Journal of the Operations Research
Society of Japan
Vol. 32, No. 4, December 1989

ANALYSIS OF POLLING SYSTEMS WITH A MIXTURE OF EXHAUSTIVE AND GATED SERVICE DISCIPLINES

Hideaki Takagi
IBM Japan, Ltd.

(Received October 11, 1988; Revised February 2, 1989)

Abstract Queueing analysis is performed for a certain type of multiple-station system attended by a single server in cyclic order. The service discipline at each station may be 'exhaustive' or 'gated', and stations with different disciplines can coexist in any order in the system. The mean and second moment of the waiting time at each station are obtained by solving a set of $O(N^3)$ and $O(N^4)$ linear equations, respectively, where N is the number of stations in the system. We consider the FCFS (first-come, first-served) and LCFS (last-come, first-served) order of service at each station. By numerical calculation of the mean waiting times (which do not depend on whether the order of service at each station is FCFS or LCFS), it is shown that gated-service stations close downstream from an exhaustive-service station receive favorable treatment, and that exhaustive-service stations close downstream from a gated-service station receive unfavorable treatment.

1. Introduction

A new development in the queueing analysis of polling models (i.e., systems of multiple stations served by a single server in cyclic order) is the study of stations with mixed service disciplines. Typical disciplines are *exhaustive* (the server continues to serve all messages at a station until it empties), *gated* (the server continuously serves only those messages that are found at a station when he inspects it; arrivals during the server's sojourn time are set aside for service in the next cycle), and *limited* (at most one message is served at a station in a cycle). Traditionally, most analyses have been carried out by assuming that the service discipline is the same for all stations in the system (see [13], [15], and references therein).

Recently, Boxma and Groenendijk [1, 2] have derived the *pseudo-conservation laws* with respect to the intensity-weighted sum of the mean waiting times for systems with mixed service disciplines. Ozawa [9] has obtained the mean waiting times for systems of two stations, station 1 being with exhaustive service and station 2 with gated service. Ozawa's method is to construct a one-dimensional Markov chain for the number of messages at station 2 when the server leaves station 1 (at this moment, the number of messages at station 1 is always zero). Thus, it seems difficult to extend his approach to systems with more than two stations.

Ozawa [9] has also noted that a classical model of head-of-line priority queues (Section 3.6 of [8]) with two classes is equivalent to the polling model with a mix of exhaustive and limited service disciplines (without switchover time). Polling systems involving two stations with a mix of exhaustive and limited disciplines in the case of nonzero switchover times are analyzed by Groenendijk [7], Ozawa [10], Skinner [11], and Srinivasan [12]. Skinner and Srinivasan assumed that the switchover time from the exhaustive-service station to the limited-service station is a constant. Groenendijk and Ozawa found the mean waiting times in the case where both switchover times are variable.

In this paper, we show that an analysis of systems with $N(\geq 2)$ stations with a mix of exhaustive and gated disciplines is possible by extending previously known techniques for systems with unmixed disciplines. In Section 2, we consider continuous-time systems without switchover times. Not only the mean waiting times but also the second moments of the waiting times are discussed. Note that the second moment of the waiting time depends on whether the FCFS (first-come, first-served) or the LCFS (last-come, first-served) service order is employed at each station. We then analyze systems with switchover times in the continuous-time model in Section 3, and those in the discrete-time model in Section 4. In particular, the mean waiting times are explicitly given for systems with two stations for these cases. The effects of mixed disciplines are discussed in Section 5, with numerical examples. We will present relatively detailed equations for the continuous-time systems without switchover times in Section 2; for other systems, we restrict ourselves to outlining the analysis and giving the mean waiting times only, because similar treatments are straightforward extensions.

We assume that N stations are indexed as $1, 2, \cdots, N$ in the order of polling. For convenience' sake, we denote by E and G the sets of indexes of the stations with exhaustive and gated disciplines, respectively.

2. Continuous-Time Systems without Switchover Times

The parameters of our system are as follows. Station i has a Poisson arrival stream of messages at rate λ_i, where $i = 1, 2, \cdots, N$. Let $\lambda = \sum_{i=1}^{N} \lambda_i$ be the total arrival rate. The Laplace-Stieltjes transform (LST) of the distribution function (DF) for the message service time at station i is denoted by $B_i^*(s)$. The mean and the nth moment of the service time at station i are denoted by b_i and $b_i^{(n)}, n = 2, 3, \cdots$, respectively. We assume an infinite capacity for each station. The server utilization of station i is given by the offered load $\rho_i = \lambda_i b_i$, and the total server utilization is given by $\rho = \sum_{i=1}^{N} \rho_i$. Our main objective is to find $E[W_i]$, the mean message waiting time at station i. This is obtained by solving a set of $O(N^3)$ linear equations. Since we have a work-conserving nonpreemptive-service system, the M/G/1 conservation law (Section 3.4 of [8]) must hold for the mean waiting times:

$$\overline{W} \equiv \sum_{i=1}^{N} \frac{\rho_i}{\rho} E[W_i] = \frac{\sum_{i=1}^{N} \lambda_i b_i^{(2)}}{2(1-\rho)} \tag{2.1}$$

In addition we show a set of $O(N^4)$ equations, the solution to which yields the second moment $E[W_i^2]$ of the waiting time at station i. Note that $E[W_i^2]$ depends on whether FCFS or LCFS is adopted at station i, while $E[W_i]$ does not.

2.1 Queue Length at Switch Points

Let us start with the analysis of the number of messages by following the approach of Cooper and Murray [5], Cooper [3], and Cooper [4 (problem 5.31)] (see also Chapter 7 of [13]). We define a *switch point* as a point in time at which the server leaves some station. Let $P_i(q_1, \cdots, q_N)$ be the joint probability that, at an arbitrary switch point, the server has just completed a visit to station i, and that q_j messages are present at station j, where $j = 1, \cdots, N$. Note that q_i is always zero for $i \in E$, by definition. Define

$$\begin{aligned} &H_i(z_1, \cdots, z_{i-1}, 0, z_{i+1}, \cdots, z_N) \\ &\equiv \sum_{q_1=0}^{\infty} \cdots \sum_{q_{i-1}=0}^{\infty} \sum_{q_{i+1}=0}^{\infty} \cdots \sum_{q_N=0}^{\infty} P_i(q_1, \cdots, q_{i-1}, 0, q_{i+1}, \cdots, q_N) \Pi_{\substack{j=1 \\ (j \neq i)}}^{N} (z_j)^{q_j} \quad i \in E \end{aligned} \tag{2.2}$$

$$H_i(z_1, \cdots, z_N) \equiv \sum_{q_1=0}^{\infty} \cdots \sum_{q_N=0}^{\infty} P_i(q_1, \cdots, q_N) \Pi_{j=1}^{N} (z_j)^{q_j} \quad i \in G \tag{2.3}$$

In [5, 13], the relations between $H_i(.)$ and $H_{i-1}(.)$ are derived when station i and station $i-1$ have the same service discipline. They are given by

$$H_i(z_1,\cdots,z_{i-1},0,z_{i+1},\cdots,z_N) = H_{i-1}(z_1,\cdots,z_{i-2},0,\Theta_i^*\left[\sum_{\substack{j=1\\(j\neq i)}}^{N}(\lambda_j-\lambda_j z_j)\right],z_{i+1},\cdots,z_N)$$

$$-P_{i-1}(0,\cdots,0)+\frac{\lambda_i}{\lambda}P(0)\Theta_i^*\left[\sum_{\substack{j=1\\(j\neq i)}}^{N}(\lambda_j-\lambda_j z_j)\right] \quad i\in E,\ i-1\in E \tag{2.4}$$

$$H_i(z_1,\cdots,z_N) = H_{i-1}(z_1,\cdots,z_{i-1},B_i^*\left[\sum_{j=1}^{N}(\lambda_j-\lambda_j z_j)\right],z_{i+1},\cdots,z_N)$$

$$-P_{i-1}(0,\cdots,0)+\frac{\lambda_i}{\lambda}P(0)B_i^*\left[\sum_{j=1}^{N}(\lambda_j-\lambda_j z_j)\right] \quad i\in G,\ i-1\in G \tag{2.5}$$

where

$$P(0)\equiv\sum_{i=1}^{N}P_i(0,\cdots,0) \tag{2.6}$$

is the probability that the whole system becomes empty at a switch point. The LST of the DF for the length Θ_i of a busy period generated by a single message at station i is denoted by $\Theta_i^*(s)$, which satisfies the equation

$$\Theta_i^*(s)=B_i^*[s+\lambda_i-\lambda_i\Theta_i^*(s)] \tag{2.7}$$

The mean and the nth moment of Θ_i are denoted by θ_i and $\theta_i^{(n)}, n=2,3,\cdots$, respectively. From (2.7), they can be expressed in terms of b_i and $b_i^{(n)}$'s.

Similarly, we can get the following relationships when the service disciplines at station i and station $i-1$ are different:

$$H_i(z_1,\cdots,z_N) = H_{i-1}(z_1,\cdots,z_{i-2},0,B_i^*\left[\sum_{j=1}^{N}(\lambda_j-\lambda_j z_j)\right],z_{i+1},\cdots,z_N)$$

$$-P_{i-1}(0,\cdots,0)+\frac{\lambda_i}{\lambda}P(0)B_i^*\left[\sum_{j=1}^{N}(\lambda_j-\lambda_j z_j)\right] \quad i\in G,\ i-1\in E \tag{2.8}$$

$$H_i(z_1,\cdots,z_{i-1},0,z_{i+1},\cdots,z_N) = H_{i-1}(z_1,\cdots,z_{i-1},\Theta_i^*\left[\sum_{\substack{j=1\\(j\neq i)}}^{N}(\lambda_j-\lambda_j z_j)\right],z_{i+1},\cdots,z_N)$$

$$-P_{i-1}(0,\cdots,0)+\frac{\lambda_i}{\lambda}P(0)\Theta_i^*\left[\sum_{\substack{j=1\\(j\neq i)}}^{N}(\lambda_j-\lambda_j z_j)\right] \quad i\in E,\ i-1\in G \tag{2.9}$$

Let us introduce

$$h_i(j)\equiv\frac{\lambda}{P(0)}\frac{\partial H_i}{\partial z_j}\Big|_{z_1=\cdots=z_N=1};\quad h_i(j,k)\equiv\frac{\lambda(1-\rho)}{P(0)}\frac{\partial^2 H_i}{\partial z_j\partial z_k}\Big|_{z_1=\cdots=z_N=1}$$

$$h_i(j,k,l)\equiv\frac{\lambda(1-\rho)}{P(0)}\frac{\partial^3 H_i}{\partial z_j\partial z_k\partial z_l}\Big|_{z_1=\cdots=z_N=1} \tag{2.10}$$

By differentiating (2.4-5) and (2.8-9), we have a set of equations for $\{h_i(j); i,j = 1,\cdots,N\}$, $\{h_i(j,k); i,j,k = 1,\cdots,N\}$, and $\{h_i(j,k,l); i,j,k,l = 1,\cdots,N\}$. For $i \in E$, we get

$$h_i(j) = h_{i-1}(j) + h_{i-1}(i)\lambda_j\theta_i + \lambda_i\lambda_j\theta_i \tag{2.11}$$

$$\begin{aligned} h_i(j,k) &= h_{i-1}(j,k) + [h_{i-1}(j,i)\lambda_k + h_{i-1}(i,k)\lambda_j]\theta_i + h_{i-1}(i,i)\lambda_j\lambda_k\theta_i^2 \\ &\quad + (1-\rho)[h_{i-1}(i) + \lambda_i]\lambda_j\lambda_k\theta_i^{(2)} \end{aligned} \tag{2.12}$$

$$\begin{aligned} h_i(j,k,l) &= h_{i-1}(j,k,l) + [h_{i-1}(j,k,i)\lambda_l + h_{i-1}(i,k,l)\lambda_j + h_{i-1}(j,i,l)\lambda_k]\theta_i \\ &\quad + [h_{i-1}(j,i,i)\lambda_k\lambda_l + h_{i-1}(i,k,i)\lambda_j\lambda_l + h_{i-1}(i,i,l)\lambda_j\lambda_k]\theta_i^2 + h_{i-1}(i,i,i)\lambda_j\lambda_k\lambda_l\theta_i^3 \\ &\quad + 3h_{i-1}(i,i)\lambda_j\lambda_k\lambda_l\theta_i\theta_i^{(2)} + [h_{i-1}(i,j)\lambda_k\lambda_l + h_{i-1}(i,k)\lambda_j\lambda_l + h_{i-1}(i,l)\lambda_j\lambda_k]\theta_i^{(2)} \\ &\quad + (1-\rho)[h_{i-1}(i) + \lambda_i]\lambda_j\lambda_k\lambda_l\theta_i^{(3)} \end{aligned} \tag{2.13}$$

For $i \in G$, we get

$$\begin{aligned} h_i(j) &= h_{i-1}(j) + h_{i-1}(i)\lambda_j b_i + \lambda_i\lambda_j b_i \quad j \neq i \\ h_i(i) &= h_{i-1}(i)\lambda_i b_i + \lambda_i^2 b_i \end{aligned} \tag{2.14}$$

$$\begin{aligned} h_i(j,k) &= h_{i-1}(j,k) + [h_{i-1}(j,i)\lambda_k + h_{i-1}(i,k)\lambda_j]b_i + h_{i-1}(i,i)\lambda_j\lambda_k b_i^2 \\ &\quad + (1-\rho)[h_{i-1}(i) + \lambda_i]\lambda_j\lambda_k b_i^{(2)} \quad j \neq i; k \neq i \\ h_i(i,k) &= h_{i-1}(i,k)\lambda_i b_i + h_{i-1}(i,i)\lambda_i\lambda_k b_i^2 + (1-\rho)[h_{i-1}(i) + \lambda_i]\lambda_i\lambda_k b_i^{(2)} \quad k \neq i \\ h_i(i,i) &= h_{i-1}(i,i)(\lambda_i b_i)^2 + (1-\rho)[h_{i-1}(i) + \lambda_i]\lambda_i^2 b_i^{(2)} \end{aligned} \tag{2.15}$$

$$\begin{aligned} h_i(j,k,l) &= h_{i-1}(j,k,l) + [h_{i-1}(j,k,i)\lambda_l + h_{i-1}(i,k,l)\lambda_j + h_{i-1}(j,i,l)\lambda_k]b_i \\ &\quad + [h_{i-1}(j,i,i)\lambda_k\lambda_l + h_{i-1}(i,k,i)\lambda_j\lambda_l + h_{i-1}(i,i,l)\lambda_j\lambda_k]b_i^2 + h_{i-1}(i,i,i)\lambda_j\lambda_k\lambda_l b_i^3 \\ &\quad + 3h_{i-1}(i,i)\lambda_j\lambda_k\lambda_l b_i b_i^{(2)} + [h_{i-1}(i,j)\lambda_k\lambda_l + h_{i-1}(i,k)\lambda_j\lambda_l + h_{i-1}(i,l)\lambda_j\lambda_k]b_i^{(2)} \\ &\quad + (1-\rho)[h_{i-1}(i) + \lambda_i]\lambda_j\lambda_k\lambda_l b_i^{(3)} \quad j \neq i;\ k \neq i;\ l \neq i \\ h_i(i,k,l) &= h_{i-1}(i,k,l)\lambda_i b_i + [h_{i-1}(i,k,i)\lambda_l + h_{i-1}(i,i,l)\lambda_k]\lambda_i b_i^2 + h_{i-1}(i,i,i)\lambda_i\lambda_k\lambda_l b_i^3 \\ &\quad + 3h_{i-1}(i,i)\lambda_i\lambda_k\lambda_l b_i b_i^{(2)} + [h_{i-1}(i,k)\lambda_l + h_{i-1}(i,l)\lambda_k]\lambda_i b_i^{(2)} \\ &\quad + (1-\rho)[h_{i-1}(i) + \lambda_i]\lambda_i\lambda_k\lambda_l b_i^{(3)} \quad k \neq i;\ l \neq i \\ h_i(i,i,l) &= h_{i-1}(i,i,l)(\lambda_i b_i)^2 + h_{i-1}(i,i,i)\lambda_i^2\lambda_l b_i^3 + 3h_{i-1}(i,i)\lambda_i^2\lambda_l b_i b_i^{(2)} + h_{i-1}(i,l)\lambda_i^2 b_i^{(2)} \\ &\quad + (1-\rho)[h_{i-1}(i) + \lambda_i]\lambda_i^2\lambda_l b_i^{(3)} \quad l \neq i \\ h_i(i,i,i) &= h_{i-1}(i,i,i)(\lambda_i b_i)^3 + 3h_{i-1}(i,i)\lambda_i^3 b_i b_i^{(2)} + (1-\rho)[h_{i-1}(i) + \lambda_i]\lambda_i^3 b_i^{(3)} \end{aligned} \tag{2.16}$$

In the above equations, $h_i(j) = 0, h_i(j,k) = 0$, and $h_i(j,k,l) = 0$ if $i \in E$ and $j = i, k = i$, or $l = i$. The number of equations in (2.11) and (2.14) is $O(N^2)$, that in (2.12) and (2.15) is $O(N^3)$, and that in (2.13) and (2.16) is $O(N^4)$. We note that, if the service discipline is the same for all stations, we can get $\{h_i(j)\}$ explicitly, and have $O(N^2)$ equations for $\{h_i(j,k)\}$ and $O(N^3)$ equations for $\{h_i(j,k,l)\}$ [3, 13, 14]. However, we have not succeeded in the similar reduction in the number of equations for systems with mixed disciplines.

2.2 Waiting Times

We now express the LST $W_i^*(s)$ of the DF for the message waiting time at station i in terms of $H_{i-1}(.)$, and the first two moments, $E[W_i]$ and $E[W_i^2]$, in terms of $\{h_i(j)\}$, $\{h_i(j,k)\}$ and $\{h_i(j,k,l)\}$ obtained above.

The LST of the DF for the message waiting time at station i with FCFS is given by [3, 4, 13]

$$W_i^*(s) = \frac{s(1-\rho)}{s-\lambda_i+\lambda_i B_i^*(s)} + \frac{\lambda_i(\rho-\rho_i)[H_{i-1}(1,\cdots,1,\zeta,1,\cdots,1) - H_{i-1}(1,\cdots,1,\zeta,1-s/\lambda_i,\cdots,1)]}{[P(0)/\lambda][s-\lambda_i+\lambda_i B_i^*(s)]h_{i-1}(i)} \quad i \in E \tag{2.17}$$

$$W_i^*(s) = 1-\rho + \frac{\lambda_i\rho[H_{i-1}(1,\cdots,1,\zeta,B_i^*(s),1,\cdots,1) - H_{i-1}(1,\cdots,1,\zeta,1-s/\lambda_i,1,\cdots,1)]}{[P(0)/\lambda][s-\lambda_i+\lambda_i B_i^*(s)]h_{i-1}(i)} \quad i \in G \tag{2.18}$$

where ζ, the $(i-1)$th argument of $H_{i-1}(.)$, is defined by

$$\zeta \equiv \begin{cases} 0 & i-1 \in E \\ 1 & i-1 \in G \end{cases} \tag{2.19}$$

We note that an incorrect expression for $W_i^*(s)$ for $i \in G$ in [3] was corrected in [4 (problem 5.31)] as in (2.18). Once we have obtained $\{h_i(j); i,j = 1,\cdots,N\}, \{h_i(j,k); i,j,k = 1,\cdots,N\}$, and $\{h_i(j,k,l); i,j,k,l = 1,\cdots,N\}$, we can calculate the mean and the second moment of the waiting times. For $i \in E$, we get

$$E[W_i] = \frac{\lambda_i b_i^{(2)}}{2(1-\rho_i)} + \frac{(\rho-\rho_i)h_{i-1}(i,i)}{2\lambda_i(1-\rho)(1-\rho_i)h_{i-1}(i)} \tag{2.20}$$

$$E[W_i^2] = \frac{\lambda_i b_i^{(3)}}{3(1-\rho_i)} + \frac{[\lambda_i b_i^{(2)}]^2}{2(1-\rho_i)^2} + \frac{\rho-\rho_i}{(1-\rho)h_{i-1}(i)}\left[\frac{h_{i-1}(i,i,i)}{3\lambda_i^2(1-\rho_i)} + \frac{h_{i-1}(i,i)b_i^{(2)}}{2(1-\rho_i)^2}\right] \tag{2.21}$$

For $i \in G$, we get

$$E[W_i] = \frac{\rho(1+\rho_i)h_{i-1}(i,i)}{2\lambda_i(1-\rho)h_{i-1}(i)} \tag{2.22}$$

$$E[W_i^2] = \frac{\rho(1+\rho_i+\rho_i^2)h_{i-1}(i,i,i)}{3\lambda_i^2(1-\rho)h_{i-1}(i)} + \frac{\rho h_{i-1}(i,i)b_i^{(2)}}{2(1-\rho)h_{i-1}(i)} \tag{2.23}$$

Note that equations (2.20) and (2.22) hold not only for FCFS but also for all other nonpreemptive service disciplines such that the order of service within each station does not depend on the service time. For $N = 2$ with $1 \in E$ and $2 \in G$, we recover a result of Ozawa [9]

$$E[W_1] = \frac{(1-\rho_1)(\lambda_1 b_1^{(2)} + \lambda_2 b_2^{(2)})}{2(1-\rho)(1-\rho_1+\rho_2)}; \quad E[W_2] = \frac{(1+\rho_2)(\lambda_1 b_1^{(2)} + \lambda_2 b_2^{(2)})}{2(1-\rho)(1-\rho_1+\rho_2)} \tag{2.24}$$

Clearly, $E[W_2] \geq E[W_1]$. We will see this trend more evidently in the numerical examples in Section 5.

In the case of LCFS order of service at station i, we can obtain $W_i^*(s)$ by extending the analysis for a single LCFS M/G/1 queue with vacations (see Appendix for an outline of derivation). Note that, at a station with LCFS gated service, only those messages that are present in the station at the time of polling are served continuously in the LCFS order. We have

$$W_i^*(s) = 1-\rho+\frac{\lambda_i[1-\Theta_i^*(s)]}{s+\lambda_i-\lambda_i\Theta_i^*(s)}$$
$$+\frac{\lambda_i(\rho-\rho_i)[H_{i-1}(1,\cdots,1,\zeta,1,1,\cdots,1)-H_{i-1}(1,\cdots,1,\zeta,\Theta_i^*(s)-s/\lambda_i,1,\cdots,1)]}{[P(0)/\lambda][s+\lambda_i-\lambda_i\Theta_i^*(s)]h_{i-1}(i)} \quad i\in E \tag{2.25}$$

$$W_i^*(s) = 1-\rho$$
$$+\frac{\lambda_i\rho[H_{i-1}(1,\cdots,1,\zeta,1,1,\cdots,1)-H_{i-1}(1,\cdots,1,\zeta,B_i^*(s)-s/\lambda_i,1,\cdots,1)]}{[P(0)/\lambda][s+\lambda_i-\lambda_i B_i^*(s)]h_{i-1}(i)} \quad i\in G \tag{2.26}$$

where ζ is defined in (2.19). From (2.25) and (2.26), we get (2.20) and (2.22) for the mean waiting time $E[W_i]$, and

$$E[W_i^2]=\frac{\lambda_i b_i^{(3)}}{3(1-\rho_i)^2}+\frac{[\lambda_i b_i^{(2)}]^2}{2(1-\rho_i)^3}+\frac{\rho-\rho_i}{(1-\rho)h_{i-1}(i)}\left[\frac{h_{i-1}(i,i,i)}{3\lambda_i^2(1-\rho_i)^2}+\frac{h_{i-1}(i,i)b_i^{(2)}}{2(1-\rho_i)^3}\right] \quad i\in E \tag{2.27}$$

$$E[W_i^2]=\frac{\rho(1+\rho_i)^2h_{i-1}(i,i,i)}{3\lambda_i^2(1-\rho)h_{i-1}(i)}+\frac{\rho h_{i-1}(i,i)b_i^{(2)}}{2(1-\rho)h_{i-1}(i)} \quad i\in G \tag{2.28}$$

Suppose that the sets of indexes E and G are given. Then, $h_{i-1}(i), h_{i-1}(i,i)$, and $h_{i-1}(i,i,i)$ are determined by solving (2.11-16), which are independent of whether the order of service at each station is FCFS or LCFS. In such a case, by comparing (2.21) with (2.27), and (2.23) with (2.28), we get

$$E[W_i^2]_{\mathrm{LCFS}}=\frac{1}{1-\rho_i}E[W_i^2]_{\mathrm{FCFS}} \quad i\in E \tag{2.29}$$

$$E[W_i^2]_{\mathrm{LCFS}}\geq E[W_i^2]_{\mathrm{FCFS}} \quad i\in G \tag{2.30}$$

which are generalizations of the same results for single queue systems.

3. Continuous-Time Systems with Switchover Times

We next consider a similar system with switchover times. The LST of the DF for the server to switch from station i to station $i+1$ is denoted by $R_i^*(s)$, with mean r_i and variance δ_i^2. For this system, the following pseudo-conservation law must hold [1]:

$$\overline{W}\equiv\sum_{i=1}^N\frac{\rho_i}{\rho}E[W_i]=\frac{\sum_{i=1}^N\lambda_i b_i^{(2)}}{2(1-\rho)}+\frac{\sum_{i=1}^N\delta_i^2}{2R}+\frac{R\left(\rho-\sum_{i=1}^N\rho_i^2+2\sum_{i\in G}\rho_i^2\right)}{2\rho(1-\rho)} \tag{3.1}$$

where $R=\sum_{i=1}^N r_i$ is the mean total switchover time. In contrast to our approach in Section 2, we now consider the points in time at which the server polls a station (Chapters 4 and 5 of [13]). We show the expressions for the mean waiting times in terms of the first and second moments of the distribution of the numbers of messages at these points. Note that the mean waiting times are independent of whether FCFS or LCFS is adopted at each station.

Let $L_j(t)$ be the number of messages at station j at time t, and let τ_i be the time when station i is polled. We define the joint generating function (GF) for the number of messages at the polling instant of station i by

$$F_i(z_1,\cdots,z_N) \equiv E\left[\Pi_{j=1}^N (z_j)^{L_j(\tau_i)}\right] \tag{3.2}$$

Depending on the service discipline at station i, we have the following relations:

$$F_{i+1}(z_1,\cdots,z_N) = R_i^*\left[\sum_{j=1}^N(\lambda_j - \lambda_j z_j)\right] F_i(z_1,\cdots,z_{i-1},\Theta_i^*\left[\sum_{\substack{j=1\\(j\neq i)}}^N (\lambda_j - \lambda_j z_j)\right], z_{i+1},\cdots,z_N) \quad i \in E \tag{3.3}$$

$$F_{i+1}(z_1,\cdots,z_N) = R_i^*\left[\sum_{j=1}^N(\lambda_j - \lambda_j z_j)\right] F_i(z_1,\cdots,z_{i-1},B_i^*\left[\sum_{j=1}^N (\lambda_j - \lambda_j z_j)\right], z_{i+1},\cdots,z_N) \quad i \in G \tag{3.4}$$

where $\Theta_i^*(s)$ is a solution to (2.7). From these equations, we can get a set of linear equations for $\{f_i(j); i,j = 1,\cdots,N\}$ and $\{f_i(j,k); i,j,k = 1,\cdots,N\}$, where

$$f_i(j) \equiv \frac{\partial F_i}{\partial z_j}\Big|_{z_1=\cdots=z_N=1}; \quad f_i(j,k) \equiv \frac{\partial^2 F_i}{\partial z_j \partial z_k}\Big|_{z_1=\cdots=z_N=1} \tag{3.5}$$

Once they have been solved, the mean message waiting times are computed through the relationships between $F_i(.)$ and the LST $W_i^*(s)$ of the DF for the waiting time. For systems with FCFS within each station, they are given by [13]

$$W_i^*(s) = \frac{\lambda_i(1-\rho_i)[1 - F_i(1,\cdots,1,1-s/\lambda_i,1,\cdots,1)]}{[s - \lambda_i + \lambda_i B_i^*(s)]f_i(i)} \quad i \in E \tag{3.6}$$

$$W_i^*(s) = \frac{\lambda_i[F_i(1,\cdots,1,B_i^*(s),1,\cdots,1) - F_i(1,\cdots,1,1-s/\lambda_i,1,\cdots,1)]}{[s - \lambda_i + \lambda_i B_i^*(s)]f_i(i)} \quad i \in G \tag{3.7}$$

Thus we get

$$E[W_i] = \frac{f_i(i,i)}{2\lambda_i f_i(i)} + \frac{\lambda_i b_i^{(2)}}{2(1-\rho_i)} \quad i \in E \tag{3.8}$$

$$E[W_i] = \frac{(1+\rho_i) f_i(i,i)}{2\lambda_i f_i(i)} \quad i \in G \tag{3.9}$$

As noted above, (3.8) and (3.9) hold for both FCFS and LCFS service disciplines within each station. Higher moments of the waiting times, which depend on the service discipline at each station, can be obtained similarly.

Again, for $N = 2$ with $1 \in E$ and $2 \in G$, we recover a result of Ozawa [9]

$$E[W_1] = \frac{(1-\rho_1)(\lambda_1 b_1^{(2)} + \lambda_2 b_2^{(2)})}{2(1-\rho)(1-\rho_1+\rho_2)} + \frac{(1-\rho_1)R}{2(1-\rho)} + \frac{(1-\rho_1)(\delta_1^2+\delta_2^2)}{2(1-\rho_1+\rho_2)R} + \frac{\rho_2(1-\rho)(1+\rho_2)\delta_1^2}{(1-\rho_1+\rho_2)R} \tag{3.10}$$

$$E[W_2] = \frac{(1+\rho_2)(\lambda_1 b_1^{(2)} + \lambda_2 b_2^{(2)})}{2(1-\rho)(1-\rho_1+\rho_2)} + \frac{(1+\rho_2)R}{2(1-\rho)} + \frac{(1+\rho_2)(\delta_1^2+\delta_2^2)}{2(1-\rho_1+\rho_2)R} - \frac{\rho_1(1-\rho)(1+\rho_2)\delta_1^2}{(1-\rho_1+\rho_2)R} \tag{3.11}$$

which can easily be shown to satisfy (3.1).

4. Discrete-Time Systems with Switchover Times

The model of discrete-time systems is as follows (Chapter 3 of [13]). The entity of service is a packet of a fixed length, and the time is slotted, with the slot size equal to the packet service time. The number of packets that arrive at station i in a slot has the GF $A_i(z)$, mean μ_i, and variance σ_i^2. The switchover time from station i to station $i+1$ has the GF $R_i(z)$, mean r_i, and variance δ_i^2. The objective of our analysis is to find the mean packet waiting time $E[W_i]$ for a randomly chosen packet at station i. For this system, the pseudo-conservation law is given by [2]

$$\overline{W} \equiv \sum_{i=1}^{N} \frac{\mu_i}{\mu} E[W_i] = \frac{\sum_{i=1}^N \sigma_i^2}{2\mu(1-\mu)} + \frac{\sum_{i=1}^N \delta_i^2}{2R} + \frac{R\Big(\mu - \sum_{i=1}^N \mu_i^2 + 2\sum_{i\in G}\mu_i^2\Big)}{2\mu(1-\mu)} - 1 \tag{4.1}$$

where $\mu = \sum_{i=1}^N \mu_i$ and $R = \sum_{i=1}^N r_i$.

For this system we may conduct an analysis similar to the one in the previous section. Now, instead of (3.3) and (3.4), we have

$$F_{i+1}(z_1,\cdots,z_N) = R_i\Big[\Pi_{j=1}^N A_j(z_j)\Big] F_i(z_1,\cdots,z_{i-1},\Theta_i\Big[\Pi_{\substack{j=1\\(j\neq i)}}^N A_j(z_j)\Big], z_{j+1},\cdots,z_N) \quad i\in E \tag{4.2}$$

$$F_{i+1}(z_1,\cdots,z_N) = R_i\Big[\Pi_{j=1}^N A_j(z_j)\Big] F_i(z_1,\cdots,z_{i-1},\Pi_{j=1}^N A_j(z_j), z_{i+1},\cdots,z_N) \quad i\in G \tag{4.3}$$

where $\Theta_i(z)$ is the GF for the length (in slots) of a busy period in the discrete-time system, and satisfies the equation

$$\Theta_i(z) = zA_i[\Theta_i(z)] \tag{4.4}$$

As before, from (4.2) and (4.3) we have a set of equations for $\{f_i(j); i,j=1,\cdots,N\}$ and $\{f_i(j,k); i,j,k=1,\cdots,N\}$ which are defined by (3.5). Again, we have expressions for the LST $W_i^*(s)$ of the DF for the packet waiting time in terms of $F_i(.)$ [13]. We then have

$$E[W_i] = \frac{f_i(i,i)}{2\mu_i f_i(i)} + \frac{1}{2\mu_i} + \frac{\sigma_i^2}{2\mu_i}\Big(\frac{1}{1-\mu_i} - \frac{1}{\mu_i}\Big) - 1 \quad i\in E \tag{4.5}$$

$$E[W_i] = \frac{(1+\mu_i)f_i(i,i)}{2\mu_i f_i(i)} + \frac{1+\mu_i}{2\mu_i} - \frac{\sigma_i^2}{2\mu_i^2} - 1 \quad i\in G \tag{4.6}$$

for both FCFS and LCFS service disciplines within each station.

In the case of $N=2$ with $1\in E$ and $2\in G$, we obtain

$$E[W_1] = \frac{\sigma_1^2(1-\mu_1-\mu_2^2) + \sigma_2^2\mu_1(1-\mu_1)^2}{2\mu_1(1-\mu_1+\mu_2)(1-\mu)}$$

$$+\frac{(1-\mu_1)(\delta_1^2+\delta_2^2)+2\delta_1^2\mu_2(1+\mu_2)(1-\mu)}{2(1-\mu_1+\mu_2)R}+\frac{(1-\mu_1)R}{2(1-\mu)}-1 \tag{4.7}$$

$$E[W_2]=\frac{\sigma_1^2\mu_2(1+\mu_2)+\sigma_2^2[(1-\mu_1)^2+\mu_2]}{2\mu_2(1-\mu_1+\mu_2)(1-\mu)}$$
$$+\frac{(1+\mu_2)(\sigma_1^2+\sigma_2^2)-2\sigma_1^2\mu_1(1+\mu_2)(1-\mu)}{2(1-\mu_1+\mu_2)R}+\frac{(1+\mu_2)R}{2(1-\mu)}-1 \tag{4.8}$$

which satisfy (4.1). The results in (4.7) and (4.8) are new.

5. Numerical Examples

To see the effects of mixed service disciplines in a system, let us consider numerical examples for continuous-time systems without switchover times, which were analyzed in Section 2. For $N = 5$, let us assume that $b_i = b_i^{(2)} = 1$ for all i and that all λ_i's are identical. In the first example, station 1 has exhaustive service and stations 2 through 5 have gated service. In the second example, station 1 has gated service and stations 2 through 5 have exhaustive service. The mean message waiting times have been computed by the method in Section 2, and are shown in Table 1 together with the intensity-weighted mean in (2.1).

In the first example, the mean waiting time at station 1 (exhaustive service) is significantly less than those at other stations with gated service. Among gated-service stations, the mean waiting times are less for stations closer to station 1 in the polling cycle direction. We note that a similar phenomenon has been pointed out by Ferguson and Aminetzah [6] as the influence of a heavily loaded station on the lightly loaded stations downstream, all with exhaustive service. In the second example, we observe an inverse phenomenon.

Table 1. Mean message waiting times for continuous-time polling systems without switchover times with mixed service disciplines. $N = 5$, $b_i = b_i^{(2)} = 1$ for all i, and all λ_i's are identical.

ρ	station 1 exhaustive	station 2 gated	station 3 gated	station 4 gated	station 5 gated	$\overline{W}$
0.05	0.02589	0.02642	0.02642	0.02642	0.02642	0.02632
0.10	0.05378	0.05598	0.05599	0.05601	0.05602	0.05556
0.15	0.08401	0.08921	0.08926	0.08932	0.08938	0.08824
0.20	0.11704	0.12679	0.12691	0.12705	0.12721	0.12500
0.25	0.15343	0.16958	0.16982	0.17009	0.17041	0.16667
0.30	0.19393	0.21869	0.21910	0.21957	0.22014	0.21429
0.35	0.23951	0.27556	0.27621	0.27697	0.27790	0.26923
0.40	0.29146	0.34215	0.34310	0.34426	0.34569	0.33333
0.45	0.35156	0.42110	0.42245	0.42412	0.42623	0.40909
0.50	0.42229	0.51613	0.51798	0.52031	0.52330	0.50000
0.55	0.50725	0.63264	0.63509	0.63824	0.64234	0.61111
0.60	0.61184	0.77870	0.78189	0.78604	0.79153	0.75000
0.65	0.74455	0.96706	0.97112	0.97648	0.98366	0.92857
0.70	0.91953	1.21891	1.22403	1.23082	1.24004	1.16667
0.75	1.16229	1.57251	1.57886	1.58735	1.59900	1.50000
0.80	1.52381	2.10432	2.11213	2.12261	2.13712	2.00000
0.85	2.12314	2.99286	3.00240	3.01521	3.03305	2.83333
0.90	3.31738	4.77379	4.78537	4.80089	4.82257	4.50000
0.95	6.89222	10.12561	10.13959	10.15824	10.18434	9.50000

ρ	station 1 gated	station 2 exhaustive	station 3 exhaustive	station 4 exhaustive	station 5 exhaustive	$\overline{W}$
0.05	0.02674	0.02621	0.02621	0.02621	0.02621	0.02632
0.10	0.05733	0.05514	0.05512	0.05510	0.05509	0.05556
0.15	0.09247	0.08726	0.08721	0.08715	0.08709	0.08824
0.20	0.13301	0.12321	0.12308	0.12294	0.12277	0.12500
0.25	0.18002	0.16374	0.16349	0.16321	0.16287	0.16667
0.30	0.23491	0.20985	0.20943	0.20892	0.20832	0.21429
0.35	0.29951	0.26283	0.26215	0.26133	0.26032	0.26923
0.40	0.37628	0.32437	0.32337	0.32212	0.32053	0.33333
0.45	0.46857	0.39679	0.39537	0.39355	0.39119	0.40909
0.50	0.58108	0.48331	0.48136	0.47882	0.47543	0.50000
0.55	0.72065	0.58857	0.58598	0.58253	0.57782	0.61111
0.60	0.89753	0.71952	0.71615	0.71159	0.70521	0.75000
0.65	1.12789	0.88704	0.88274	0.87682	0.86837	0.92857
0.70	1.43868	1.10924	1.10381	1.09627	1.08532	1.16667
0.75	1.87853	1.41866	1.41190	1.40244	1.38848	1.50000
0.80	2.54472	1.88033	1.87198	1.86025	1.84273	2.00000
0.85	3.66441	2.64586	2.63562	2.62124	2.59954	2.83333
0.90	5.91930	4.16993	4.15741	4.13994	4.11341	4.50000
0.95	12.71835	8.72539	8.71014	8.68909	8.65703	9.50000

References

[1] Boxma, O.J. and Groenendijk, W.P.: Pseudo-Conservation Laws in Cyclic-Service Systems. *Journal of Applied Probability*, Vol. 24, No. 4 (1987), 949–964.

[2] Boxma, O.J. and Groenendijk, W.P.: Waiting Times in Discrete-Time Cyclic-Service Systems. *IEEE Transactions on Communications*, Vol. COM-36, No. 2 (1988), 223–231.

[3] Cooper, R.B.: Queues Served in Cyclic Order: Waiting Times. *The Bell System Technical Journal*, Vol. 49, No. 3 (1970), 399–413.

[4] Cooper, R.B.: *Introduction to Queueing Theory*. Second Edition, North-Holland, New York, 1981.

[5] Cooper, R.B. and Murray, G.: Queues Served in Cyclic Order. *The Bell System Technical Journal*, Vol. 48, No. 3 (1969), 675–689.

[6] Ferguson, M.J. and Aminetzah, Y.J.: Exact Results for Nonsymmetric Token Ring Systems. *IEEE Transactions on Communications*, Vol. COM-33, No. 3 (1985), 223–231.

[7] Groenendijk, W.P.: Waiting-Time Approximations for Cyclic-Service Systems with Mixed Service Strategies. *Twelfth International Teletraffic Congress*, 1.4B.3.1-1.4B.3.8, Torino, Italy, 1988.

[8] Kleinrock, L.: *Queueing Systems, Volume 2: Computer Applications*. John Wiley and Sons, New York, 1976.

[9] Ozawa, T.: An Analysis for Multi-Queueing Systems with Cyclic-Service Discipline—Models with Exhaustive and Gated Services. IEICE Technical report, Vol. 87, No. 4 (1987), 19–24, The Institute of Electronics, Information and Communication Engineers of Japan (in Japanese).

[10] Ozawa, T.: Analysis of a Single Server Model with Two Queues Having Different Service Disciplines. *The Transactions of the Institute of Electronics, Information and Communication Engineers*, Vol. J71-A, No. 5 (1988), 1193–1200 (in Japanese).

[11] Skinner, C.E.: A Priority Queueing System with Server-Walking Time. *Operations Research*, Vol. 15, No. 2 (1967), 278–285.

[12] Srinivasan, M.M. and Lee, H.-S.: An Analysis of a Prioritized Polling Mechanism for Cyclic Server Systems. Technical Report No. 87-6, 1988, Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, Michigan.

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

[14] Takagi, H.: Queueing Analysis of Polling Systems with Zero Switchover Times. TRL Research Report, TR87-0034, 1987, IBM Japan, Ltd., Tokyo.

[15] Takagi, H.: Queuing Analysis of Polling Models. *ACM Computing Surveys*, Vol. 20, No. 1 (1988), 5–28.

Appendix: Derivation of (2.25) and (2.26)

Since polling systems with LCFS service discipline have not been treated commonly, we give here a brief derivation of (2.25) and (2.26).

When $i \in E$, we consider three types of messages that arrive at station i. Messages that arrive during an idle period, which occur with probability $1 - \rho$, have zero waiting time. Messages that arrive while station i is in service, which occur with probability ρ_i, have the waiting time whose LST of the DF is given by

$$\frac{1 - \Theta_i^*(s)}{[s + \lambda_i - \lambda_i \Theta_i^*(s)]b_i} \tag{A.1}$$

Messages that arrive while one of the other stations is in service, which occur with probability $\rho - \rho_i$, have the waiting time whose LST of the DF is given by

$$\frac{1 - I_i^*[s + \lambda_i - \lambda_i\Theta_i^*(s)]}{[s + \lambda_i - \lambda_i\Theta_i^*(s)]E[I_i]} \tag{A.2}$$

where $I_i^*(s)$ and $E[I_i]$ are given by

$$I_i^*(\lambda_i - \lambda_i z) = \frac{H_{i-1}(1, \cdots, 1, \zeta, z, 1, \cdots, 1)}{H_{i-1}(1, \cdots, 1, \zeta, 1, 1, \cdots, 1)} \tag{A.3}$$

$$E[I_i] = \frac{P(0)}{\lambda} \frac{h_{i-1}(i)}{\lambda_i H_{i-1}(1, \cdots, 1, \zeta, 1, 1, \cdots, 1)} \tag{A.4}$$

Here ζ, defined in (2.19), and z are the $(i-1)$th and ith arguments of $H_{i-1}(.)$, respectively. $I_i^*(s)$ is the LST of the DF for the ***intervisit time*** I_i for station i, which is defined as the time interval from the instant when the server leaves station i to the instant when the server returns to station i for the first time. We note that expressions in (A.1) and (A.2) appear in the analysis of a single M/G/1 queue with LCFS and vacations (Section 3.5 of [8]). Unconditioning on the three types of messages, we obtain

$$W_i^*(s) = (1-\rho) \times 1 + \rho_i \times \frac{1 - \Theta_i^*(s)}{[s + \lambda_i - \lambda_i\Theta_i^*(s)]b_i} + (\rho - \rho_i) \times \frac{1 - I_i^*[s + \lambda_i - \lambda_i\Theta_i^*(s)]}{[s + \lambda_i - \lambda_i\Theta_i^*(s)]E[I_i]} \tag{A.5}$$

which reduces to (2.25).

When $i \in G$, we consider two types of messages that arrive at station i. Messages that arrive during an idle period, which occur with probability $1 - \rho$, have zero waiting time. Messages that arrive when the server is busy, which occur with probability ρ, have the waiting time whose LST of the DF is given by

$$\frac{1 - C_i^*[s + \lambda_i - \lambda_i B_i^*(s)]}{[s + \lambda_i - \lambda_i B_i^*(s)]E[C_i]} \tag{A.6}$$

where

$$C_i^*(\lambda_i - \lambda_i z) = \frac{H_{i-1}(1, \cdots, 1, \zeta, z, 1, \cdots, 1)}{H_{i-1}(1, \cdots, 1, \zeta, 1, 1, \cdots, 1)} \tag{A.7}$$

$$E[C_i] = \frac{P(0)}{\lambda} \frac{h_{i-1}(i)}{\lambda_i H_{i-1}(1, \cdots, 1, \zeta, 1, 1, \cdots, 1)} \tag{A.8}$$

$C_i^*(s)$ is the LST of the DF for the *cycle time* C_i for station i, which is defined as the time interval from the polling instant of station i to the next polling instant of the same station i. Unconditioning on the message types yields

$$W_i^*(s) = (1-\rho) \times 1 + \rho \times \frac{1 - C_i^*[s + \lambda_i - \lambda_i B_i^*(s)]}{[s + \lambda_i - \lambda_i B_i^*(s)]E[C_i]} \tag{A.9}$$

which reduces to (2.26).

Hideaki TAKAGI: IBM Research,
Tokyo Research Laboratory, 5-19 Sanban-cho
Chiyoda-ku, Tokyo 102, Japan