform of the distribution function of the waiting time at stage i is $\tilde{W}_i(1 - (s/\lambda_{R_i}))$. We differentiate $[-\tilde{W}_i(1 - (s/\lambda_{R_i}))]$ and let $s = 0$ to obtain the mean waiting time $\tilde{w}_i$, given by

$$\tilde{w}_i = \frac{1}{\lambda_{R_i}} \cdot \left[\frac{s_i}{1 - \pi_i^0} - 1 \right]. \tag{B.1}$$

On the other hand, w_r can be regarded as the weighted sum of $\tilde{w}_i$ for all i such that $R_i = r$; w_r is expressed as

$$w_r = \frac{1}{\lambda_r} \sum_{\{i \mid R_i = r\}} \beta_i \tilde{w}_i, \tag{B.2}$$

where β_i is the departure rate of customers from stage i and is given by

$$\beta_i = (1 - \pi_i^0)/c_0. \tag{B.3}$$

Then, from (B.1)–(B.3) and noting that $\sum_{\{i \mid R_i = r\}}(1 - \pi_i^0) = \lambda_r c_0$, w_r can be obtained by

$$w_r = \frac{1}{\lambda_r^2 c_0} \cdot \sum_{\{i \mid R_i = r\}} s_i - \frac{1}{\lambda_r}. \tag{B.4}$$

From (B.1) and (B.4), $\tilde{w}_i$ can be expressed in terms of w_{R_i} as

$$\tilde{w}_i = \frac{\lambda_{R_i} c_0 s_i}{(1 - \pi_i^0)\sum_{\{k \mid R_k = R_i\}} s_k} \cdot \left(w_{R_i} + \frac{1}{\lambda_{R_i}} \right) - \frac{1}{\lambda_{R_i}}. \tag{B.5}$$

Appendix C. The estimations of π_i^0 and s_i

Denote $\tilde{C}_i^{k+}$ ($\tilde{C}_i^{k-}$) to be the conditional pseudocycle time of stage i if there is a service (no service) at stage k and its mean is $\tilde{c}_i^{k+}$ ($\tilde{c}_i^{k-}$); denote $A_{r_{j+1}}^{r_j+}$ ($A_{r_{j+1}}^{r_j-}$) to be the number of customers arriving at station r during $\tilde{C}_{r_{j+1}}^{r_j+}$ ($\tilde{C}_{r_{j+1}}^{r_j-}$) and its pgf is $A_{r_{j+1}}^{r_j+}(z)$ ($A_{r_{j+1}}^{r_j-}(z)$); and denote α_i^{k+} (α_i^{k-}) to be the probability of stage i having customers at a scan instant of stage i if there was a service (no service) at stage k [9].

$S_{r_{j+1}}$ is equal to $A_{r_{j+1}}^{r_j-}$ if $S_{r_j} = 0$ or equal to the sum of ($S_{r_j} - 1$) and $A_{r_{j+1}}^{r_j+}$ if $S_{r_j} > 0$. This relation can be expressed as

$$S_{r_{j+1}}(z) = \pi_{r_j}^0 \cdot A_{r_{j+1}}^{r_j-}(z) + \sum_{m=1}^{\infty} \pi_{r_j}^m \cdot z^{m-1} \cdot A_{r_{j+1}}^{r_j+}(z). \tag{C.1}$$

By comparing the coefficients of z^{n-1} on both sides of (C.1), we have

$$\pi_{r_j}^n = \frac{\pi_{r_{j+1}}^{n-1} - \sum_{m=1}^{n-1}\left\{\pi_{r_j}^m \cdot \Pr\left[A_{r_{j+1}}^{r_j+} = n - m\right]\right\} - \pi_{r_j}^0 \cdot \Pr\left[A_{r_{j+1}}^{r_j-} = n - 1\right]}{\Pr\left[A_{r_{j+1}}^{r_j+} = 0\right]}. \tag{C.2}$$

Substituting $n = 1, n = 2, \ldots$, into (C.2) and performing a simple algebraic manipulation, we obtain $\pi_{r_j}^n$ expressed in terms of $\pi_{r_1}^0, \pi_{r_2}^0, \ldots, \pi_{r_j}^0, \ldots, \pi_{r_{m_r}}^0$ for $n \geq 1$. Again, substituting $\pi_{r_j}^n$ into m_r summability-to-one criteria given by

$$\sum_{n=0}^{\infty} \pi_{r_j}^n = 1 \quad \text{for } 1 \leq j \leq m_r, \tag{C.3}$$

we obtain a total of m_r simultaneous equations consisting of $\pi_{r_1}^0, \pi_{r_2}^0, \ldots, \pi_{r_{m_r}}^0$. The $\pi_{r_j}^0$, $1 \leq j \leq m_r$, can thus be obtained by solving these m_r simultaneous equations via numerical algorithm if $A_{r_{j+1}}^{r_j+}(z)$ and $A_{r_{j+1}}^{r_j-}(z)$ are known.

Under the assumption that the visit times of stages are independent, $A^{r_j+}_{r_{j+1}}(z)$ and $A^{r_j-}_{r_{j+1}}(z)$ can be obtained by

$$\begin{aligned} A^{r_j+}_{r_{j+1}}(z) &= H^r_r(z) \cdot U^r_{r_j}(z) \cdot \prod_{i=r_j+1}^{r_{j+1}-1} \left\{[\alpha_i^{r_j+} \cdot H^r_{R_i}(z) + 1 - \alpha_i^{r_j+}] \cdot U^r_i(z)\right\}, \\ A^{r_j-}_{r_{j+1}}(z) &= U^r_{r_j}(z) \cdot \prod_{i=r_j+1}^{r_{j+1}-1} \left\{[\alpha_i^{r_j-} \cdot H^r_{R_i}(z) + 1 - \alpha_i^{r_j-}] \cdot U^r_i(z)\right\}, \end{aligned} \tag{C.4}$$

where $H^r_{r'}(z)$ denotes the pgf of the number of customers arriving at a certain station r during $H_{r'}$, $1 \le r, r' \le R$; $U^r_i(z)$ denotes the pgf of the number of customers arriving at station r during U_i, where $1 \le r \le R$ and $1 \le i \le P$; and $\prod$ is defined as

$$\prod_{i=m}^{n} x_i = \begin{cases} \prod_{i=m}^{n} x_i & \text{if } m \le n \text{ and } (n-m+1) < P, \\ 1 & \text{if } m = n+1, \\ \prod_{i=m}^{P} x_i \cdot \prod_{i=1}^{n} x_i & \text{if } m > n+1. \end{cases}$$

If we regard our system as an approximately cyclic polling system with P independent stages and assume the arrival rate for stage i is equal to the departure rate for stage i, then we have

$$\alpha_i^{k+} = \min(\beta_i \bar{c}_i^{k+}, 1), \tag{C.5a}$$

$$\alpha_i^{k-} = \frac{1}{\pi_k^0}[1 - \pi_i^0 - (1 - \pi_k^0)\alpha_i^{k+}], \tag{C.5b}$$

where

$$\bar{c}_i^{k+} = u + h_{R_k} + \min[(\lambda_{R_k} - \beta_k)\bar{c}_i^{k+}, m_{R_k} - 1]h_{R_k} + \sum_{r \ne R_k} \min(\lambda_r \bar{c}_i^{k+}, m_r)h_r. \tag{C.6}$$

Eq. (C.5) comes from the equation

$$(1 - \pi_k^0) \cdot \alpha_i^{k+} + \pi_k^0 \cdot \alpha_i^{k-} = 1 - \pi_i^0 \tag{C.7}$$

and from [9, (9d) and (9e)] rewritten as

$$\alpha_i^{k+} = \min(\beta_i \bar{c}_i^{k+}, 1), \tag{C.8a}$$

$$\alpha_i^{k-} = \min(\beta_i \bar{c}_i^{k-}, 1). \tag{C.8b}$$

Eq. (C.7) always holds under any given π_i^0 and π_k^0. However, its validity is violated if $\beta_i \bar{c}_i^{k+}$ or $\beta_i \bar{c}_i^{k-}$ in (C.8) are greater than 1. Because $\beta_i \bar{c}_i^{k-} > 1$ will certainly result in $\beta_i \bar{c}_i^{k+} > 1$, we modify (C.8b) to be (C.5b) in our analysis in order to preserve the validity of (C.7). Note that the mean conditional cycle time $\bar{c}_i^{k+}$ in (C.6) is computed by an iterative method.

From (C.2)–(C.6), we find that there is a recursive relationship among π_i^0, $A^{r_j+}_{r_{j+1}}(z)$ and $A^{r_j-}_{r_{j+1}}(z)$. The solutions for π_i^0, $A^{r_j+}_{r_{j+1}}(z)$ and $A^{r_j-}_{r_{j+1}}(z)$ can be found by an iterative method, which is described in Section 4. Consequently, we substitute the resulting $\pi_{r_j}^0$'s and the distributions of $A^{r_j+}_{r_{j+1}}$'s and $A^{r_j-}_{r_{j+1}}$'s into (C.2) to obtain $\pi_{r_j}^n$, for all r, j and n, and then the mean queue lengths at scan instants s_i for all i can be obtained.

Appendix D. The derivations of c_r^{+r}, v_r^{+r}, $\check{v}_{r,-r}$, $\hat{v}_{r,+r}$ and $\hat{v}_{r,-r}$

The mean conditional cycle time of station r if there is a service at station r, c_r^{+r} can be expressed as a weighted sum of the mean conditional pseudocycle time of stage i if there is a service at stage i, denoted by $\tilde{c}_i^{i+}$. c_r^{+r} is given by

$$c_r^{+r} \cong \frac{1}{\lambda_r} \sum_{\{i \mid R_i = r\}} \beta_i \cdot \tilde{c}_i^{i+}, \tag{D.1}$$

where $\tilde{c}_i^{i+}$ is obtained by

$$\tilde{c}_i^{i+} \cong \sum_{k=i}^{a_i-1} u_k + h_{R_i} + \sum_{\eta \neq R_i} \min\left(c_0^{i+} \sum_{k=i+1}^{a_i-1} I_{R_k}^{\eta} \beta_k, \sum_{k=i+1}^{a_i-1} I_{R_k}^{\eta} \right) h_\eta. \tag{D.2}$$

The function $I_{R_k}^{\eta}$ in (D.2) is equal to 1 if $\eta = R_k$ and is equal to 0 otherwise; c_0^{i+} denotes the mean conditional whole cycle time if there is a service at stage i and is given by

$$c_0^{i+} \cong u + h_{R_i} + \min[c_0^{i+}(\lambda_{R_i} - \beta_i), m_{R_i} - 1] h_{R_i} + \sum_{\eta \neq R_i} \min(c_0^{i+} \lambda_\eta, m_\eta) h_\eta. \tag{D.3}$$

Consequently, the mean conditional vacation time of station r if there is a service at station r before this vacation, v_r^{+r} can be obtained by $v_r^{+r} = c_r^{+r} - h_r$.

The mean conditional residual vacation time of station r if the server is on vacation from station r when the labeled customer arrives, $\check{v}_{r,-r}$ is given by [2, (15)]

$$\check{v}_{r,-r} \cong \sum_{\{i \mid R_i \neq r\}} \frac{\varphi_i}{1 - \rho_r} \cdot \check{v}_{r,i+} + \sum_{i=1}^{P} \frac{u_i / c_0}{1 - \rho_r} \cdot \check{v}_{r,i-}, \tag{D.4}$$

where φ_i is the load of stage i and $\varphi_i = \beta_i h_{R_i}$, and $\check{v}_{r,i+}$ ($\check{v}_{r,i-}$) denotes the mean conditional residual vacation time of station r if the server is in service at (in switchover time from) stage i when the labeled customer arrives. $\check{v}_{r,i+}$ for stage i, $R_i \neq r$, and $\check{v}_{r,i-}$ for stage i are obtained by

$$\begin{aligned}
\check{v}_{r,i+} &\cong \frac{h_{R_i}^{(2)}}{2h_{R_i}} + \sum_{k=i}^{r^i-1} u_k + \sum_{\eta \neq r} \min\left(c_{0,i+}^{r^i+} \sum_{k=i+1}^{r^i-1} I_{R_k}^{\eta} \beta_k, \sum_{k=i+1}^{r^i-1} I_{R_k}^{\eta} \right) h_\eta, \quad R_i \neq r, \\
\check{v}_{r,i-} &\cong \frac{u_i^{(2)}}{2u_i} + \sum_{k=i+1}^{r^i-1} u_k + \sum_{\eta \neq r} \min\left(c_{0,i-}^{r^i+} \sum_{k=i+1}^{r^i-1} I_{R_k}^{\eta} \beta_k, \sum_{k=i+1}^{r^i-1} I_{R_k}^{\eta} \right) h_\eta,
\end{aligned} \tag{D.5}$$

where $c_{0,i+}^{r^i+}$ ($c_{0,i-}^{r^i+}$) is the mean conditional whole cycle time if the server is in service at (in switchover time from) stage i when the labeled customer arrives and there is a service at stage r^i; r^i is a notation for the first stage after stage i that corresponds to station "r", $1 \leq i, r^i \leq P$. $c_{0,i+}^{r^i+}$ for stage i, $R_i \neq r$ and $c_{0,i-}^{r^i+}$ are given by

$$c_{0,i+}^{r^i+} \cong u + \frac{h_{R_i}^{(2)}}{h_{R_i}} + h_r + \min(c_{0,i+}^{r^i+}(\lambda_{R_i} - \beta_i), m_{R_i} - 1)h_{R_i} + \min(c_{0,i+}^{r^i+}(\lambda_r - \beta_{r^i}), m_r - 1)h_r$$
$$+ \sum_{\eta \neq R_i, r} \min(c_{0,i+}^{r^i+}\lambda_\eta, m_\eta)h_\eta, \quad R_i \neq r, \tag{D.6}$$
$$c_{0,i-}^{r^i+} \cong u - u_i + \frac{u_i^{(2)}}{u_i} + h_r + \min(c_{0,i-}^{r^i+}(\lambda_r - \beta_{r^i}), m_r - 1)h_r + \sum_{\eta \neq r} \min(c_{0,i-}^{r^i+}\lambda_\eta, m_\eta)h_\eta.$$

The mean conditional vacation time of station r after the HOL customer in station r if the server is in service at station r when the labeled customer arrives, $\hat{v}_{r,+r}$ is obtained by

$$\hat{v}_{r,+r} = \frac{1}{\lambda_r} \sum_{\{i \mid R_i = r\}} \beta_i \hat{v}_{r,i+}, \tag{D.7}$$

where $\hat{v}_{r,i+}$ is the mean conditional vacation time of station r after the HOL customer if the server is in service at stage i when the labeled customer arrives. $\hat{v}_{r,i+}$ for stage i, $R_i = r$, can be obtained by

$$\hat{v}_{r,i+} \cong \sum_{k=i}^{a_i - 1} u_k + \sum_{\eta \neq R_i} \min\left(c_{0,i+}^{a_i+} \sum_{k=i+1}^{a_i-1} I_{R_k}^{\eta} \beta_k, \sum_{k=i+1}^{a_i-1} I_{R_k}^{\eta}\right) h_\eta, \quad R_i = r, \tag{D.8}$$

where $c_{0,i+}^{a_i+}$ is the mean conditional whole cycle time if the server is in service at stage i and $R_i = r$ when the labeled customer arrives and there is also a service at stage a_i. $c_{0,i+}^{a_i+}$ for $R_i = r$ is given by

$$c_{0,i+}^{a_i+} \cong \begin{cases} u + \dfrac{h_{R_i}^{(2)}}{h_{R_i}} + \sum_{\eta \neq R_i} \min(c_{0,i+}^{a_i+}\lambda_\eta, m_\eta)h_\eta & m_{R_i} = 1, \\ u + \dfrac{h_{R_i}^{(2)}}{h_{R_i}} + h_{R_i} + \min[c_{0,i+}^{a_i+}(\lambda_{R_i} - \beta_i - \beta_{a_i}), m_{R_i} - 2]h_{R_i} & \\ \quad + \sum_{\eta \neq R_i} \min(c_{0,i+}^{a_i+}\lambda_\eta, m_\eta)h_\eta & m_{R_i} > 1. \end{cases} \tag{D.9}$$

The mean conditional vacation time of station r after the HOL customer if the server is on vacation from station r when the labeled customer arrives, $\hat{v}_{r,-r}$ is given by [2, (17)]

$$\hat{v}_{r,-r} \cong \sum_{\{i \mid R_i \neq r\}} \frac{\varphi_i}{1 - \rho_r} \cdot \hat{v}_{r,i+} + \sum_{i=1}^{P} \frac{u_i / c_0}{1 - \rho_r} \cdot \hat{v}_{r,i-}, \tag{D.10}$$

where $\hat{v}_{r,i+}$ ($\hat{v}_{r,i-}$) denotes the mean conditional vacation time of station r after the HOL customer if the server is in service at (in switchover time from) stage i when the labeled customer arrives. $\hat{v}_{r,i+}$ for stage i, $R_i \neq r$ and $\hat{v}_{r,i-}$ for i are obtained by

$$
\begin{aligned}
\hat{v}_{r,i+} &\cong \sum_{k=r^i}^{a_{r^i}-1} u_k + \sum_{\eta \neq r} \min\left(c_{0,i+}^{r^i+}, \sum_{k=r^i+1}^{a_{r^i}-1} I_{R_k}^{\eta} \beta_k, \sum_{k=r^i+1}^{a_{r^i}-1} I_{R_k}^{\eta} \right) h_\eta, \quad R_i \neq r, \\
\hat{v}_{r,i-} &\cong \sum_{k=r^i}^{a_{r^i}-1} u_k + \sum_{\eta \neq r} \min\left(c_{0,i+}^{r^i+}, \sum_{k=r^i+1}^{a_{r^i}-1} I_{R_k}^{\eta} \beta_k, \sum_{k=r^i+1}^{a_{r^i}-1} I_{R_k}^{\eta} \right) h_\eta,
\end{aligned}
\tag{D.11}
$$

where $c_{0,i+}^{r^i+}$ $(c_{0,i-}^{r^i+})$ is the mean conditional whole cycle time if the server is in service at (in switchover time from) stage i when the labeled customer arrives and there is a service at stage r^i and is given in (D.6). Note that $\hat{v}_{r,-r}$ should be greater than v_r^{+r}.

References

[1] H. Takagi, Queueing analysis of polling models: an update, in: *Stochastic Analysis of Computer and Communication Systems*, Elsevier, North Holland (1990) 267–318.

[2] M.M. Srinivasan, An approximation for mean waiting time in cyclic service systems with nonexhaustive service, *Performance Eval.* **9**(1) (1988) 17–33.

[3] S.W. Fuhrmann and Y.T. Wang, Analysis of cyclic service systems with limited service: bounds and approximations, *Performance Eval.* **9**(1) (1988) 35–54.

[4] M. Eisenberg, Queues with periodic service and changeover time, *Oper. Res.* **20** (1972) 440–451.

[5] O.J. Boxma, W.P. Groenendijk and J.A. Weststrate, A pseudoconservation law for service systems with a polling table, *IEEE Trans. Comm.* **38**(10) (1990) 1865–1870.

[6] R.W. Wolff, *Stochastic Modeling and the Theory of Queues*, Prentice-Hall, Englewood Cliffs, NJ (1989).

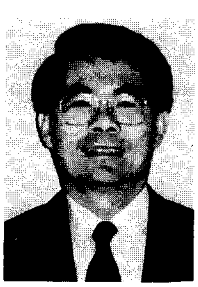
[7] P.H. Welch, On a generalized M/G/1 queueing process in which the first customer in a busy period receives exceptional service, *Oper. Res.* **12** (1964) 736–752.

[8] O.J. Boxma and B.W. Meister, Waiting-time approximations for cyclic-service system with switchover times, *Performance Eval.* **7**(4) (1987) 299–308.

[9] P.J. Kuehn, Multiqueue systems with nonexhaustive cyclic service, *Bell Syst. Tech. J.* **58**(3) (1979) 671–698.

[10] R.V. Hogg and A.T. Craig, *Introduction to Mathematical Statistics*, 4th ed., Macmillan, New York (1989).

[11] O.J. Boxma and W.P. Groenendijk, Pseudo-conservation law in cyclic-service systems, *J. Appl. Probab.* **24**(4) (1987) 949–964.



Chung-Ju Chang was born in Taiwan in 1950. He received the B.E. and M.E. degrees in Electronics Engineering from National Chiao Tung University, Hsinchu, Taiwan, in 1972 and 1976, respectively, and Ph.D. degree in Electrical Engineering from National Taiwan University in 1985.

From 1976 to 1988 he was with Telecommunication Laboratories, Directorate General of Telecommunications, Ministry of Communications, Republic of China, as a Design Engineer, Supervisor, Project Manager and then Division Director. There, he was involved in designing digital switching system, ISDN user-network interface and ISDN service and technology trials. In the meantime, he also acted as a Science and Technical Advisor for Minister of the Ministry of Communications from 1987 to 1989. In August 1988, he joined the faculty of the Department of Communication Engineering and Center for Telecommunications Research, National Chiao Tung University, where he is currently Professor and Director of the Institute of Communication Engineering. His research interests include performance evaluation, ATM (asynchronous transfer mode) networks and mobile radio networks.

Dr. Chang is a member of the Chinese Institute of Engineers (CIE) and IEEE.

Lain-Chyr Hwang was born in Taiwan on 26 December 1965. He received the B.S. degree in Electrical Engineering from National Sun Yat-Sen University, Kaohsiung, Taiwan, in 1988, the M.S. degree in Communication Engineering and the Ph.D. degree in Electronic Engineering from National Chiao-Tung University, Hsinchu, Taiwan, in 1990 and 1994, respectively. His research interests include queueing theory and computer communication networks.