Journal of the Operations Research
Society of Japan
Vol. 35, No. 3, September 1992

ANALYSIS OF AN M/G/1//N QUEUE WITH MULTIPLE SERVER VACATIONS, AND ITS APPLICATION TO A POLLING MODEL

Hideaki Takagi
IBM Research, Tokyo Research Laboratory

(Received July 15, 1991; Revised January 23, 1992)

Abstract. Queues with a finite population of customers and the server's occasional unavailable periods (called vacations) are studied in detail. We first consider M/G/1//N queueing system where the server takes repeated vacations until it finds a customer in the queue after emptying the queue. For the steady state, we obtain the performance measures such as the system throughput and mean waiting time from the known analysis of a regenerative cycle of the busy and vacation periods. We also obtain the Laplace-Stieltjes transform of the distribution function for the waiting time of a customer by applying the method of supplementary variables to the joint distribution of the queue size and the elapsed service or vacation times at an arbitrary point in time. These results are then applied to the steady-state analysis of a multiple-queue, cyclic-service (polling) model with a finite population of customers, which can represent a token ring network for several computers each with a finite number of interactive users. Some numerical results for symmetric systems are shown.

1. Introduction

Recently, queues with server's vacations have been studied extensively from its own theoretical interest as well as its applicability to many engineering systems such as computers, communication networks, and manufacturing systems. According to surveys by Doshi [1, 2], all previous studies of vacation models assume an infinite population of customers. However, it is equally or more important to study queueing systems with a finite, fixed population of customers as a moderate number of service requestors are usually involved in the above mentioned engineering systems.†

In the first part of this paper, we consider an M/G/1//N queueing system with multiple vacations and exhaustive service, which is described specifically as follows. We assume that N is the total number of customers in the system. Our system consists of a "source" and a "service facility," where the service facility contains the queue and a server. Each customer is either in the source or in the service facility at any time. A customer in source arrives at the service facility at rate λ. In other words, the time until the arrival of each customer in the source is exponentially distributed with mean $1/\lambda$. The service time of a customer is assumed to be generally distributed. The server begins a vacation each time when the queue

† Doshi [3] studies rather general vacation models for which the waiting time distribution can be decomposed into the waiting time distribution in corresponding queues without vacations and some other distribution. He assumes that the sequences of interarrival times are stochastically equivalent between the queue with vacations and the corresponding queue without vacations. The M/G/1//N queue with multiple vacations considered in this paper does not fall into a category of Doshi's general vacation model because the arrival pattern is different from the M/G/1//N queue without vacations. Hence, we need a new analysis for the M/G/1//N queue with vacations.

becomes empty (*exhaustive service*). If the server returns from a vacation to find the queue not empty, the vacation period ends; otherwise it begins another vacation, and continues in this manner until it finds at least one customer in the queue upon returning from a vacation (*multiple vacations*). We assume that the length of each vacation is an independent and identically distributed random variable. Performance measures in this system include the mean customer response time and the throughput of the system.

As an application of the steady-state analysis of an M/G/1//N system, we can consider a system of M queues that are attended by a single server in cyclic order. Here we assume that each queue has a separate source of a finite, fixed population of customers. Customers of each queue return to its original source after service completion. A certain time is required for the server to switch from one queue to another. Such cyclic order of service to multiple queues has been called a *polling model*. See Takagi [13, 15] for surveys of queueing analysis of polling models and its applications to computer and communication systems, including the token-ring local area network. Most previous analysis of polling models assumes an infinite population of customers in the source. As far as the author knows, only Ibe and Trivedi [4] analyze polling systems with a finite population of customers; they use a technique called *stochastic Petri nets* which must presume exponential distributions for both service times and switchover times.

In the second part of the paper, we provide for the first time a probabilistic analysis of a polling model with a finite population of customers and generally distributed service and switchover times. The performance measures for each queue in a polling model are similar to those for the M/G/1//N queue. An overall measure is the mean time it takes the server to complete one cycle of service to all queues, which is called the *mean cycle time*. A classical application of the M/G/1//N queue to a computer system is the interactive processing system, in which each user can place at most one outstanding service request at any time [7 (sec. 4.11)]. By combining this model and the polling model for the token-ring network, our result for the finite-population polling model can be used for the performance analysis of a group of computers, each with a finite number of interactive usres, connected by the token ring.

The rest of the paper is organized as follows. In Section 2, we analyze a single M/G/1//N queue with multiple vacations and exhaustive service. We obtain the mean response time and the system throughput by capitalizing on the known results for the busy period of an M/G/1//N queue and the regenerative consideration of vacation cycles. We then employ the method of supplementary variables to analyze the joint distribution of the queue size and the elapsed service or vacation times at an arbitrary point in time, and obtain the distribution of the customer waiting time. In Section 3, we show that these results can be applied to a polling model with a finite population of customers. Simplified formulation and numerical results are presented for symmetric polling models. We conclude in Section 4 by reviewing the results obtained in this paper and suggesting future research subjects.

2. M/G/1//N with Multiple Server Vacations and Exhaustive Service

We consider a single-server queueing system with N customers, each of which if in the source arrives at the service facility at rate λ. We denote by $B(x), b(x)$, and b the distribution function (DF), the probability density function (pdf), and the mean of the service time, respectively. Let $B^*(s)$ be the Laplace-Stieltjes transform (LST) of $B(x)$. The length V of each vacation is independent and identically distributed with its DF, pdf, and mean denoted by $V(x), v(x), E[V]$, respectively. Let $V^*(s)$ be the LST of $V(x)$. We assume that $B(x)$ and $V(x)$ are continuous for $x \geq 0$, and that $B(0) = V(0) = 0$.

2.1 PERFORMANCE MEASURES.

In the steady state, one of the performance measures in our system is the ***mean response time*** $E[T]$ defined as the mean time from the arrival of a customer to its service completion, namely, the mean time a customer spends in the service facility. Otherwise the customer stays in the source. The mean time that each customer passes a cycle of staying in the source and in the service facility is given by $E[T] + 1/\lambda$. Therefore, the mean number of customers served per unit time in the system, called the ***throughput*** γ of the system, is given by $N/(E[T]+1/\lambda)$. Note that $\rho' = \gamma b$ is called the ***carried load***, that is, the long-run fraction of the time that the server is busy. Furthermore, the throughput is the fraction of the arrival rate of $N - E[L]$ customers on the average, where $E[L]$ is the mean number of customers in the service facility. Thus, we have the following relationship among these performance measures:

$$\gamma = \frac{N}{E[T] + 1/\lambda} = \frac{\rho'}{b} = \lambda(N - E[L]) \tag{2.1}$$

from which we get

$$E[T] = \frac{Nb}{\rho'} - \frac{1}{\lambda} \tag{2.2}$$

From (2.1) and (2.2) we get

$$\gamma E[T] = E[L] \tag{2.3}$$

which is an instance of Little's theorem applied to those customers that are present in the service facility. We are also interested in the waiting time W of a customer in the queue. The LST of the DF for W is denoted by $W^*(s)$. The mean waiting time is given by

$$E[W] = E[T] - b \tag{2.4}$$

Let $E[\Theta]$ be the mean length of a busy period and $E[I]$ be the mean length of a vacation period. We consider the number of customers in the service facility as the system state. It then follows from the assumption of memoryless arrival process that those points in time at which each vacation period is started are the regeneration points of the system state. Therefore, the long-run fraction of time that the server is in the busy period is given by

$$\rho' = \frac{E[\Theta]}{E[\Theta] + E[I]} \tag{2.5}$$

Let α_k be the probability that there are k arrivals during a vacation period in which at least one arrival has occurred, where $k = 1, 2, \cdots, N$. In other words, there are k customers at the beginning of a busy period with probability α_k. We then have

$$E[\Theta] = \sum_{k=1}^{N} \alpha_k E[\Theta_k] \tag{2.6}$$

where Θ_k denotes the length of a busy period which is started with k customers in the M/G/1//N system, where $k = 1, 2, \cdots, N$. The LST $\Theta_k^*(s)$ of the DF for Θ_k is given by Jaiswal [5 (sec. II.2)] as

$$\Theta_k^*(s) = \frac{\sum_{n=0}^{N-k} \binom{N-k}{n} \frac{1}{\varphi_{n-1}^*(s)}}{\sum_{n=0}^{N} \binom{N}{n} \frac{1}{\varphi_{n-1}^*(s)}} \qquad k = 1, 2, \cdots, N \tag{2.7}$$

where

$$\varphi^*_{-1}(s) \triangleq 1 \quad ; \quad \varphi^*_n(s) \triangleq \prod_{j=0}^{n} \frac{B^*(s+j\lambda)}{1-B^*(s+j\lambda)} \quad n \geq 0 \tag{2.8}$$

The mean of Θ_k is then given by

$$E[\Theta_k] = b \sum_{n=1}^{N} \left[\binom{N}{n} - \binom{N-k}{n} \right] \frac{1}{\zeta_{n-1}} \tag{2.9}$$

where

$$\zeta_0 \triangleq 1 \quad ; \quad \zeta_n \triangleq \prod_{j=1}^{n} \frac{B^*(j\lambda)}{1-B^*(j\lambda)} \quad n \geq 1 \tag{2.10}$$

The probability f_k that there are k customers in the queue at the end of a vacation is given by

$$\begin{aligned} f_k &= \binom{N}{k} \int_0^\infty (1-e^{-\lambda x})^k (e^{-\lambda x})^{N-k} dV(x) \\ &= \binom{N}{k} \sum_{n=0}^{k} \binom{k}{n} (-1)^n V^*[(N-k+n)\lambda] \quad k = 0,1,2,\cdots,N \end{aligned} \tag{2.11}$$

Using (2.11), we have

$$\alpha_k = \frac{f_k}{1-f_0} \quad k = 1,2,\cdots,N \tag{2.12}$$

Let $I^*(s)$ be the LST of the DF for the length of a vacation period. For a multiple vacation model, this is given by†

$$I^*(s) = \frac{V^*(s) - V^*(s+N\lambda)}{1-V^*(s+N\lambda)} \tag{2.13}$$

which yields

$$E[I] = \frac{E[V]}{1-V^*(N\lambda)} \tag{2.14}$$

Substituting (2.9) and (2.12) into (2.6), we get

$$E[\Theta] = \frac{b}{1-V^*(N\lambda)} \sum_{n=1}^{N} \binom{N}{n} \frac{1-V^*(n\lambda)}{\zeta_{n-1}} \tag{2.15}$$

Using (2.14) and (2.15) in (2.5), we obtain

$$\rho' = \frac{b \sum_{n=1}^{N} \binom{N}{n} \frac{1-V^*(n\lambda)}{\zeta_{n-1}}}{E[V] + b \sum_{n=1}^{N} \binom{N}{n} \frac{1-V^*(n\lambda)}{\zeta_{n-1}}} \tag{2.16}$$

The mean response time $E[T]$ can then be calculated from (2.2) as

$$E[T] = \frac{N E[V]}{\sum_{n=1}^{N} \binom{N}{n} \frac{1-V^*(n\lambda)}{\zeta_{n-1}}} + Nb - \frac{1}{\lambda} \tag{2.17}$$

† The expression for $I^*(s)$ given by Levy and Yechiali [8] is incorrect.

2.2 DISTRIBUTION OF THE WAITING TIME.

We proceed to the analysis of the system state at an arbitrary time using the method of supplementary variables. Let $L(t)$ be the number of customers in the service facility at time t. Let $X^*(t)$ be the amount of service already received by a customer in service (the elapsed service time) at time t in the busy period. Similarly, let $U^*(t)$ be the elapsed vacation time at time t in the vacation period. In addition, we define the state $\xi(t)$ of the server at time t by

$$\xi(t) \triangleq \begin{cases} 0 & \text{on vacation at } t \\ 1 & \text{busy at } t \end{cases} \tag{2.18}$$

We consider the steady-state joint distributions

$$Q_k(x)dx \triangleq \lim_{t\to\infty} \text{Prob}[\xi(t)=0, L(t)=k, x<U^*(t)\le x+dx] \quad 0\le k\le N \tag{2.19a}$$

$$P_k(x)dx \triangleq \lim_{t\to\infty} \text{Prob}[\xi(t)=1, L(t)=k, x<X^*(t)\le x+dx] \quad 1\le k\le N \tag{2.19b}$$

Equations for $\{Q_k(x)\}$ and $\{P_k(x)\}$ can be obtained by extending the arguments for the M/G/1 system by Keilson and Kooharian [6] to our system. They are given by

$$\frac{\partial Q_0(x)}{\partial x} + [N\lambda + \overline{v}(x)]Q_0(x) = 0 \tag{2.20a}$$

$$\frac{\partial Q_k(x)}{\partial x} + [(N-k)\lambda + \overline{v}(x)]Q_k(x) = (N-k+1)\lambda Q_{k-1}(x) \quad 1\le k\le N-1 \tag{2.20b}$$

$$\frac{\partial P_1(x)}{\partial x} + [(N-1)\lambda + \overline{b}(x)]P_1(x) = 0 \tag{2.20c}$$

$$\frac{\partial P_k(x)}{\partial x} + [(N-k)\lambda + \overline{b}(x)]P_k(x) = (N-k+1)\lambda P_{k-1}(x) \quad 2\le k\le N \tag{2.20d}$$

where

$$\overline{b}(x) \triangleq \frac{b(x)}{1-B(x)} \quad ; \quad \overline{v}(x) \triangleq \frac{v(x)}{1-V(x)} \tag{2.21}$$

The boundary conditions are given by

$$Q_0(0) = \int_0^\infty Q_0(x)\overline{v}(x)dx + \int_0^\infty P_1(x)\overline{b}(x)dx \tag{2.22a}$$

$$Q_k(0) = 0 \quad 1\le k\le N \tag{2.22b}$$

$$P_k(0) = \int_0^\infty Q_k(x)\overline{v}(x)dx + \int_0^\infty P_{k+1}(x)\overline{b}(x)dx \quad 1\le k\le N-1 \tag{2.22c}$$

$$P_N(0) = \int_0^\infty Q_N(x)\overline{v}(x)dx \tag{2.22d}$$

The notmalization condition is given by

$$\sum_{k=0}^{N}\int_0^\infty Q_k(x)dx + \sum_{k=1}^{N}\int_0^\infty P_k(x)dx = 1 \tag{2.23}$$

To solve the above equations, we introduce the transforms by

$$\overline{Q}(z,x) \triangleq \frac{\sum_{k=0}^{N} Q_k(x)z^{N-k}}{1-V(x)} \quad ; \quad \overline{P}(z,x) \triangleq \frac{\sum_{k=1}^{N} P_k(x)z^{N-k}}{1-B(x)} \tag{2.24}$$

Equations (2.20)-(2.23) are then converted into

$$\frac{\partial \overline{Q}(z,x)}{\partial x} + \lambda(z-1)\frac{\partial \overline{Q}(z,x)}{\partial z} = 0 \tag{2.25a}$$

$$\frac{\partial \overline{P}(z,x)}{\partial x} + \lambda(z-1)\frac{\partial \overline{P}(z,x)}{\partial z} = 0 \tag{2.25b}$$

$$\overline{Q}(z,0) = Q_0(0)z^N \tag{2.25c}$$

$$\overline{P}(z,0) + Q_0(0)z^N = \int_0^\infty \overline{Q}(z,x)dV(x) + z\int_0^\infty \overline{P}(z,x)dB(x) \tag{2.25d}$$

$$\int_0^\infty \overline{Q}(1,x)[1-V(x)]dx + \int_0^\infty \overline{P}(1,x)[1-B(x)]dx = 1 \tag{2.25e}$$

The following analysis is an extension of that for the M/G/1//N queue without vacations by Sztrik [12] (see also Takagi [14]; this method is an alternative to Jaiswal's discrete transform technique [5]) to the present system with multiple vacations. From (2.25a), it is clear that $\overline{Q}(z,x)$ is a function of only $e^{-\lambda x}(z-1)$. From this and (2.25c), we immediately get

$$\overline{Q}(z,x) = Q_0(0)[1+e^{-\lambda x}(z-1)]^N \tag{2.26}$$

where we determine $Q_0(0)$ later. From the same reasoning for $\overline{P}(z,x)$ which satisfies (2.25b), we assume

$$\overline{P}(z,x) = \sum_{k=0}^{N-1} c_k[e^{-\lambda x}(z-1)]^k \tag{2.27}$$

for constants $\{c_k; 0 \le k \le N-1\}$. Substituting (2.26) and (2.27) into (2.25d), and comparing the coefficients of $(z-1)^k$, we get

$$c_{k-1}B^*[(k-1)\lambda] - c_k[1-B^*(k\lambda)] = Q_0(0)\binom{N}{k}[1-V^*(k\lambda)] \quad 1 \le k \le N-1 \tag{2.28}$$

This recurrence relation is satisfied by

$$c_k = \frac{Q_0(0)\zeta_k}{B^*(k\lambda)} \sum_{n=k+1}^{N} \binom{N}{n} \frac{1-V^*(n\lambda)}{\zeta_{n-1}} \qquad 0 \le k \le N-1 \tag{2.29}$$

Since (2.25e) now takes the form

$$Q_0(0)E[V] + c_0 b = 1 \tag{2.30}$$

we obtain

$$Q_0(0) = \left[E[V] + b\sum_{n=1}^{N}\binom{N}{n}\frac{1-V^*(n\lambda)}{\zeta_{n-1}}\right]^{-1} \tag{2.31}$$

which completes the solution. Note that

$$c_0 = Q_0(0)\sum_{n=1}^{N}\binom{N}{n}\frac{1-V^*(n\lambda)}{\zeta_{n-1}} = \frac{\rho'}{b} = \gamma \tag{2.32}$$

Using the above solution, we obtain the joint distribution of $\xi(t)$, $L(t)$, and the remaining service time $Y^*(t)$ or the remaining vacation time $V^*(t)$ at an arbitrary time t as $t \to \infty$. In the form of LST, we have

$$\begin{aligned}\Omega^*(z,s) &\triangleq \lim_{t\to\infty} \sum_{k=0}^{N} z^{N-k} \int_0^\infty e^{-sy}\text{Prob}[\xi(t)=0, L(t)=k, y<V^*(t)\le y+dy] \\ &= \sum_{k=0}^{N} z^{N-k} \int_0^\infty Q_k(x)dx \int_0^\infty e^{-sy}\frac{v(x+y)}{1-V(x)}dy \\ &= \int_0^\infty \overline{Q}(z,x)dx \int_0^\infty e^{-sy} v(x+y)dy \end{aligned} \tag{2.33a}$$

$$\begin{aligned}\Pi^*(z,s) &\triangleq \lim_{t\to\infty} \sum_{k=1}^{N} z^{N-k} \int_0^\infty e^{-sy}\text{Prob}[\xi(t)=1, L(t)=k, y<Y^*(t)\le y+dy] \\ &= \sum_{k=1}^{N} z^{N-k} \int_0^\infty P_k(x)dx \int_0^\infty e^{-sy}\frac{b(x+y)}{1-B(x)}dy \\ &= \int_0^\infty \overline{P}(z,x)dx \int_0^\infty e^{-sy} b(x+y)dy \end{aligned} \tag{2.33b}$$

Substituting (2.26) and (2.27) into (2.33a) and (2.33b) respectively, we get

$$\Omega^*(z,s) = Q_0(0)\sum_{k=0}^{N}\binom{N}{k}\frac{V^*(s)-V^*(k\lambda)}{k\lambda - s}(z-1)^k \tag{2.34a}$$

$$\Pi^*(z,s) = \sum_{k=0}^{N-1} c_k \frac{B^*(s)-B^*(k\lambda)}{k\lambda - s}(z-1)^k \tag{2.34b}$$

The distribution of the number of customers in the service facility is obtained by considering the marginal distribution of (2.34) as

$$\begin{aligned} Q_k &\triangleq \lim_{t\to\infty} \text{Prob}[\xi(t)=0, L(t)=k] \\ &= Q_0(0)\binom{N}{k}\int_0^\infty (1-e^{-\lambda x})^k e^{-\lambda(N-k)x}[1-V(x)]dx \quad 0\le k\le N \end{aligned} \tag{2.35a}$$

$$\begin{aligned} P_k &\triangleq \lim_{t\to\infty} \text{Prob}[\xi(t)=1, L(t)=k] \\ &= \frac{1}{(N-k)\lambda}\sum_{n=N-k}^{N-1} c_n \binom{n-1}{N-k-1}(-1)^{k-n+1}[1-B^*(n\lambda)] \quad 1\le k\le N \end{aligned} \tag{2.35b}$$

Further marginal distributions yield

$$\lim_{t\to\infty}\text{Prob}[\xi(t)=0] = \Omega^*(1,0) = Q_0(0)E[V] = 1-\rho' \tag{2.36a}$$

$$\lim_{t\to\infty}\text{Prob}[\xi(t)=1] = \Pi^*(1,0) = c_0 b = \rho' \tag{2.36b}$$

which provides a check on the present derivation. The mean number of customers in the source at an arbitrary time is given by

$$\begin{aligned} N - E[L] &= \frac{\partial}{\partial z}[\Omega^*(z,0)+\Pi^*(z,0)]|_{z=1} \\ &= Q_0(0)N\frac{1-V^*(\lambda)}{\lambda} + c_1\frac{1-B^*(\lambda)}{\lambda} = \frac{\gamma}{\lambda} \end{aligned} \tag{2.37}$$

which coincides with (2.1).

Note that the state of the system seen by an arriving customer is different from the state at an arbitrary time. Let us denote by $\overline{\xi}$, $\overline{L}$, $\overline{V}$, ane $\overline{Y}$ the state of server, the number of customers in the service facillity, the remaining vacation time, and the remaining service time, respectively, at an arrival time. Their joint distributions are obtained by noting that the arrival rate when there are k customers in the service facility is given by $(N-k)\lambda$. Thus we get

$$\begin{aligned}\overline{\Omega}^*(z,s) &\triangleq \sum_{k=0}^{N} z^{N-k}\int_0^\infty e^{-sy}\text{Prob}[\overline{\xi}=0,\overline{L}=k,y<\overline{V}\le y+dy]\\ &= \frac{\sum_{k=0}^{N} z^{N-k}(N-k)\lambda\int_0^\infty Q_k(x)dx\int_0^\infty e^{-sy}\frac{v(x+y)}{1-V(x)}dy}{\sum_{k=0}^{N}(N-k)\lambda Q_k}\\ &= \frac{\lambda z\frac{\partial\Omega^*(z,s)}{\partial z}}{Q_0(0)N[1-V^*(\lambda)]}\\ &= \frac{\lambda z}{N[1-V^*(\lambda)]}\sum_{k=1}^{N} k\binom{N}{k}\frac{V^*(s)-V^*(k\lambda)}{k\lambda-s}(z-1)^{k-1} \qquad (2.38a)\end{aligned}$$

$$\begin{aligned}\overline{\Pi}^*(z,s) &\triangleq \sum_{k=1}^{N} z^{N-k}\int_0^\infty e^{-sy}\text{Prob}[\overline{\xi}=1,\overline{L}=k,y<\overline{Y}\le y+dy]\\ &= \frac{\sum_{k=1}^{N} z^{N-k}(N-k)\lambda\int_0^\infty P_k(x)dx\int_0^\infty e^{-sy}\frac{b(x+y)}{1-B(x)}dy}{\sum_{k=1}^{N}(N-k)\lambda P_k}\\ &= \frac{\lambda z\frac{\partial\prod^*(z,s)}{\partial z}}{c_1[1-B^*(\lambda)]}\\ &= \frac{\lambda}{c_1[1-B^*(\lambda)]}\sum_{k=1}^{N-1} kc_k\frac{B^*(s)-B^*(k\lambda)}{k\lambda-s}(z-1)^{k-1} \qquad (2.38b)\end{aligned}$$

The marginal distribution of the number of customers in the service facility at the time of arrival is explicitly given by

$$\begin{aligned}\overline{Q}_k &\triangleq \text{Prob}[\overline{\xi}=0,\overline{L}=k] = \frac{(N-k)\lambda Q_k}{\sum_{j=0}^{N-1}(N-j)\lambda Q_j}\\ &= \frac{\lambda}{1-V^*(\lambda)}\binom{N-1}{k}\int_0^\infty(1-e^{-\lambda x})^k e^{-\lambda(N-k)x}[1-V(x)]dx\\ &\quad 0\le k\le N-1 \qquad (2.39a)\end{aligned}$$

$$\begin{aligned}\overline{P}_k &\triangleq \text{Prob}[\overline{\xi}=1,\overline{L}=k] = \frac{(N-k)\lambda P_k}{\sum_{j=1}^{N-1}(N-j)\lambda P_j}\\ &= \frac{1}{c_1[1-B^*(\lambda)]}\sum_{n=N-k}^{N-1} c_n\binom{n-1}{N-k-1}(-1)^{k-n+1}[1-B^*(n\lambda)]\\ &\quad 1\le k\le N-1 \qquad (2.39b)\end{aligned}$$

We note that

$$\sum_{k=0}^{N}(N-k)\lambda Q_k+\sum_{k=1}^{N}(N-k)\lambda P_k = Q_0(0)N[1-V^*(\lambda)]+c_1[1-B^*(\lambda)]=\gamma \qquad (2.40)$$

By using this arrival time distribution, the LST $W^*(s)$ of the DF for the waiting time of a customer in the first-come first-served (FCFS) system is given by

$$\begin{aligned}\gamma W^*(s) &= \sum_{k=0}^{N}[B^*(s)]^k(N-k)\lambda\int_0^\infty Q_k(x)dx\int_0^\infty e^{-sy}\frac{v(x+y)}{1-V(x)}dy \\ &+ \sum_{k=1}^{N}[B^*(s)]^{k-1}(N-k)\lambda\int_0^\infty P_k(x)dx\int_0^\infty e^{-sy}\frac{b(x+y)}{1-B(x)}dy \\ &= Q_0(0)\lambda\sum_{k=1}^{N}k\binom{N}{k}\frac{V^*(s)-V^*(k\lambda)}{k\lambda-s}[1-B^*(s)]^{k-1}[B^*(s)]^{N-k} \\ &+ \lambda\sum_{k=1}^{N-1}kc_k\frac{B^*(s)-B^*(k\lambda)}{k\lambda-s}[1-B^*(s)]^{k-1}[B^*(s)]^{N-k-1} \qquad (2.41)\end{aligned}$$

The mean waiting time $E[W]$ obtained from (2.41) coincides with (2.4) and (2.17).

We refer to Takagi [16] for a further treatment of an M/G/1//N system with multiple vacations, including the time-dependent analysis.

3. Application to a Polling Model

Our polling model is composed of M queues and a single server. M queues are identified as queue 1 through queue M in the order of server movement. Queue i has $N_i(<\infty)$ customers, and consists of the source and the service facility as an $M/G/1//N_i$ queueing system considered above, where $i=1,2,\cdots,M$. In queue i, each customer in the source arrives at its service facility at rate λ_i. The service time of a customer in queue i has a general DF $B_i(x)$ and its LST $B_i^*(s)$. The service is given to the M queues in cyclic order of indices. The service at each queue is given exhaustively, namely, once the server visits queue i, the service continues at queue i until the service facility of queue i becomes empty. [For simplicity, "the service facility of queue i" is often abbreviated as "queue i" hereafter.] When queue i becomes empty, the server switches to queue $i+1$, continues to serve queue $i+1$ until it becomes empty, and so on. We denote by $R_i(x)$ and $R_i^*(s)$ the DF and its LST, respectively, for the server's switchover time from queue i to queue $i+1$. We will show how to apply our steady-state analysis of the M/G/1//N queue given in Section 2 to obtain the mean $E[W_i]$ and the LST $W_i^*(s)$ of the DF for the waiting time (excluding the service time) of a customer at queue $i, i=1,2,\cdots,M$, in the polling model.

If we focus our attention on a particular queue, say queue i, it is clear that the period during which the server is switching or serving other queues can be viewed as the server's *vacation* time, that is, the period in which the server becomes unavailable to the queue. The length of this vacation time depends on the length of the preceding service period, which in turn depends on the length of the previous vacation time; therefore, successive vacation times are positively correlated. However, at every point in time when the server begins a vacation for queue i, queue i is always empty. The length of a vacation for queue i stochastically determines the queueing process in queue i during that vacation and the following service period. Hence, we can use our previous results for the M/G/1//N queue with vacations by supplying the distribution of each vacation time from the analysis of queue sizes in the polling model.

3.1 SOLUTION FOR AN ASYMMETRIC SYSTEM.

Let $\pi_i(k_1,\cdots,k_{i-1},0,k_{i+1},\cdots,k_M)$ be the probability that the number of customers at queue j is $k_j(j\neq i)$ when the server leaves queue i. Let $f_i(k_1,\cdots,k_M)$ be the probability that the number of customers at queue j is k_j when the server arrives at queue i. We first derive a set of linear equations to determine $\{\pi_i(k_1, \cdots, k_{i-1}, 0, k_{i+1}, \cdots, k_M); 0 \le k_j \le N_j (j \ne i)\}$ and $\{f_i(k_1, \cdots, k_M); 0 \le k_j \le N_j\}$. Note that the number of unknowns is

$$\sum_{i=1}^{M} \prod_{\substack{j=1 \\ (j \ne i)}}^{M} (N_j + 1) + \sum_{i=1}^{M} \prod_{j=1}^{M} (N_j + 1) \tag{3.1}$$

Let $\theta_i(x;k)$ be the pdf for the length of a busy period at queue i that is started with k customers, where $1 \le k \le N_i$. We can find the Laplace transform $\Theta_i^*(s;k)$ of $\theta_i(x;k)$ from (2.7) by substituting λ_i, $B_i^*(s)$, and N_i for $\lambda, B^*(s)$, and N, respectively. Also, let $a_j(x;q,k)$ be the probability that the number of customers in the service facility of queue j increases from q to k (owing to new arrivals) during a time interval x in which queue j is not being served. This is given by

$$\begin{aligned} a_j(x;q,k) &= \binom{N_j - q}{k - q}(1 - e^{-\lambda_j x})^{k-q} e^{-\lambda_j x(N_j - k)} \\ &= \binom{N_j - q}{k - q} \sum_{n=0}^{k-q} \binom{k-q}{n} (-1)^n e^{-\lambda_j (N_j - k + n)x} \qquad q \le k \le N_j \end{aligned} \tag{3.2}$$

Consider the events that occur between the instant when the server visits queue i and the instant when the server leaves queue i. Suppose that there are q_i customers in queue j when the server visits queue i, where $j = 1, \cdots, M$. If $q_i = 0$, the server immediately leaves queue i. Otherwise, the length of a service period at queue i has a pdf $\theta_i(x;q_i)$. If $k_j - q_j$ customers arrive at queue j during the service period of queue i, there are k_j customers at queue j when the server leaves queue i, where $j \ne i$. Of course, queue i is empty at that point. Therefore, we get a relation

$$\begin{aligned} \pi_i(k_1, \cdots, k_{i-1}, 0, k_{i+1}, \cdots, k_M) &= f_i(k_1, \cdots, k_{i-1}, 0, k_{i+1}, \cdots, k_M) \\ &+ \sum_{q_1=0}^{k_1} \cdots \sum_{q_{i-1}=0}^{k_{i-1}} \sum_{q_i=1}^{N_i} \sum_{q_{i+1}=0}^{k_{i+1}} \cdots \sum_{q_M=0}^{k_M} f_i(q_1, \cdots, q_{i-1}, q_i, q_{i+1}, \cdots, q_M) \\ & \int_0^\infty [\prod_{\substack{j=1 \\ (j \ne i)}}^{M} a_j(x;q_j,k_j)]\theta_i(x;q_i)dx \end{aligned} \tag{3.3}$$

The integration part in (3.3) can be expressed in terms of known $\Theta_i^*(s;q_i)$ by substituting (3.2) as

$$\begin{aligned} \int_0^\infty [\prod_{\substack{j=1 \\ (j \ne i)}}^{M} a_j(x;q_j,k_j)]\theta_i(x;q_i)dx &= [\prod_{\substack{j=1 \\ (j \ne i)}}^{M} \binom{N_j - q_j}{k_j - q_j}] \\ &\times \sum_{n_1=0}^{k_1 - q_1} \cdots \sum_{n_{i-1}=0}^{k_{i-1}-q_{i-1}} \sum_{n_{i+1}=0}^{k_{i+1}-q_{i+1}} \cdots \sum_{n_M=0}^{k_M - q_M} [\prod_{\substack{j=1 \\ (j \ne i)}}^{M} \binom{k_j - q_j}{n_j} (-1)^{n_j}] \\ &\times \Theta_i^*[\sum_{\substack{j=1 \\ (j \ne i)}}^{M} (N_j - k_j + n_j)\lambda_j; q_i] \end{aligned} \tag{3.4}$$

Note that we do not need numerical integration. By similar arguments for the events during the switchover from queue i to queue $i+1$, we get

$$\begin{aligned} &f_{i+1}(k_1,\cdots,k_{i-1},k_i,k_{i+1},\cdots,k_M) \\ &= \sum_{q_1=0}^{k_1}\cdots\sum_{q_{i-0}=0}^{k_{i-1}}\sum_{q_{i+1}=0}^{k_{i+1}}\cdots\sum_{q_M=0}^{k_M}\pi_i(q_1\cdots,q_{i-1},0,q_{i+1},\cdots,q_M) \\ &\int_0^\infty[\prod_{\substack{j=1\\(j\neq i)}}^M a_j(x;q_j,k_j)]a_i(x;0,k_i)dR_i(x) \end{aligned} \tag{3.5}$$

where

$$\begin{aligned} &\int_0^\infty[\prod_{\substack{j=1\\(j\neq i)}}^M a_j(x;q_j,k_j)]a_i(x;0,k_i)dR_i(x) = [\prod_{\substack{j=1\\(j\neq i)}}^M \binom{N_j-q_j}{k_j-q_j}]\binom{N_i}{k_i} \\ &\times\sum_{n_1=0}^{k_1-q_1}\cdots\sum_{n_{i-1}=0}^{k_{i-1}-q_{i-1}}\sum_{n_i=0}^{k_i}\sum_{n_{i+1}=0}^{k_{i+1}-q_{i+1}}\cdots\sum_{n_M=0}^{k_M-q_M}[\prod_{\substack{j=1\\(j\neq i)}}^M\binom{k_j-q_j}{n_j}(-1)^{n_j}]\binom{k_i}{n_i}(-1)^{n_i} \\ &\times R_i^*[\sum_{j=1}^M(N_j-k_j+n_j)\lambda_j] \end{aligned} \tag{3.6}$$

The normalization conditions are given by

$$\sum_{k_1=0}^{N_1}\cdots\sum_{k_{i-1}=0}^{N_{i-1}}\sum_{k_{i+1}=0}^{N_{i+1}}\cdots\sum_{k_M=0}^{N_M}\pi_i(k_1,\cdots,k_{i-1},0,k_{i+1},\cdots,k_M)=1 \tag{3.7a}$$

$$\sum_{k_1=0}^{N_1}\cdots\sum_{k_M=0}^{N_M}f_i(k_1,\cdots,k_M)=1 \quad i=1,\cdots,M \tag{3.7b}$$

Thus we can compute $\{\pi_i(k_1,\cdots,k_{i-1},0,k_{i+1},\cdots,k_M); 0\le k_j\le N_j (j\neq i)\}$ and $\{f_i(k_1,\cdots,k_M); 0\le k_j\le N_j\}$ by solving equations (3.3), (3.5) and (3.7a,b).

Let $V_i^*(s;k_1,\cdots,k_{i-1},0,k_{i+1},\cdots,k_M)$ be the LST of the DF for the length of a "vacation" for queue i which is started with the state that there are k_j customers in queue j, where $j\neq i$. The vacation for queue i, also called the *intervisit time* for queue i, is the time interval between the instant when the server leaves queue i and the instant when the server next arrives at queue i. We will show how to obtain $V_i^*(s;k_1,\cdots,k_{i-1},0,k_{i+1},\cdots,k_M)$ shortly. Once it is obtained, we can calculate the unconditional LST $V_i^*(s)$ of the DF for the vacation length for queue i by

$$\begin{aligned} V_i^*(s) = &\sum_{k_1=0}^{N_1}\cdots\sum_{k_{i-1}=0}^{N_{i-1}}\sum_{k_{i+1}=0}^{N_{i+1}}\cdots\sum_{k_M=0}^{N_M}\pi_i(k_1,\cdots,k_{i-1},0,k_{i+1},\cdots,k_M) \\ \times\ &V_i^*(s;k_1,\cdots,k_{i-1},0,k_{i+1},\cdots,k_M) \end{aligned} \tag{3.8}$$

Using $V_i^*(s)$ and appropriate parameters of queue i in our previous analysis of the M/G/1//N queue, we can obtain from (2.41) the LST $W_i^*(s)$ of the DF for the customer waiting time, and from (2.4) and (2.17) its mean $E[W_i]$. Other performance measures for each queue can be obtained similarly.

For the sake of convenience, we give an expression for the conditional LST of the DF for the vacation length for queue 1. Considering the events after the server leaves queue 1 and until it next arrives at queue 1, we get

$$V_1^*(s;0,k_2^{(0)},\cdots,k_M^{(0)}) = \sum_{k_2^{(1)}=k_2^{(0)}}^{N_2} \sum_{k_3^{(1)}=k_3^{(0)}}^{N_3} \sum_{k_3^{(2)}=k_3^{(1)}}^{N_3} \cdots \sum_{k_M^{(1)}=k_M^{(0)}}^{N_M} \sum_{k_M^{(2)}=k_M^{(1)}}^{N_M} \cdots \sum_{k_M^{(M-1)}=k_M^{(M-2)}}^{N_M}$$
$$\int_0^\infty e^{-sx_1} \prod_{j=2}^{M} a_j(x_1;k_j^{(0)},k_j^{(1)})dR_1(x_1)$$
$$\times \prod_{i=2}^{M} \int_0^\infty e^{-sx_i}[\prod_{j=i+1}^{M} a_j(x_i;k_j^{(i-1)},k_j^{(i)})]\omega_i(x_i;k_i^{(i-1)})dx_i \tag{3.9}$$

where $k_j^{(i)}$ represents the number of customers in queue j when the server arrives at queue $i+1$, $1 \le i \le M-1$, and

$$\begin{aligned}\omega_i(x;k) &\triangleq \int_0^\infty \theta_i(x-y;k)dR_i(y) && k \ge 1 \\ &\triangleq dR_i(x)/dx && k=0\end{aligned} \tag{3.10}$$

is the pdf for the convolution of the service period started with k customers and the following switchover time for queue i. Integration parts in (3.9) can be expressed in terms of LSTs $\{\Theta_i^*(s;q)\}$ and $\{R_i^*(s)\}$ in a way similar to (3.4) and (3.6).

A system-wide performance measure in a polling model is the *polling cycle time*. The cycle time C_i for queue i is defined as the time between the server's visits to queue i in successive cycles. Considering queue 1, for example, let $C_1^*(s;k_1,\cdots,k_M)$ be the LST of the DF for cycle time C_1 which is started with the state that there are k_j customers in queue j, where $j=1,2,\cdots,M$. The unconditional LST $C_1^*(s)$ is given by

$$C_1^*(s) = \sum_{k_1=0}^{N_1} \cdots \sum_{k_M=0}^{N_M} f_1(k_1,\cdots,k_M)C_1^*(s;k_1,\cdots,k_M) \tag{3.11}$$

where

$$C_1^*(s;k_1^{(0)},\cdots,k_M^{(0)}) = \sum_{k_2^{(1)}=k_2^{(0)}}^{N_2} \sum_{k_3^{(1)}=k_3^{(0)}}^{N_3} \sum_{k_3^{(2)}=k_3^{(1)}}^{N_3} \cdots \sum_{k_M^{(1)}=k_M^{(0)}}^{N_M} \sum_{k_M^{(2)}=k_M^{(1)}}^{N_M} \cdots \sum_{k_M^{(M-1)}=k_M^{(M-2)}}^{N_M}$$
$$\prod_{i=1}^{M} \int_0^\infty e^{-sx_i}[\prod_{j=i+1}^{M} a_j(x_i;k_j^{(i-1)},k_j^{(i)})]\omega_i(x_i;k_i^{(i-1)})dx_i \tag{3.12}$$

3.2 NUMERICAL RESULTS FOR SYMMETRIC SYSTEMS.

For a symmetric system in which all parameters are statistically identical for all queues, we can focus on queue 1 without loss of generality. In this case, we omit subscripts from all system parameters.

From (3.3), $\pi_1(0,k_2,\cdots,k_M)$ can be expressed in terms of

$$f_1(q_1,\cdots,q_{M-1},q_M) = f_2(q_M,q_1,\cdots,q_{M-1}) \tag{3.13}$$

which results from symmetry. Furthermore, from (3.5), $f_2(q_M, q_1, \cdots, q_{M-1})$ can be expressed in terms of $\pi_1(0, r_1, \cdots, r_{M-1})$. Therefore, we obtain the following set of linear equations for $(N+1)^{M-1}$ unknowns $\{\pi_1(0, k_2, \cdots, k_M); 0 \le k_j \le N (2 \le j \le M)\}$:

$$\pi_1(0, k_2, \cdots, k_M) = \sum_{r_1=0}^{N} \sum_{r_2=0}^{N} \cdots \sum_{r_{M-1}=0}^{N} \pi_1(0, r_1, r_2, \cdots, r_{M-1}) \times \mathbf{R}(r_1, r_2, \cdots, r_{M-1}; k_2, \cdots, k_{M-1}, k_M) \tag{3.14a}$$

$$\sum_{k_2=0}^{N} \cdots \sum_{k_M=0}^{N} \pi_1(0, k_2, \cdots, k_M) = 1 \tag{3.14b}$$

where

$$\mathbf{R}(0, r_2, \cdots, r_{M-1}; k_2, \cdots, k_{M-1}, k_M) \triangleq \mathbf{Q}(0, r_2, \cdots, r_{M-1}; k_M, 0, k_2, \cdots, k_{M-1}) + \sum_{q_1=1}^{N} \sum_{q_2=r_2}^{k_2} \cdots \sum_{q_{M-1}=r_{M-1}}^{k_{M-1}} \sum_{q_M=0}^{k_M} \mathbf{Q}(0, r_2, \cdots, r_{M-1}; q_M, q_1, \cdots, q_{M-1}) \mathbf{P}(q_1, \cdots, q_M; k_2, \cdots, k_M)$$
$$0 \le r_2 \le k_2 \le N; \cdots; 0 \le r_{M-1} \le k_{M-1} \le N; 0 \le k_M \le N \tag{3.15a}$$

$$\mathbf{R}(r_1, r_2, \cdots, r_{M-1}; k_2, \cdots, k_{M-1}, k_M) \triangleq \sum_{q_1=r_1}^{N} \sum_{q_2=r_2}^{k_2} \cdots \sum_{q_{M-1}=r_{M-1}}^{k_{M-1}} \sum_{q_M=0}^{k_M} \mathbf{Q}(r_1, r_2 \cdots, r_{M-1}; q_M, q_1, \cdots, q_{M-1}) \mathbf{P}(q_1, \cdots, q_M; k_2, \cdots, k_M)$$
$$1 \le r_1 \le N; 0 \le r_2 \le k_2 \le N; \cdots; 0 \le r_{M-1} \le k_{M-1} \le N; 0 \le k_M \le N \tag{3.15b}$$

and

$$\mathbf{P}(q_1, \cdots, q_M; k_2, \cdots, k_M) \triangleq \prod_{j=2}^{M} \binom{N-q_j}{k_j-q_j} \times \sum_{n_2=0}^{k_2-q_2} \cdots \sum_{n_M=0}^{k_M-q_M} [\prod_{j=2}^{M} \binom{k_j-q_j}{n_j} (-1)^{n_j}] \Theta^*_{q_1}[\sum_{j=2}^{M} (N-k_j+n_j)\lambda] \tag{3.16a}$$

$$\mathbf{Q}(r_1, \cdots, r_{M-1}; q_M, q_1, \cdots, q_{M-1}) \triangleq [\prod_{j=1}^{M-1} \binom{N-r_j}{q_j-r_j}] \binom{N}{q_M} \times \sum_{n_M=0}^{q_M} \sum_{n_1=0}^{q_1-r_1} \cdots \sum_{n_{M-1}=0}^{q_{M-1}-r_{M-1}} [\prod_{j=1}^{M-1} \binom{q_j-r_j}{n_j} (-1)^{n_j}] \binom{q_M}{n_M} (-1)^{n_M} R^*[\sum_{j=1}^{M} (N-q_j+n_j)\lambda] \tag{3.16b}$$

where $\Theta^*_k(s)$ is defined in (2.7). Using $\{\pi_1(0, k_2, \cdots, k_M); 0 \le k_j \le N (2 \le j \le M)\}$, we can determine the unconditional LST of the DF for the vacation length for queue 1 by (3.8), namely,

$$V_1^*(s) = \sum_{k_2=0}^{N} \cdots \sum_{k_M=0}^{N} \pi_1(0, k_2, \cdots, k_M) V_1^*(s; 0, k_2, \cdots, k_M) \tag{3.17}$$

where, from (3.9), we have

$$V_1^*(s; 0, k_2^{(0)}, \cdots, k_M^{(0)}) = \sum_{k_2^{(1)}=k_2^{(0)}}^{N} \cdots \sum_{k_M^{(1)}=k_M^{(0)}}^{N} V_1^{(1)}(s; k_2^{(1)}, \cdots, k_M^{(1)}) \prod_{j=2}^{M} \binom{N-k_j^{(0)}}{k_j^{(1)}-k_j^{(0)}}$$

$$\times \sum_{n_2=0}^{k_2^{(1)}-k_2^{(0)}} \cdots \sum_{n_M=0}^{k_M^{(1)}-k_M^{(0)}} [\prod_{j=2}^{M} \binom{k_j^{(1)}-k_j^{(0)}}{n_j}(-1)^{n_j}]R^*[s+\sum_{j=2}^{M}(N-k_j^{(1)}+n_j)\lambda] \quad (3.18a)$$

$$V_1^{(i)}(s;k_{i+1}^{(i)},\cdots,k_M^{(i)}) = \sum_{k_{i+2}^{(i+1)}=k_{i+2}^{(i)}}^{N} \cdots \sum_{k_M^{(i+1)}=k_M^{(i)}}^{N} V_1^{(i+1)}(s;k_{i+2}^{(i+1)},\cdots,k_M^{(i+1)})$$

$$\prod_{j=i+2}^{M}\binom{N-k_j^{(i)}}{k_j^{(i+1)}-k^{(i)}{}_j} \sum_{n_{i+2}=0}^{k_j^{(i+1)}-k_j^{(i)}} \cdots \sum_{n_M=0}^{k_M^{(i+1)}-k_M^{(i)}} [\prod_{j=i+2}^{M}\binom{k_j^{(i+1)}-k_j^{(i)}}{n_j}(-1)^{n_j}]$$

$$\times \Omega^*[s+\sum_{j=i+2}^{M}(N-k_j^{(i+1)}+n_j)\lambda; k_{i+1}^{(i)}]$$

$$i=1,2,\cdots,M-2 \quad (3.18b)$$

$$V_1^{(M-1)}(s;k_M^{(M-1)}) = \Omega^*(s;k_M^{(M-1)}) \quad (3.18c)$$

and

$$\Omega^*(s;k) \triangleq \begin{cases} \Theta_k^*(s)R^*(s) & k \geq 1 \\ R^*(s) & k=0 \end{cases} \quad (3.19)$$

Note that $V_1^{(i)}(s;k_{i+1}^{(i)},\cdots,k_M^{(i)})$ is the LST of the DF for the length of an interval from the time at which the server arrives at queue $i+1$ and there are $k_{i+1}^{(i)},\cdots,k_M^{(i)}$ customers in queues $i+1,\cdots,M$, respectively, to the time at which the server comes to queue 1. Starting with $V_1^{(M-1)}(s;k_M^{(M-1)})$ given by (3.18c), we can determine $V_1^{(i)}(s;k_{i+1}^{(i)},\cdots,k_M^{(i)})$ in the decreasing order of i by (3.18b) until $i=1$. Using $V_1^{(1)}(s;k_2^{(1)},\cdots,k_M^{(1)})$ thus obtained in (3.18a), we get $V_1^*(s;0,k_2^{(0)},\cdots,k_M^{(0)})$, which is substituted into (3.17) to determine $V_1^*(s)$. A similar simplification is possible for the cycle time distribution.

Using the above formulation, we have computed the mean response time $E[T]$ for symmetric polling systems with $M=4$ queues, constant service time $b=1.0$, constant switchover time $r=0.1$, and population size (per queue) $N=1,3$, and 5. For a symmetric system with constant service time b, constant switchover time r, and $N=1$ (which is equivalent to a single-buffer model), the exact expression for $E[T]$ is available [9,11](also in [13,15]):

$$E[T] = Mb - \frac{1}{\lambda} + \frac{Mr[1+\sum_{m=1}^{M}\binom{M}{m}\prod_{j=0}^{m-1}\{e^{\lambda(Mr+jb)}-1\}]}{\sum_{m=0}^{M-1}\binom{M-1}{m}\prod_{j=0}^{m}\{e^{\lambda(R+jb)}-1\}} \quad (3.20)$$

Our numerical results are given in Table 1 against the offered load $\rho = MN\lambda b$. For $N=1$, the numerical values based on the above solution and those from (3.20) agree down to digits shown in the table. For comparison, we attach the mean response times in the corresponding polling system with an infinite population ($N \to \infty$ and $\lambda \to 0$ with ρ kept fixed), which is given by (see [13, 15])

$$E[T] = \frac{\rho b + r(M-\rho)}{2(1-\rho)} + b \quad (3.21)$$

For small values of ρ in systems with $N=3$ and 5 (the arrival rate per customer is very small), we have encountered low precision in the numerical calculation. We have confirmed that the numerical values in the table fall within the 90 percent confidence interval of RESQ [10] simulation results in almost all cases. In the table, it is clear that $E[T] \approx b+Mr/2$ as $\rho \to$

0. For $N = 1$, $E[T]$ grows almost linear in ρ within the range shown in the table. For ρ fixed at a relatively small value (for example, $\rho = 0.2$), the variation in $E[T]$ with increasing N in not monotonous. This anomaly seems to result from the trade-off between the effects of decreasing the arrival rate per customer and increasing the total population size.

Table 1. Mean response times in polling systems.

ρ	$N = 1$	$N = 3$	$N = 5$	$N = \infty$
.05	1.22	-	-	1.234
.10	1.25	1.24	-	1.272
.15	1.284	1.29	-	1.315
.20	1.314	1.34	1.26	1.363
.25	1.345	1.38	1.35	1.417
.30	1.377	1.43	1.42	1.479
.35	1.410	1.49	1.49	1.550
.40	1.444	1.55	1.57	1.633
.45	1.479	1.61	1.64	1.732
.50	1.515	1.68	1.73	1.850
.55	1.552	1.76	1.83	1.994
.60	1.589	1.84	1.94	2.175
.65	1.627	1.94	2.06	2.407
.70	1.665	2.04	2.20	2.717
.75	1.704	2.15	2.36	3.150
.80	1.744	2.27	2.54	3.800
.85	1.784	2.40	2.75	4.883
.90	1.824	2.54	2.98	7.050
.95	1.864	2.70	3.24	13.55
1.00	1.904	2.86	3.52	∞

4. Concluding Remarks

In this paper, we have studied a single $M/G/1//N$ queue with the server's vacations. The results have been applied to a polling model with a finite population of customers, which can be used for the performance evaluation of token-ring networks connecting several computers, each of which supports a finite number of interactive users. Problems in the numerical analysis of a polling model include the computational time that grows exponentially with the number of queues, and the low precision resulting from very small values of the arrival rate per customer when the system size is large. Therefore, we have shown numerical results for systems of small size. One way to overcome these difficulties may be the development of approximation techniques, which is a future research subject. Our exact solution will provide a useful benchmark tool for them.

Although the exhaustive service has been assumed throughout the paper, other policies such as *gated* and *limited* are often considered in vacation and polling models with an infinite population of customers [1, 2, 13, 15]. In the *gated* service, only customers that are present in the service facility when the server visits queue i are served continuously; those customers that arrive during this service period are set aside, and served in the next polling cycle. In the *limited* service, only at most one customer is served from each queue. It will be challenging to extend the approach of this paper to those systems with gated or limited service policies.

ACKNOWLEDGMENT

The author is thankful to Mr. Jun Nakano of IBM Tokyo Research Laboratory for his programming support.

REFERENCES

1. Doshi, B. T.: Queueing systems with vacations – a survey. *Queueing Systems 1*, 1 (June 1986), 29–66.
2. Doshi, B. T.: Single server queues with vacations. *Stochastic Analysis of Computer and Communication Systems*, 217–265, H. Takagi (editor), Elsevier Science Publishers B. V. (North-Holland), Amsterdam 1990.
3. Doshi, B.: Generalizations of the stochastic decomposition results for single server queues with vacations. *Stochastic Models 6*, 2 (1990), 307–333.
4. Ibe, O. C., and Trivedi, K. S.: Stochastic Petri net models of polling systems. *IEEE Journal on Selected Areas in Communications 8*, 9 (December 1990), 1649–1657.
5. Jaiswal, N. K.: *Priority Queues.* Academic Press, New York, 1968.
6. Keilson, J., and Kooharian, A.: On time dependent queuing processes. *The Annals of Mathematical Statistics 31*, 1 (March, 1960), 104–112.
7. Kleinrock, L.: *Queueing Systems, Volume 2: Computer Applications.* John Wiley and Sons, New York, 1976.
8. Levy, Y., and Yechiali, U.: Utilization of idle time in an $M/G/1$ queueing system. *Management Science 22*, 2 (October 1975), 202–211.
9. Mack, C., Murphy, T., and Webb, N.L.: The efficiency of N machines uni-directionally patrolled by one operative when walking time and repair times are constants. *Journal of the Royal Statistical Society B 19*, 1 (1957), 166–172.
10. Sauer, C. H., McNair, E. A., and Kurose, J. F.: Queueing network simulations of computer communication. *IEEE Journal on Selected Areas in Communications SAC-2*, 1 (January 1984), 203–220.
11. Scholl, M., and Potier, D.: Finite and infinite source models for communication systems under polling. IRIA Rapport de Recherche, No. 308, Institut de Recherche en Informatique et en Automatique, Le Chesnay, France, 1978.
12. Sztrik, J.: On machine interference. *Publicationes Mathematicae 30*, 1-2 (1983) 165–175, Institutum Mathematicum Universitatis Debreceniensis, Debrecen, Hungary.
13. Takagi, H.: Queuing analysis of polling models. *ACM Computing Surveys 20*, 1 (March, 1988), 5–28.
14. Takagi, H.: Queueing analysis of vacation models, Part IV: $M/G/1//N$. TRL Research Report TR87-0043, IBM Japan, Tokyo, 1988.
15. Takagi, H.: Queueing analysis of polling models: An update. *Stochastic Analysis of Computer and Communication Systems*, 267–318, H. Takagi (editor), Elsevier Science Publishers B. V. (North-Holland), Amsterdam, 1990.
16. Takagi, H.: Analysis of an $M/G/1//N$ queue with server's multiple vacations and exhaustive service, and its application to a polling model. Research Report RT 0033, IBM Tokyo Research Laboratory, 1990.

Hideaki Takagi
IBM Research,
Tokyo Research Laboratory
5-19 Sanbancho, Chiyoda-ku
Tokyo 102, Japan