Journal of the Operations Research
Society of Japan
Vol. 43, No. 2, June 2000

SECOND MOMENTS OF THE WAITING TIME IN SYMMETRIC POLLING SYSTEMS

Seiya Kudoh Hideaki Takagi On Hashida
University of Tsukuba

(Received February 26, 1999; Revised September 30, 1999)

Abstract Multiple queue, cyclic service systems (called polling systems) have often been used as performance evaluation models in communication and production systems with cyclic resource allocation. However, most research has focused only on the mean waiting times due to prohibitively growing complexity in computing higher-order moments of the waiting time. This paper presents the explicit expressions for the second moments of the waiting time in symmetric exhaustive and gated service systems of two, three, and four queues with Poisson arrival processes. Numerical comparison reveals that they are ordered in the number of queues (increasingly/decreasingly for exhaustive/gated service systems, respectively) and bounded fairly tightly by those for single queue systems and by those for systems with infinitely many queues. Conjecture of their heavy traffic limits is also made.

1. Introduction

In polling systems, N queues are served by a single server in cyclic order. Such systems have diverse applications in communication networks and production systems in which a single resource is shared by multiple users with cyclic allocation. A particularly noteworthy application is the performance evaluation of token ring and FDDI protocols for high speed LANs. Two service disciplines among others with respect to the set of customers served at each visit of the server to a queue are exhaustive and gated service ones; they differ in respect to whether those customers that arrive at a queue in service are served during the current visit or reserved for the next visit. As usual in queueing systems, of main interest from customer's viewpoint is his/her waiting time. A special case in which $N = 1$ is referred to as a system with server vacations. Another extreme case in which $N = \infty$ (the load at each queue is infinitesimally small) is called a continuous polling system. Early studies of analysis and applications of polling systems are surveyed in [9, 14, 15, 16].

Consider a polling system with exhaustive or gated service discipline in which each queue has a Poisson arrival process and general service time distribution (like in an M/G/1 queue) and the switchover times of the server also has general distribution. Then it can be shown that the moments of any order of the customer waiting time in each queue can be evaluated, in principle, through the solution to a set of linear equations. However, explicit analytical formulas are unavailable for the waiting time moments, even for the mean waiting time except for a few special cases (for example, systems with $N = 2$ queues and symmetric systems). Thus major efforts have focused on the mean waiting times, such as the reduction in the number of the linear equations [6, 10] as well as efficient algorithms for their computation [8], and linear dependence relationships among them called pseudoconservation laws [2].

This paper is concerned with explicit evaluation of the second moments of the waiting time in exhaustive and gated service polling systems. The second moments generally give measures of variability and can be used in Chebyshev inequality for estimating the probability distribution. In our evaluation of the second moments, we confine ourselves to *symmetric* systems such that all the queues are statistically identical. The explicit formulas are available for the mean waiting times in symmetric polling systems of N queues $(1 \leq N \leq \infty)$ [7]. To the authors's best knowledge, the second moments (or variances) of the waiting time have been obtained for the following special cases only (besides for server vacation systems $N = 1$): an asymmetric exhaustive service system of $N = 2$ queues and zero switchover times [13], the same system but with nonzero switchover times [12], symmetric systems with deterministic service and switchover times of $N = 2$ and 3 queues [1], and symmetric continuous polling systems $(N = \infty)$ [3, 5]. Numerical computation procedures for the variance are given in [5], [8], and [11]. We present the explicit expressions for the second moments of the waiting time in symmetric exhaustive and gated service systems with $N = 2, 3$, and 4 queues. They have been obtained by solving the traditional set of linear equations using the software package *Mathematica* [19] for symbolic formula manipulation. We show by numerical examples that the second moments are ordered in N (increasingly/decreasingly for exhaustive/gated service systems, respectively) and fairly tightly bounded by those for the continuous polling systems $(N = \infty)$ and those for single queue systems $(N = 1)$. We also conjecture the heavy traffic limit forms of the second moments for general N. The heavy traffic case of an exhaustive service system of $N = 2$ queues and zero switchover times is studied in [4].

The rest of the paper is organized as follows. In Section 2 we describe our models more specifically and introduce notation used in the paper. In Sections 3 and 4, we consider symmetric exhaustive and gated service system, respectively, and give the second moments of the waiting time for systems of $N = 2, 3$ and 4 queues. In Section 5, they are plotted and compared numerically. We also discuss limiting forms of the second moments in heavy traffic conditions.

2. Models and Notation

Let us describe our systems and introduce notation for system parameters. The number of queues in the system is denoted by N. Queues are indexed by i, $i = 1, 2, \ldots, N$, in the order of server movement. We assume a Poisson arrival process of customers at rate λ_i for queue i. The Laplace-Stieltjes transform (LST) of the distribution function (DF), the mean, and the nth moment of service time of a customer at queue i are denoted by $B_i^*(s), b_i, b_i^{(n)}$, respectively. The total load offered to the system is then given by

$$\bar{\rho} \ := \ \sum_{i=1}^{N} \rho_i \quad ; \quad \rho_i := \lambda_i b_i \tag{1}$$

In each queue, customers are served on first-come first-served (FCFS) basis. The buffer of each queue has infinite capacity. We assume $\bar{\rho} < 1$ which makes the whole system stable. The LST of the DF, the mean, and the nth moment of time needed by the server to switch from queue i to queue $i+1$ (switchover time) are denoted by $R_i^*(s), r_i, r_i^{(n)}$, respectively. We also use the variance $\delta_i^2 := r_i^{(2)} - r_i^2$. The mean total switchover time is then given by

$$\bar{r} \ := \ \sum_{i=1}^{N} r_i \tag{2}$$

The switchover times are assumed to be independent of the arrival and service processes. The LST of the DF and the nth moment of the waiting time W_i of a customer at queue i are denoted by $W_i^*(s)$ and $E[W_i^n]$, respectively. A system in which the arrival and service processes in all queues and all the switchover times are independent of queue index i is called *symmetric*. For symmetric systems, we let $\rho_i = \rho = \lambda b$, and omit subscript i from other parameters; thus we get $\bar{\rho} = N\rho$ and $\bar{r} = Nr$.

3. Exhaustive Service Systems

We first consider exhaustive service systems. In such systems, the server continues to serve each queue until no customers remain there. Customers arriving at the queue in service are also served during the current visit of the server.

Let $L_i(t)$ denote the number of customers present in queue i at time t. We define the joint generating function (GF) for $[L_1(t), L_2(t), \ldots, L_N(t)]$ at time $t = \tau_i(m)$, i.e., at the instant when the server visits queue i in the mth polling cycle as $m \to \infty$ by

$$F_i(z_1, z_2, \ldots, z_N) \;:=\; \lim_{m\to\infty} E\left[\prod_{j=1}^{N} z_j^{L_j(\tau_i(m))}\right] \tag{3}$$

By counting the number of customers in each queue at $t = \tau_i(m)$ and $t = \tau_{i+1}(m)$, we get the relation

$$F_{i+1}(z_1, z_2, \ldots, z_N) = R_i^*\left[\sum_{j=1}^{N}(\lambda_j - \lambda_j z_j)\right] F_i\left(z_1, z_2, \ldots, z_{i-1}, \Theta_i^*\left[\sum_{\substack{j=1\\ j\neq i}}^{N}(\lambda_j - \lambda_j z_j)\right], z_{i+1}, \ldots, z_N\right) \tag{4}$$

where $\Theta_i^*(s)$ is the LST of the DF for the length of a busy period at queue i, and satisfies the equation

$$\Theta_i^*(s) \;=\; B_i^*[s + \lambda_i - \lambda_i \Theta_i^*(s)] \tag{5}$$

The LST of the DF for the waiting time W_i in an exhaustive service FCFS system is given by

$$W_i^*(s) \;=\; \frac{1-\bar{\rho}}{\bar{r}} \cdot \frac{1 - F_i(1 - s/\lambda_i)}{s - \lambda_i + \lambda_i B_i^*(s)} \tag{6}$$

where the marginal GF for $L_i(t)$ at time $t = \tau_i(m)$ as $m \to \infty$ is defined by

$$F_i(z) \;:=\; \lim_{m\to\infty} E[z^{L_i(\tau_i(m))}] = F_i(1, \ldots, 1, z, 1, \ldots, 1) \tag{7}$$

and z is the i th argument in $F_i(1, \ldots, 1, z, 1, \ldots 1)$. See [14] for the derivation of these formulas.

In order to find up to the second moments of the waiting time, let

$$f_i(j) \;:=\; \left.\frac{\partial F_i(z_1, z_2, \ldots, z_N)}{\partial z_j}\right|_{z_1=z_2=\ldots=z_N=1} \tag{8a}$$

$$f_i(j,k) := \left.\frac{\partial^2 F_i(z_1,z_2,\ldots,z_N)}{\partial z_j \partial z_k}\right|_{z_1=z_2=\ldots=z_N=1} \tag{8b}$$

$$f_i(j,k,l) := \left.\frac{\partial^3 F_i(z_1,z_2,\ldots,z_N)}{\partial z_j \partial z_k \partial z_l}\right|_{z_1=z_2=\ldots=z_N=1} \tag{8c}$$

$$i,j,k,l = 1,2,\ldots,N$$

where it is known that

$$f_i(i) = \frac{\lambda_i(1-\rho_i)\bar{r}}{1-\bar{\rho}} \quad ; \quad f_i(j) = \lambda_j \left[\sum_{k=j}^{i-1} r_k + \frac{\bar{r}\sum_{k=j+1}^{i-1}\rho_k}{1-\bar{\rho}}\right] \qquad j \neq i \tag{9}$$

From (6) with (9), the second moment of W_i is expressed as

$$E[W_i^2] = \frac{\lambda_i b_i^{(3)}}{3(1-\rho_i)} + \frac{\left[\lambda_i b_i^{(2)}\right]^2}{2(1-\rho_i)^2} + \frac{(1-\bar{\rho})f_i(i,i,i)}{3(1-\rho_i)\lambda_i^3\bar{r}} + \frac{(1-\bar{\rho})b_i^{(2)}f_i(i,i)}{2(1-\rho_i)^2\lambda_i\bar{r}} \tag{10}$$

Thus we need $f_i(i,i)$ and $f_i(i,i,i)$ to obtain $E[W_i^2]$. To do so, a set of N^3 equations for $\{f_i(j,k); i,j,k = 1,2,\ldots,N\}$ is derived from the partial derivative of (4) at $z_1 = z_2 = \cdots = z_N = 1$ as follows.

$$f_{i+1}(j,k) = \lambda_j\lambda_k r_i^{(2)} + r_i\lambda_k f_i(j) + r_i\lambda_j f_i(k) + f_i(i)\lambda_j\lambda_k\left[\frac{2r_ib_i}{1-\rho_i} + \frac{b_i^{(2)}}{(1-\rho_i)^3}\right]$$
$$+\frac{b_i}{1-\rho_i}\left[f_i(i,j)\lambda_k + f_i(i,k)\lambda_j\right] + f_i(j,k) + \frac{f_i(i,i)\lambda_j\lambda_k b_i^2}{(1-\rho_i)^2} \qquad j\neq i, k\neq i \tag{11a}$$

$$f_{i+1}(i,j) = \lambda_i\lambda_j r_i^{(2)} + \lambda_i r_i\left[f_i(j) + \frac{f_i(i)\lambda_j b_i}{1-\rho_i}\right] \qquad j \neq i \tag{11b}$$

$$f_{i+1}(i,i) = \lambda_i^2 r_i^{(2)} \tag{11c}$$

Similarly, a set of N^4 equations for $\{f_i(j,k,l); i,j,k,l = 1,2,\ldots,N\}$ is given by

$$f_{i+1}(j,k,l) = \lambda_j\lambda_k\lambda_l r_i^{(3)} + \left[\lambda_j\lambda_k f_i(l) + \lambda_j\lambda_l f_i(k) + \lambda_k\lambda_l f_i(j) + \frac{3\lambda_j\lambda_k\lambda_l b_i f_i(i)}{1-\rho_i}\right] r_i^{(2)}$$
$$+\left[\lambda_j f_i(k,l) + \lambda_k f_i(j,l) + \lambda_l f_i(j,k) + \frac{2b_i(\lambda_j\lambda_k f_i(i,l) + \lambda_j\lambda_l f_i(i,k) + \lambda_k\lambda_l f_i(i,j))}{1-\rho_i}\right] r_i$$
$$+3\lambda_j\lambda_k\lambda_l r_i\left\{\frac{b_i^2 f_i(i,i)}{(1-\rho_i)^2} + \frac{b_i^{(2)}f_i(i)}{(1-\rho_i)}\right\} + \lambda_j\lambda_k\lambda_l\left(\frac{3\lambda_i {b_i^{(2)}}^2}{(1-\rho_i)^5} + \frac{b_i^{(3)}}{(1-\rho_i)^4}\right) f_i(i)$$
$$+\left[\lambda_j\lambda_k f_i(i,l) + \lambda_j\lambda_l f_i(i,k) + \lambda_k\lambda_l f_i(i,j) + \frac{3\lambda_j\lambda_k\lambda_l b_i f_i(i,i)}{1-\rho_i}\right]\frac{b_i^{(2)}}{(1-\rho_i)^3}$$
$$+\frac{\lambda_j\lambda_k\lambda_l b_i^3 f_i(i,i,i)}{(1-\rho_i)^3} + \left[\lambda_j\lambda_k f_i(i,i,l) + \lambda_j\lambda_l f_i(i,i,k) + \lambda_k\lambda_l f_i(i,i,j)\right]\frac{b_i^2}{(1-\rho_i)^2}$$
$$+\left[\lambda_j f_i(i,k,l) + \lambda_k f_i(i,j,l) + \lambda_l f_i(i,j,k)\right]\frac{b_i}{1-\rho_i} + f_i(j,k,l) \qquad j\neq i, k\neq i, l\neq i \tag{12a}$$

$$f_{i+1}(i,j,k) = \lambda_i\lambda_j\lambda_k r_i^{(3)} + \lambda_i\left[\lambda_j f_i(k) + \lambda_k f_i(j) + \frac{2\lambda_j\lambda_k b_i f_i(i)}{1-\rho_i}\right] r_i^{(2)} + \lambda_i r_i f_i(j,k)$$

$$+\frac{(\lambda_j f_i(i,k)+\lambda_k f_i(i,j))\rho_i r_i}{1-\rho_i}+\frac{\lambda_i\lambda_j\lambda_k r_i b_i^2 f_i(i,i)}{(1-\rho_i)^2}+\frac{\lambda_i\lambda_j\lambda_k r_i b_i^{(2)} f_i(i)}{(1-\rho_i)^3} \qquad j\neq i, k\neq i \qquad (12b)$$

$$f_{i+1}(i,i,j)=\lambda_i^2\lambda_j r_i^{(3)}+\lambda_i^2 r_i^{(2)} f_i(j)+\frac{\lambda_i\lambda_j\rho_i f_i(i)}{1-\rho_i} \qquad j\neq i \qquad (12c)$$

$$f_{i+1}(i,i,i)=\lambda_i^3 r_i^{(3)} \qquad (12d)$$

For a symmetric system, all subscripts are dropped from $\lambda_i, b_i, b_i^{(n)}, r_i, r_i^{(n)}, \rho_i$ in (11a–c) and (12a–d). Although no explicit solution to (11) is available in general, $f_i(i,i)$ for a symmetric system is given by

$$f_i(i,i) = \frac{\delta^2\lambda^2 N(1-\rho)}{1-N\rho}+\frac{N(N-1)\lambda^3 r b^{(2)}}{(1-N\rho)^2}+\frac{N^2r^2\lambda^2(1-\rho)^2}{(1-N\rho)^2} \qquad (13)$$

where $\delta^2=r^{(2)}-r^2$.

We have not been able to derive an explicit expression for $f_i(i,i,i)$ even for a symmetric system with a general value of N. However, we can get $f_i(i,i,i)$ for specific values of N. Here we have solved these equations using *Mathematica* for symmetric systems of $N=2,3$, and 4 queues. The resulting second moments are as follows:

$N=2:$

$$\begin{aligned}E[W^2]=&\frac{2-\rho+2\rho^2}{3(1-\rho+\rho^2)}\left[\frac{r^{(3)}}{2r}+\frac{\lambda b^{(3)}}{1-2\rho}\right]\\&+\frac{(2-\rho+2\rho^2)\lambda b^{(2)}}{(1-2\rho)(1-\rho+\rho^2)}\left[\frac{\delta^2}{2r}+\frac{\lambda b^{(2)}}{1-2\rho}\right]+\frac{(2-3\rho+4\rho^2)\delta^2}{2(1-2\rho)(1-\rho+\rho^2)}\\&+\frac{(2-5\rho+6\rho^2-4\rho^3)r^2}{2(1-2\rho)^2(1-\rho+\rho^2)}+\frac{(3-2\rho)\lambda r b^{(2)}}{(1-2\rho)^2}\end{aligned} \qquad (14a)$$

$N=3:$

$$\begin{aligned}E[W^2]=&\frac{1-2\rho+2\rho^2-3\rho^3}{1-3\rho+4\rho^2-3\rho^3}\left[\frac{r^{(3)}}{3r}+\frac{\lambda b^{(3)}}{1-3\rho}\right]\\&+\frac{(9-28\rho+39\rho^2-45\rho^3+27\rho^4)\lambda b^{(2)}}{2(1-\rho)(1-3\rho)(1-3\rho+4\rho^2-3\rho^3)}\left[\frac{\delta^2}{3r}+\frac{\lambda b^{(2)}}{1-3\rho}\right]\\&+\frac{(2-7\rho+13\rho^2-12\rho^3)\delta^2}{(1-3\rho)(1-3\rho+4\rho^2-3\rho^3)}+\frac{(8-37\rho+76\rho^2-93\rho^3+54\rho^4)r^2}{3(1-3\rho)^2(1-3\rho+4\rho^2-3\rho^3)}\\&+\frac{3(5-3\rho)\lambda r b^{(2)}}{2(1-3\rho)^2}\end{aligned} \qquad (14b)$$

$N=4:$

$$\begin{aligned}E[W^2]=&\frac{4-14\rho+28\rho^2-51\rho^3+46\rho^4-40\rho^5}{3(1-2\rho)(1-3\rho+7\rho^2-7\rho^3+5\rho^4)}\left[\frac{r^{(3)}}{4r}+\frac{\lambda b^{(3)}}{1-4\rho}\right]\\&+\frac{(8-22\rho+51\rho^2-72\rho^3+66\rho^4-40\rho^5)\lambda b^{(2)}}{(1-\rho)(1-4\rho)(1-3\rho+7\rho^2-7\rho^3+5\rho^4)}\left[\frac{\delta^2}{4r}+\frac{\lambda b^{(2)}}{1-4\rho}\right]\\&+\frac{3(4-22\rho+68\rho^2-127\rho^3+130\rho^4-80\rho^5)\delta^2}{4(1-2\rho)(1-4\rho)(1-3\rho+7\rho^2-7\rho^3+5\rho^4)}\\&+\frac{(20-134\rho+444\rho^2-943\rho^3+1278\rho^4-1064\rho^5+480\rho^6)r^2}{4(1-2\rho)(1-4\rho)^2(1-3\rho+7\rho^2-7\rho^3+5\rho^4)}\end{aligned}$$

$$+\frac{2(7-4\rho)\lambda r b^{(2)}}{(1-4\rho)^2} \tag{14c}$$

4. Gated Service Systems

We next consider gated service systems. Here, the server serves only those customers that are found at the instant when the queue is visited. Those customers that arrive during the service period are set aside to be served in the next round of visit. For a gated service system, the joint GF $F_i(z_1, z_2, \dots, z_N)$ defined in (3) satisfies the equation

$$F_{i+1}(z_1, z_2, \dots, z_n) = R_i^*\left[\sum_{j=1}^{N}(\lambda_j - \lambda_j z_j)\right] F_i\left(z_1, z_2, \dots, z_{i-1}, B_i^*\left[\sum_{j=1}^{N}(\lambda_j - \lambda_j z_j)\right], z_{i+1}, \dots, z_N\right) \tag{15}$$

from which we can get $f_i(i)$, $f_i(i,i)$ and $f_i(i,i,i)$ in the same way as in Section 3. It is known that

$$f_i(i) = \frac{\lambda_i \bar{r}}{1-\bar{\rho}} \quad ; \quad f_i(j) = \lambda_j\left[\sum_{k=j}^{i-1} r_k + \frac{\bar{r}\sum_{k=j}^{i-1}\rho_k}{1-\bar{\rho}}\right] \qquad j \neq i \tag{16}$$

A set of N^3 equations for $\{f_i(j,k); i,j,k = 1,2,\dots,N\}$ is given by

$$\begin{aligned} f_{i+1}(j,k) &= \lambda_j\lambda_k r_i^{(2)} + r_i\lambda_k f_i(j) + r_i\lambda_j f_i(k) + f_i(i)\lambda_j\lambda_k(2r_ib_i + b_i^{(2)}) \\ &\quad + f_i(j,k) + b_i\lambda_k f_i(i,j) + b_i\lambda_j f_i(i,k) + b_i^2\lambda_j\lambda_k f(i,i) \qquad j \neq i, k \neq i \qquad (17a) \\ f_{i+1}(i,j) &= \lambda_i\lambda_j r_i^{(2)} + r_i\lambda_i f_i(j) + f_i(i)\lambda_i\lambda_j(2r_ib_i + b_i^{(2)}) + \lambda_i b_i f_i(i,j) \\ &\quad + \lambda_i\lambda_j b_i^2 f_i(i,i) \qquad j \neq i \qquad (17b) \\ f_{i+1}(i,i) &= \lambda_i^2 r_i^{(2)} + f_i(i)\lambda_i^2(2r_ib_i + b_i^{(2)}) + (\lambda_i b_i)^2 f_i(i,i) \qquad (17c) \end{aligned}$$

A set of N^4 equations for $\{f_i(i,j,k,l); i,j,k,l = 1,2,\dots,N\}$ is given by

$$\begin{aligned} f_{i+1}(j,k,l) &= \lambda_i\lambda_j\lambda_k r_i^{(3)} + [\lambda_k\lambda_l f_i(j) + \lambda_j\lambda_k f_i(l) + \lambda_j\lambda_l f_i(k) + 3\lambda_j\lambda_k\lambda_l b_i f_i(i)]\, r_i^{(2)} \\ &\quad + \Big[3\lambda_j\lambda_k\lambda_l b_i^{(2)} f_i(i) + 3\lambda_j\lambda_k\lambda_l b_i^2 f_i(i,i) + \lambda_j f_i(k,l) + \lambda_k f_i(j,l) + \lambda_l f_i(j,k) \\ &\quad + \{2\lambda_k\lambda_l f_i(i,j) + 2\lambda_j\lambda_l f_i(i,k) + 2\lambda_j\lambda_k f_i(i,l)\}\, b_i\Big]\, r_i \\ &\quad + \lambda_j\lambda_k\lambda_l f_i(i) b_i^{(3)} + [\lambda_k\lambda_l f_i(i,j) + \lambda_j\lambda_l f_i(i,k) + \lambda_j\lambda_k f_i(i,l) + 3\lambda_j\lambda_k\lambda_l b_i f_i(i,i)]\, b_i^{(2)} \\ &\quad + [\lambda_j f_i(i,k,l) + \lambda_k f_i(i,j,l) + \lambda_l f_i(i,j,k)]\, b_i \\ &\quad + [\lambda_k\lambda_l f_i(i,i,j) + \lambda_j\lambda_l f_i(i,i,k) + \lambda_j\lambda_k f_i(i,i,l)]\, b_i^2 + \lambda_j\lambda_k\lambda_l b_i^3 f_i(i,i,i) \\ &\qquad j \neq i, k \neq i, l \neq i \qquad (18a) \\ f_{i+1}(i,j,k) &= \lambda_i\lambda_j\lambda_k r_i^{(3)} + \lambda_i\,[3\lambda_j\lambda_k b_i f_i(i) + \lambda_k f_i(j) + \lambda_j f_i(k)]\, r_i^{(2)} \\ &\quad + \lambda_i\Big[3\lambda_j\lambda_k b_i^{(2)} f_i(i) + b_i\{2\lambda_k f_i(i,j) + 2\lambda_j f_i(i,k) + 3b_i\lambda_j\lambda_k f_i(i,i)\} + f_i(j,k)\Big]\, r_i \\ &\quad + \lambda_i\lambda_j\lambda_k b_i^{(3)} f_i(i) + \lambda_i\,[\lambda_k f_i(i,j) + \lambda_j f_i(i,k) + 3\lambda_j\lambda_k b_i f_i(i,i)]\, b_i^{(2)} \\ &\quad + \lambda_i\,[f_i(i,j,k) + \lambda_j b_i f_i(i,i,k) + \lambda_k b_i f_i(i,i,j)]\, b_i + \lambda_i\lambda_j\lambda_k b_i^3 f_i(i,i,i) \\ &\qquad j \neq i, k \neq i \qquad (18b) \\ f_{i+1}(i,i,j) &= \lambda_i^2\lambda_j r_i^{(3)} + \lambda_i^2\,[3\lambda_j b_i f_i(i) + f_i(j)]\, r_i^{(2)} \end{aligned}$$

$$+\lambda_i^2\left[3\lambda_j\left\{b_i^{(2)}f_i(i)+b_i^2 f_i(i,i)\right\}+2b_i f_i(i,j)\right]r_i$$
$$+\lambda_i^2\lambda_j b_i^{(3)}f_i(i)+3\lambda_i^2\lambda_j b_i b_i^{(2)}f_i(i,i)+\lambda_i^2 b_i^{(2)}f_i(i,j)$$
$$+\lambda_i^2 b_i^2 f_i(i,i,j)+\lambda_i^2\lambda_j b_i^3 f_i(i,i,i) \qquad j\neq i \tag{18c}$$
$$f_{i+1}(i,i,i)=\lambda_i^3 r_i^{(3)}+\lambda_i^3 b_i^{(3)}f_i(i)+3\lambda_i^3 b_i r_i^{(2)}f_i(i)+3\lambda_i^3 b_i^2 r_i f_i(i,i)$$
$$+3\lambda_i^3\left(r_i f_i(i)+b_i f_i(i,i)\right)b_i^{(2)}+\lambda_i^3 b_i^3 f_i(i,i,i) \tag{18d}$$

For a symmetric system we get

$$f_i(i,i) = \frac{\delta^2\lambda^2 N}{(1-N\rho)(1+\rho)}+\frac{N^2\lambda^3 r b^{(2)}}{(1-N\rho)^2(1+\rho)}+\frac{N^2 r^2\lambda^2}{(1-N\rho)^2} \tag{19}$$

The LST of the DF for the waiting time W_i in a gated service FCFS system is given by

$$W_i^*(s) = \frac{1-\bar{\rho}}{\bar{r}}\cdot\frac{F_i\left[B_i^*(s)\right]-F_i\left(1-s/\lambda_i\right)}{s-\lambda_i+\lambda_i B_i^*(s)} \tag{20}$$

from which we get

$$E[W_i^2] = \frac{(1-\bar{\rho})(1+\rho_i+\rho_i^2)f_i(i,i,i)}{3\lambda_i^3\bar{r}}+\frac{(1-\bar{\rho})b_i^{(2)}f_i(i,i)}{2\lambda_i\bar{r}} \tag{21}$$

We have obtained the second moments of the waiting time in symmetric gated service systems of $N=2,3,$ and 4 queues as follows:

$N=2:$

$$E[W^2]=\frac{(1+\rho+\rho^2)(2+\rho-2\rho^2)}{3(1+\rho)(1+\rho^2-\rho^3)}\left[\frac{r^{(3)}}{2r}+\frac{\lambda b^{(3)}}{1-2\rho}\right]$$
$$+\frac{(2+4\rho+6\rho^2-3\rho^4-2\rho^5)\lambda b^{(2)}}{(1+\rho)^2(1-2\rho)(1+\rho^2-\rho^3)}\left[\frac{\delta^2}{2r}+\frac{\lambda b^{(2)}}{1-2\rho}\right]$$
$$+\frac{(1+\rho+\rho^2)(2+5\rho+2\rho^2-4\rho^3)r^2}{2(1+\rho)(1-2\rho)^2(1+\rho^2-\rho^3)}+\frac{(1+\rho+\rho^2)(2+3\rho+8\rho^2-8\rho^3)\delta^2}{2(1+\rho)(1-2\rho)(1+\rho^2-\rho^3)}$$
$$+\frac{(5+3\rho+2\rho^2)\lambda r b^{(2)}}{(1+\rho)(1-2\rho)^2} \tag{22a}$$

$N=3:$

$$E[W^2]=\frac{(1+\rho+\rho^2)(1+\rho-\rho^3-3\rho^4)}{(1-\rho)(1+\rho)(1+\rho+4\rho^2+2\rho^3+3\rho^4)}\left[\frac{r^{(3)}}{3r}+\frac{\lambda b^{(3)}}{1-3\rho}\right]$$
$$+\frac{(9+20\rho+54\rho^2+18\rho^3-8\rho^4-75\rho^5-63\rho^6-27\rho^7)\lambda b^{(2)}}{2(1-\rho)(1+\rho)^2(1-3\rho)(1+\rho+4\rho^2+2\rho^3+3\rho^4)}\left[\frac{\delta^2}{3r}+\frac{\lambda b^{(2)}}{1-3\rho}\right]$$
$$+\frac{(1+\rho+\rho^2)(8+14\rho+24\rho^2+\rho^3-12\rho^4-27\rho^5)r^2}{3(1-\rho)(1+\rho)(1-3\rho)^2(1+\rho+4\rho^2+2\rho^3+3\rho^4)}$$
$$+\frac{(1+\rho+\rho^2)(2+2\rho+12\rho^2-5\rho^3+3\rho^4-18\rho^5)\delta^2}{(1-\rho)(1+\rho)(1-3\rho)(1+\rho+4\rho^2+2\rho^3+3\rho^4)}$$
$$+\frac{3(7+4\rho+3\rho^2)\lambda r b^{(2)}}{2(1+\rho)(1-3\rho)^2} \tag{22b}$$

$N = 4:$

$$E[W^2] = \frac{(4+2\rho-2\rho^2-15\rho^3-31\rho^4+7\rho^5+18\rho^6+40\rho^7)}{3\,(1+\rho)\,(1-\rho+4\rho^2-5\rho^3)\,(1-\rho+\rho^2+2\rho^3)}\left[\frac{r^{(3)}}{4r}+\frac{\lambda b^{(3)}}{1-4\rho}\right]$$
$$+\frac{(8+4\rho+41\rho^2-85\rho^3-9\rho^4-144\rho^5+66\rho^6+116\rho^7+80\rho^8)\,\lambda b^{(2)}}{(1+\rho)^2\,(1-4\rho)\,(1-\rho+4\rho^2-5\rho^3)\,(1-\rho+\rho^2-2\rho^3)}$$
$$\times\left[\frac{\delta^2}{4r}+\frac{\lambda b^{(2)}}{1-4\rho}\right]+\frac{2\,(9+5\rho+4\rho^2)\,\lambda r b^{(2)}}{(1+\rho)\,(1-4\rho)^2}$$
$$+\frac{(20+10\rho+70\rho^2-75\rho^3-83\rho^4-197\rho^5+50\rho^6+104\rho^7+160\rho^8)\,r^2}{4\,(1+\rho)\,(1-4\rho)^2\,(1-\rho+4\rho^2-5\rho^3)\,(1-\rho+\rho^2-2\rho^3)}$$
$$+\frac{(12-2\rho+90\rho^2-121\rho^3+51\rho^4-355\rho^5+138\rho^6-16\rho^7+320\rho^8)\,\delta^2}{4\,(1+\rho)\,(1-4\rho)\,(1-\rho+4\rho^2-5\rho^3)\,(1-\rho+\rho^2-2\rho^3)} \qquad (22c)$$

5. Numerical Results, Comparison, and Heavy Traffic Limits

Let us compare the waiting times in various symmetric systems with the same total load $\bar{\rho} = N\rho$. To avoid the effects from the variability in switchover times, we assume that all the switchover times are deterministic, thus $r^{(2)} = r^2$, $r^{(3)} = r^3$, and $\delta^2 = 0$. In such a case, we have the second moment of the waiting time in a continuous polling system, that is a limiting form obtained by making $N \to \infty$ with $\bar{\rho} = N\lambda b$ and $\bar{r} = Nr$ held at fixed values [5]:

$$E[W^2] = \frac{2\bar{\rho}b^{(3)}}{3(2-\bar{\rho})(1-\bar{\rho})b}+\frac{\bar{\rho}^2(3-\bar{\rho})\left[b^{(2)}\right]^2}{3(2-\bar{\rho})(1-\bar{\rho})^2b^2}+\frac{\bar{\rho}\bar{r}b^{(2)}}{(1-\bar{\rho})^2b}+\frac{\bar{r}^2}{3(1-\bar{\rho})^2} \qquad (23)$$

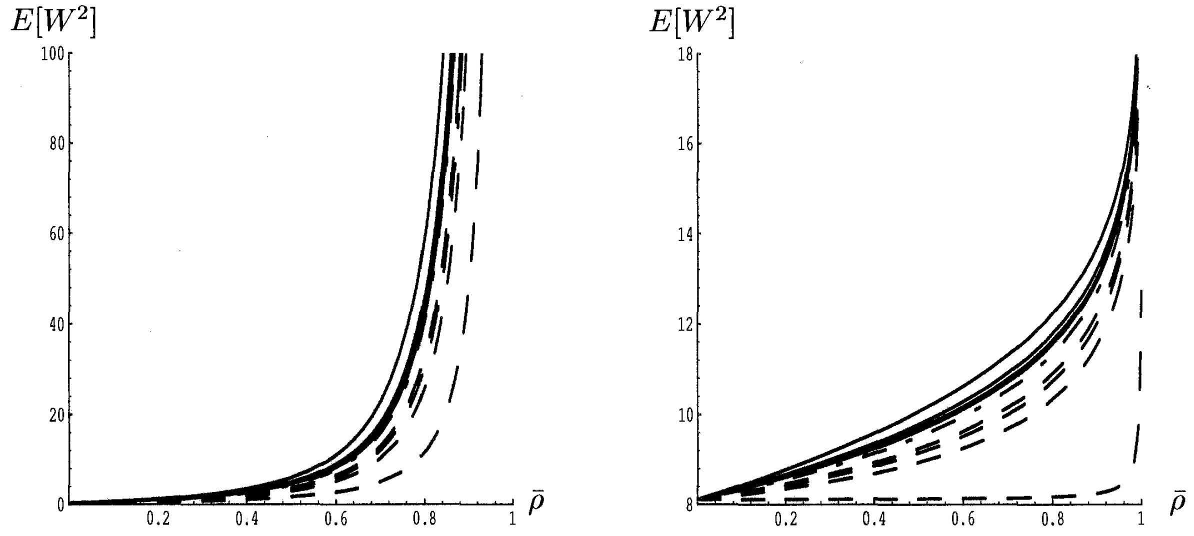
Note that a continuous polling system has no distinction of exhaustive and gated service because the load at each queue is infinitesimally small. The second moment of the waiting time in single queue systems ($N = 1$) without the assumption of deterministic switchover times are given by [17]

$$\text{exhaustive service}: E[W^2] = \frac{1}{3}\left[\frac{r^{(3)}}{r}+\frac{\lambda b^{(3)}}{1-\rho}\right]+\frac{\lambda b^{(2)}}{2(1-\rho)}\left[\frac{r^{(2)}}{r}+\frac{\lambda b^{(2)}}{1-\rho}\right] \qquad (24a)$$

$$\text{gated service}: E[W^2] = \frac{1}{3}\left[\frac{r^{(3)}}{r}+\frac{\lambda b^{(3)}}{1-\rho}\right]+\frac{\lambda b^{(2)}}{2(1-\rho)}\left[\frac{r^{(2)}}{r}+\frac{\lambda b^{(2)}}{1-\rho}\right]$$
$$+\frac{(1+\rho+\rho^2)\lambda r b^{(2)}}{(1-\rho)(1-\rho^2)}+\frac{\rho(1+2\rho)r^{(2)}}{1-\rho^2}$$
$$+\frac{2\rho^3r^2}{(1-\rho)(1-\rho^2)} \qquad (24b)$$

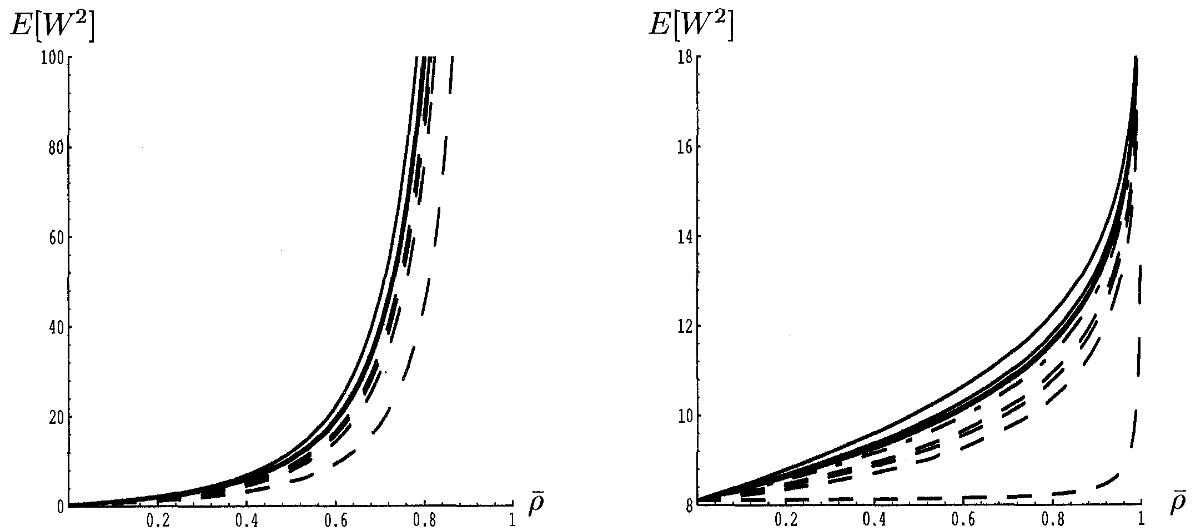
In Figures 1 and 2, we plot $E[W^2]$ against $\bar{\rho}$ with fixed b ($=1$) and $\bar{r}$ ($=1$ or $=100$) for different number N of queues. The service times are either deterministic or exponentially distributed. To compare, let us use $E[W^2]_{e,N}$, $E[W^2]_{g,N}$, and $E[W^2]_\infty$ to denote the second moment of the waiting time in an exhaustive service system of N queues, in a gated service system of N queues, and in a continuous polling system, respectively. We observe in the figures monotone ordering for the second moment of the waiting time:

$$E[W^2]_{e,1} < E[W^2]_{e,2} < E[W^2]_{e,3} < E[W^2]_{e,4} < E[W^2]_\infty$$
$$< E[W^2]_{g,4} < E[W^2]_{g,3} < E[W^2]_{g,2} < E[W^2]_{g,1} \qquad (25)$$



(a) $\bar{r} = 1$ (b) $\bar{r} = 100$

Figure 1: Second moments of the waiting time in exhaustive and gated service systems (service times are deterministic).



(a) $\bar{r} = 1$ (b) $\bar{r} = 100$

Figure 2: Second moments of the waiting time in exhaustive and gated service systems (service times are exponentially distributed).

Legend for the figures

– – – – – –	:	exhaustive service systems $N = 1$, 2, 3, and 4 from below
————	:	gated service systems $N = 1$, 2, 3, and 4 from top
— - — - —	:	continuous polling systems

We also see that the second moments are fairly tightly bounded by those for the continuous polling systems ($N = \infty$) and those for single queue systems ($N = 1$) if the switchover times are small and the total load $\overline{\rho}$ is not so large. This suggests possible use of the formula (23), (24a) and (24b) for approximating $E[W^2]$ in a system of N queues where $2 \leq N < \infty$.

While it seems difficult to obtain the second moments of the waiting time for general values of N at present, we may conjecture the limiting forms in the heavy traffic condition $\overline{\rho} \to 1$. In (14a–c) and (22a–c), we see that the most significant contributions in this limit come from terms of order $O(1/(1-\overline{\rho})^2)$. From the coefficients of these terms, we induce the heavy traffic limits

$$\text{exhaustive service}: E[W^2] \approx \frac{2\left[N\lambda b^{(2)}\right]^2}{3(1-N\rho)^2} + \frac{N(N-1)\lambda r b^{(2)}}{(1-N\rho)^2} + \frac{[(N-1)r]^2}{3(1-N\rho)^2} \tag{26a}$$

$$\text{gated service}: E[W^2] \approx \frac{2(N^2+N+1)\left[N\lambda b^{(2)}\right]^2}{3(N+1)^2(1-N\rho)^2} + \frac{N(N^2+N+1)\lambda r b^{(2)}}{(N+1)(1-N\rho)^2} + \frac{(N^2+N+1)r^2}{3(1-N\rho)^2} \tag{26b}$$

We note that these agree as $N \to \infty$ with the heavy traffic limit in a continuous polling system which can be obtained from (23) as

$$\text{continuous polling system}: E[W^2] \approx \frac{2\left[b^{(2)}\right]^2}{3(1-\overline{\rho})^2 b^2} + \frac{\overline{r} b^{(2)}}{(1-\overline{\rho})^2 b} + \frac{\overline{r}^2}{3(1-\overline{\rho})^2} \tag{26c}$$

For the systems with zero switchover times, the above heavy traffic limits reduce to

$$\text{exhaustive service}: E[W^2] \approx \frac{2[N\lambda b^{(2)}]^2}{3(1-N\rho)^2} \tag{27a}$$

$$\text{gated service}: E[W^2] \approx \frac{2(N^2+N+1)[N\lambda b^{(2)}]^2}{3(N+1)^2(1-N\rho)^2} \tag{27b}$$

which agree with the special cases of the recent result for the n th moments of the waiting time [18]

$$\text{exhaustive service}: E[W^n] \approx \frac{n![N\lambda b^{(2)}]^n}{(n+1)(1-N\rho)^n} \tag{28a}$$

$$\text{gated service}: E[W^n] \approx \frac{n!(N^n + N^{n-1} + \cdots + N^2 + N + 1)[N\lambda b^{(2)}]^n}{(n+1)(N+1)^n(1-N\rho)^n} \tag{28b}$$

where $n = 1, 2, \ldots$.

References

[1] Y. J. Aminetzah: *An Exact Approach to the Polling System* (Ph.D. dissertation. Department of Electrical Engineering, McGill University, Montreal, Quebec, Canada, 1975).

[2] O. J. Boxma: Workloads and waiting times in single-server systems with multiple customer classes. *Queueing Systems*, **5** (1989) 185–214.

[3] E. G. Coffman, Jr. and E. N. Gilbert: A continuous polling system with constant service times. *IEEE Transactions on Information Theory*, **IT-33** (1986) 584–591.

[4] E. G. Coffman, Jr., A. A. Puhalskii, and M. I. Reiman: Polling systems with zero switchover times: a heavy-traffic averaging principle. *The Annals of Applied Probability*, **5** (1995) 681–719.

[5] M. J. Ferguson: Computation of the variance of the waiting time for token rings. *IEEE Journal on Selected Areas in Communications*, **SAC-4** (1986) 775–782.

[6] M. J. Ferguson and Y. J. Aminetzah: Exact results for nonsymmetric token ring systems. *IEEE Transactions on Communications*, **COM-33** (1985) 223–231. Correction in: G. L. Choudhury and H. Takagi: Comments on "Exact results for nonsymmetric token ring systems." *IEEE Transactions on Communications*, **38** (1990) 1125–1127.

[7] O. Hashida: Analysis of multiqueue. *Review of the Electrical Communication Laboratories*, **20** (1972) 189–199.

[8] A. G. Konheim, H. Levy, and M. M. Srinivasan: Descendant set: An efficient approach for the analysis of polling systems. *IEEE Transactions on Communications*, **42** (1994) 1245–1253.

[9] H. Levy and M. Sidi: Polling systems: applications, modeling, and optimization. *IEEE Transactions on Communications*, **38** (1990) 1750-1760.

[10] D. Sarkar and W. I. Zangwill: Expected waiting time for nonsymmetric cyclic queueing systems - Exact results and applications. *Management Science*, **35** (1989) 1463–1474.

[11] D. Sarkar and W. I. Zangwill: Waiting time variance for nonsymmetric cyclic queueing systems. (Preprint, AT&T Bell Laboratories, 1991).

[12] M. M. Srinivasan, H. Levy, and A. G. Konheim: The individual station technique for the analysis of cyclic polling systems. *Naval Research Logistics*, **43** (1986) 79–101.

[13] L. Takács: Two queues attended by a single server. *Operations Research*, **16** (1968) 639–650.

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

[15] H. Takagi: Queuing analysis of polling models. *ACM Computing Surveys*, **20** (1988) 5–28.

[16] H. Takagi: Application of polling models to computer networks. *Computer Networks and ISDN Systems*, **22** (1991) 193–211.

[17] H. Takagi: *Queueing Analysis, A Foundation of Performance Evaluation, Volume 1: Vacation and Priority Systems, Part 1* (Elsevier, Amsterdam, 1991).

[18] R. D. van der Mei: Polling systems in heavy traffic: higher moments of the delay. In V. Ramaswami and P. E. Wirth (eds.): *Teletraffic Analysis for the Communication Age* (Elsevier, Amsterdam, 1997), 275–284.

[19] S. Wolfram: *Mathematica: A System for Doing Mathematics by Computer*, Second edition (Addison-Wesley, Reading, Massachusetts, 1991).

Hideaki Takagi
Institute of Policy and Planning Sciences
University of Tsukuba
1-1-1 Tennoudai, Tsukuba-shi
Ibaraki 305-8573, Japan
E-mail: `takagi@shako.sk.tsukuba.ac.jp`