Relating Polling Models with Zero and Nonzero Switchover Times

Mandyam M. Srinivasan
Management Science Program
College of Business Administration
The University of Tennessee
Knoxville, TN 37996-0562

Shun-Chen Niu[1]
School of Management
The University of Texas at Dallas
P. O. Box 830688
Richardson, TX 75083-0688

Robert B. Cooper[1]
Department of Computer Science and Engineering
Florida Atlantic University
P. O. Box 3091
Boca Raton, FL 33431-0991

November 9, 1993
Revised: March 29, 1994; August 12, 1994

To appear in *Queueing Systems.*

[1]Research supported in part by the National Science Foundation under grant DDM-9001751.

Abstract

We consider a system of N queues served by a single server in cyclic order. Each queue has its own distinct Poisson arrival stream and its own distinct general service-time distribution (asymmetric queues); and each queue has its own distinct distribution of switchover time (the time required for the server to travel from that queue to the next). We consider two versions of this classical polling model: In the first, which we refer to as the zero-switchover-times model, it is assumed that all switchover times are zero and the server stops traveling whenever the system becomes empty. In the second, which we refer to as the nonzero-switchover-times model, it is assumed that the sum of all switchover times in a cycle is nonzero and the server does not stop traveling when the system is empty. After providing a new analysis for the zero-switchover-times model, we obtain, for a host of service disciplines, transform results that completely characterize the relationship between the waiting times in these two, operationally-different, polling models. These results can be used to derive simple relations that express (all) waiting-time moments in the nonzero-switchover-times model in terms of those in the zero-switchover-times model. Our results, therefore, generalize corresponding results for the expected waiting times obtained recently by Fuhrmann [8] and Cooper, Niu, and Srinivasan [4].

Key words. Polling models, cyclic queues, waiting times, decomposition, switchover times, vacation models.

1 Introduction

A *polling model* is a system of multiple queues attended by a single server that travels from queue to queue in some prescribed manner. Polling models have many important applications (computer networks, telephone switching systems, materials handling, etc.) and, in general, they are extremely complicated. Consequently, there is a huge literature dealing with various versions of these models and their numerical analysis (see, for example, Takagi [16, 17]).

In this paper, we assume that the queues are attended by the server in cyclic order, and we consider two versions of the polling model: the *zero-switchover-times* model and the *nonzero-switchover-times* model. In the former model, it is assumed that the switchover times (i.e., the times required for the server to travel between queues) are equal to zero; the server stops switching whenever the system becomes empty, and then instantaneously switches to the queue where the next customer arrives, to begin service on this customer. In the latter model, the switchover times are assumed to be arbitrarily-distributed random variables, and the server continues traveling even when the system is empty. In each case, the queue discipline at each queue is First-In-First-Out (FIFO).

In most of the existing literature on polling models, the zero-switchover-times model and the nonzero-switchover-times model are considered separately. This tradition is closely related to the confusing fact, first noted by Eisenberg [5] (p. 441), that since the sum of all switchover times in a cycle (i.e., the elapsed time between two successive visits to a queue by the server) must remain nonzero in the latter model, corresponding results for the former model cannot be duplicated. (Also see Eisenberg [6], where he revisits this issue and analyzes several models in which the server stops when the system is empty.)

Our purpose in this paper is to uncover, for a host of service disciplines, simple transform results that characterize the relationship between the waiting times in these two, *operationally different*, models. Understanding this relationship is important, because it is clearly desirable to be able to translate back and forth between these two models. In relating these models, our primary objectives are: (1) to establish continuity, in distribution, of the waiting times, and (2) to do so in a way that reveals the underlying "link" between these models. With respect to these objectives, it must be emphasized that not *all* functionals of these models are continuous. For example, it is well known that the average cycle times in these models are not continuous. That is, if one lets the sums of the switchover times in a sequence of nonzero-switchover-times models converge to zero, then the corresponding average cycle times will converge to zero, since each time the system becomes empty, the server will execute, in the limit, an infinite number of cycles in any ensuing positive finite time interval; and this limit clearly is not consistent with the zero-switchover-times model defined above. For the waiting times, continuity has either been *assumed* or investigated only numerically in most of the previous work; see, for example, Garner [10], Levy and Kleinrock [14], Takagi [16] (p. 294), and Tayur and Sarkar [18]. Our transform results, which are established via two separate, but parallel, analyses of these two models, can be used to provide a rigorous proof of the continuity (in distribution) of the waiting times.

Our work is a continuation of two recent papers by Fuhrmann [8] and by Cooper, Niu, and Srinivasan [4]. In Fuhrmann [8], it is shown that, for constant switchover times, the population of customers present (represented by a vector whose components are the numbers of waiting customers at each queue) in the nonzero-switchover-times model enjoys a stochastic decomposition at certain points in time. This leads to a simple decomposition relationship between the expected waiting times in the zero- and nonzero-switchover-times models. Using a different argument, Cooper, Niu, and Srinivasan [4] obtain the same decomposition for constant switchover times, and further show that, with an appropriate modification of the service-time variances, the decomposition holds, in fact, even when the switchover times are arbitrarily-distributed random variables. These analyses, however, do not extend to the higher moments of waiting times. Our transform results can be used to derive, in a straightforward manner, higher-moment generalizations of these expected-waiting-times results.

Our transform results will be derived via a new analysis of the zero-switchover-times model. The zero-switchover-times model was originally studied in Cooper and Murray [3] and Cooper [1] for a system with either *exhaustive service* at all queues (that is, the server switches from a queue only when it is empty), or *gated service* at all queues (the server closes a "gate" behind the waiting customers when it arrives at a queue and switches to the next queue upon completion of service of all customers in front of the gate). In particular, the FIFO waiting-time distribution for the exhaustive-service system was derived in Cooper [1] (eq. (27), p. 407) based on the idea that, from the viewpoint of a particular queue, the time spent serving the other queues can be interpreted as a "vacation." This formulation led to a *decomposition* of the waiting times into a sum of two independent variables: the waiting time in the standard $M/G/1$ queue without vacations, and an additional time that depends only on the duration of the vacation. This decomposition was subsequently generalized by Fuhrmann and Cooper [9] (Propositions 3 and 4, pp. 1125 and 1126) to cover disciplines other than exhaustive service. Our analysis in this paper will also be based on these decompositions: We will study the durations of vacations using the concept of "descendant set," as explicated recently by Konheim, Levy, and Srinivasan [12] for the analysis of the nonzero-switchover-times model. (See also Fuhrmann [7].) For expositional ease, we will first present our results assuming that the service discipline (at every queue) is exhaustive. Subsequently, we will show that our method easily extends to polling models with gated service, and to many other cases.

In Section 2, we summarize notation and preliminaries. In Section 3, we focus on the zero-switchover-times model with exhaustive service at all queues. In Section 4, we establish a transform result that relates the waiting times in the zero- and nonzero-switchover-time models with the exhaustive-service discipline; in particular, we show that in the nonzero-switchover-times model, if the switchover times are all constants, then the waiting-time distribution depends only on the sum of the switchover times, and not on their individual values. In Section 5, we extend our analysis to the gated-service case, and to other disciplines as well. In Section 6, we develop a new computational technique for calculating all waiting-time moments in the zero-switchover-times model; and from this we derive several explicit expressions that relate these moments to those of the nonzero-switchover-times model.

2 Notation and Preliminaries

A single server serves in cyclic order a sequence of N infinite-capacity queues. The arrival process to queue i is Poisson at rate λ_i, the time required to serve a customer at queue i is denoted by B_i, and a switchover time (possibly of zero duration) at queue i (i.e., the time required for the server to travel from queue i to queue $i+1$) is denoted by R_i. The arrival processes, service times, and switchover times are all mutually independent.

Unless otherwise stated, the Laplace-Stieltjes transform (LST) of a nonnegative random variable A, defined as $E[\exp(-sA)]$, is denoted by $\tilde{A}(s)$; and when A is discrete, its probability-generating function (PGF), defined as $E[z^A]$, is denoted by $A(z)$. Multivariate LSTs and PGFs are defined similarly. It is assumed that the index used for a summation over the queues is (a) reset to 1 if it increases to $N+1$ and (b) reset to N if it decreases to 0. We adopt the convention that an empty product equals 1 and an empty sum equals 0.

The total arrival rate to the system is denoted by $\lambda \equiv \sum_{i=1}^{N} \lambda_i$, and the sum of the switchover times by $R \equiv \sum_{i=1}^{N} R_i$. The traffic intensity at queue i is denoted by $\rho_i \equiv \lambda_i E[B_i]$, and the server utilization by $\rho \equiv \sum_{i=1}^{N} \rho_i$. For the polling system to be stable, ρ must be less than 1, and this is assumed to be the case. A busy period generated by a single customer at queue i is denoted by Θ_i. The waiting times at queue i in the zero- and nonzero-switchover-times models are denoted by W_i^0 and W_i, respectively. The waiting time in the standard $M/G/1$ queue with arrival rate λ_i and service times distributed as B_i is denoted by W_i^*.

We conclude this section with the statements of two elementary results that will be used repeatedly later. Consider a random interval of duration S. For each $i = 1, \cdots, n$, let A_i denote the number of arrivals during S of a Poisson process at rate λ_i. Suppose these n Poisson processes are independent of each other and of S. Then, the (multivariate) PGF of $(A_1, \cdots, A_n)$ is given by

$$E\left[\prod_{i=1}^{n} z_i^{A_i}\right] = \tilde{S}\left(\sum_{i=1}^{n} \lambda_i (1 - z_i)\right). \tag{1}$$

Next, consider a random vector $(Q_1, \cdots, Q_n)$ with PGF $f(z_1, \cdots, z_n)$. For each $i = 1, \cdots, n$, let $\{K_{ij},\, j = 1, 2, \cdots\}$ be a corresponding sequence of i.i.d. random variables, each with PGF $g_i(z)$; and let $\hat{Q}_i = \sum_{j=1}^{Q_i} K_{ij}$ (that is, replace the "jth count" in Q_i by the cardinality K_{ij} of a corresponding "batch"). Suppose the n sequences $\{K_{ij},\, j = 1, 2, \cdots\}$

are independent of each other and of $(Q_1, \cdots, Q_n)$. Then, the PGF of $(\hat{Q}_1, \cdots, \hat{Q}_n)$ is given by

$$E\left[\prod_{i=1}^{n} z_i^{\hat{Q}_i}\right] = f\left(g_1(z_1), \cdots, g_n(z_n)\right). \tag{2}$$

These two results are multivariate extensions of, e.g., equation (6.10), p. 199, and Exercise 4, p. 30, in Cooper [2].

3 The Zero-Switchover-Times Model

The starting point of the analysis in [3] and [1] is a Markov chain embedded at a set of *switch points*, defined as the time epochs at which the server *finishes* service (if any) at a queue and is ready to switch from that queue to the next. Our objective in this section is to develop (for this zero-switchover-times model with exhaustive service) a parallel analysis that is based on an embedding at *polling epochs*, defined as the time epochs at which the server *arrives* at a queue. We are motivated to work with this different set of embedding epochs because most existing analyses for the nonzero-switchover-times model are based on polling epochs. As will be seen in Section 4, a unified embedding scheme greatly simplifies comparison of results for these two models.

Define a time interval during which the server is away from queue i as a *vacation* from queue i, and consider the time epoch at which the server returns to that queue after a randomly-selected (or "tagged") vacation. We will call such an epoch an *i-polling epoch.* Let $G_i(z_1, \cdots, z_N)$ denote the PGF of the numbers of customers waiting for service at the various queues at this polling epoch, and note that, in particular, the PGF of the number of customers waiting at queue i is given by $G_i(1, \cdots, 1, z_i, 1, \cdots, 1)$. Then, it is shown in [1] (eq. (18), p. 404) that the LST of the waiting time of a randomly-selected departing customer from queue i has the following decomposition:

$$\tilde{W}_i^0(s) = \frac{1 - G_i\left(1, \cdots, 1, 1 - s/\lambda_i, 1, \cdots, 1\right)}{(s/\lambda_i) G_i'(1, \cdots, 1, z_i, 1, \cdots, 1)\mid_{z_i=1}} \frac{s(1-\rho_i)}{s - \lambda_i + \lambda_i \tilde{B}_i(s)}, \tag{3}$$

where the second term is the well-known Pollaczek-Khintchine transform for the waiting time in the standard $M/G/1$ queue. It follows from (3) that $\tilde{W}_i^0(s)$ can be calculated once $G_i(1, \cdots, 1, z_i, 1, \cdots, 1)$ is given.

We will focus on queue 1 and thus address the problem of determining $G_1(z_1, 1, \cdots, 1)$. Consider a randomly-selected 1-polling epoch, and call it the *reference point.* Let Y_1 denote the number of waiting customers at queue 1 at the reference point; then, Y_1 has PGF $G_1(z_1, 1, \cdots, 1)$.

Define a *cycle* as the elapsed time between two successive 1-polling epochs; and let the reference point correspond to the initiation of a cycle with cycle index $c = -1$. Clearly, every waiting customer (if any) at queue 1 at the reference point contributes a count of exactly 1 in Y_1, and none of the waiting customers at the other queues contributes any count in Y_1. (Note that the system is never empty at a polling epoch.) Thus, with $L_{i,-1}$ denoting the "contribution" to Y_1 from each waiting customer at queue i at the reference point and with $L_{i,-1}(z)$ denoting the PGF of $L_{i,-1}$, we have $L_{1,-1} = 1$, with $L_{1,-1}(z) = z$, and $L_{i,-1} = 0$, with $L_{i,-1}(z) = 1$, for $i > 1$. Therefore, it follows from (2), by first replacing each z_i in $G_1(z_1, \cdots, z_N)$ by $L_{i,-1}(z_i)$ and then setting all z_is to z_1, that

$$G_1(L_{1,-1}(z_1), L_{2,-1}(z_1), \cdots, L_{N,-1}(z_1)) = G_1(z_1, 1, \cdots, 1) \,. \tag{4}$$

This key observation (the starting point of the descendant-set approach in [12]) suggests that we can evaluate $G_1(z_1, 1, \cdots, 1)$ by considering, recursively, contributions to Y_1 from waiting customers at all queues at each of the *past* polling epochs, working backward from the reference point.

Look backward in time and consider the cth cycle prior to the reference point, where $c = 0,\, 1,\, \cdots$. Let $\mathcal{C}_{i,c}$ be a customer (if any) served at queue i during the cth cycle. Define the immediate *children* of $\mathcal{C}_{i,c}$ as the set of all customers arriving to the system (at all queues) during the service to $\mathcal{C}_{i,c}$, and define the *descendant set* of $\mathcal{C}_{i,c}$ as the set consisting of $\mathcal{C}_{i,c}$, its children (if any), and the descendants of its children. Let $L_{i,c}$ be the number of waiting customers at queue 1 at the reference point that are also in the descendant set of $\mathcal{C}_{i,c}$, and let $L_{i,c}(z)$ denote its PGF. As is the case for $L_{i,-1}$, we will say that $L_{i,c}$ is the contribution to Y_1 from customer $\mathcal{C}_{i,c}$.

For $c \geq 0$, the PGF of $L_{i,c}$ can be expressed in terms of the PGF of the contributions to Y_1 made by customers who arrive during the service to $\mathcal{C}_{i,c}$, as follows. Customers arriving during this service at a queue $j \geq i$ are served in the current cycle; these are, by definition, $\mathcal{C}_{j,c}$ customers, and hence each has a contribution with PGF $L_{j,c}(z)$. Similarly, customers arriving at a queue $j < i$ are served in the next cycle; these are, by definition, $\mathcal{C}_{j,c-1}$ customers, and hence each has a contribution with PGF $L_{j,c-1}(z)$. Therefore, it follows from (1) (with $S = B_i$) and (2) (similar to (4)) that

$$L_{i,c}(z) = \tilde{B}_i \left(\sum_{j=i}^{N} \lambda_j \left[1 - L_{j,c}(z)\right] + \sum_{j=1}^{i-1} \lambda_j \left[1 - L_{j,c-1}(z)\right] \right), \tag{5}$$

for $i = 1, \cdots, N$ and $c \geq 0$.

To see how (5) relates to the calculation of $G_1(z_1, 1, \cdots, 1)$, we consider the i-polling epoch during the cth cycle, where $c = -1,\, 0\,,\, 1,\, \cdots$. Let $G_{i,c}(z_1, \cdots, z_N)$ be the PGF of the numbers of waiting customers at the various queues at this epoch and note, in particular, that $G_{1,-1}(z_1, \cdots, z_N) \equiv G_1(z_1, \cdots, z_N)$. Then, an argument similar to that for (4) immediately yields that the PGF of the total contributions to Y_1 made by these waiting customers is given by

$$G_{i,c}(L_{1,c-1}(z), \cdots, L_{i-1,c-1}(z), L_{i,c}(z), \cdots, L_{N,c}(z)) \,, \tag{6}$$

where the PGFs $L_{i,c}(z)$, $i = 1, \cdots, N$ and $c \geq 0$, are determined by the recursion (5), with the initial condition

$$(L_{1,-1}(z), L_{2,-1}(z), \cdots, L_{N,-1}(z)) = (z, 1, \cdots, 1)\,. \tag{7}$$

The idea, then, is to calculate the PGF $G_{1,-1}(z, 1, \cdots, 1)$ by relating it recursively to the $G_{i,c}(\cdot)$s in (6) for $i = 1, \cdots, N$ and $c \geq 0$, via a Markov chain embedded at polling epochs, monitoring the numbers of waiting customers at the various queues. Note that by stationarity, we have, for all $c \geq 0$,

$$G_{i,c}(L_{1,c-1}(z), \cdots, L_{i-1,c-1}(z), L_{i,c}(z), \cdots, L_{N,c}(z)) = \\ G_{i,-1}(L_{1,c-1}(z), \cdots, L_{i-1,c-1}(z), L_{i,c}(z), \cdots, L_{N,c}(z))\,, \tag{8}$$

a fact that will be needed later.

Define a *server-departure epoch* from queue i as a time epoch at which the server has just completed service (if any) at queue i and is ready to leave that queue. (The set of server-departure epochs includes, as a subset, the set of switch points used in [3]. The difference between the two sets occurs when the system becomes empty at the time the server completes service at a queue. This results in a single switch point, and a server-departure epoch will be registered at this instant as well; however, this will be followed by the simultaneous registration of a random number of server-departure epochs when the next customer arrives.) With $\mathbf{0}$ denoting a vector of zeros, let $P_{i,c}(\mathbf{0})$ be the probability that the system is empty at the server-departure epoch at queue i during the cth cycle. Note that, again by stationarity, $P_{i,c}(\mathbf{0})$ is independent of c. The following lemma gives a recursion for the PGFs in (6).

Lemma 1 *For $i = 1, \cdots, N$ and $c \geq -1$,*

$$G_{i,c}(L_{1,c-1}(z), \cdots, L_{i-1,c-1}(z), L_{i,c}(z), \cdots, L_{N,c}(z)) = \\ G_{i-1,c}(L_{1,c-1}(z), \cdots, L_{i-2,c-1}(z), L_{i-1,c}(z), \cdots, L_{N,c}(z)) \\ - P_{i-1,c}(\mathbf{0}) \left(\sum_{j=i}^{N} a_j \left[1 - L_{j,c}(z)\right] + \sum_{j=1}^{i-1} a_j \left[1 - L_{j,c-1}(z)\right] \right)\,, \tag{9}$$

where $a_j \equiv \lambda_j / \lambda$.

We note that if $i = 1$ on the left-hand side of (9), then the corresponding indices i and c on the right-hand side need to be replaced by $N + 1$ and $c + 1$ respectively (that is, the 1-polling epoch in the cth cycle can be thought of as the $(N + 1)$-polling epoch in the $(c + 1)$th cycle); and that if $c = -1$, then $L_{j,c-1}(z)$ is defined as 1 for any $j = 1, \cdots, N$.

Lemma 1 is proved as follows. Consider the server-departure epoch from queue $i - 1$ in the cth cycle, and let $D_{i-1,c}(z_1, \cdots, z_N)$ be the PGF of the numbers of waiting customers at the various queues at this epoch. Then, similar to (6), the PGF of the total contributions to Y_1 made by these waiting customers is given by

$$D_{i-1,c}(L_{1,c-1}(z), \cdots, L_{i-2,c-1}(z), 1, L_{i,c}(z), \cdots, L_{N,c}(z)) . \tag{10}$$

If the entire system is not empty at this epoch, then this epoch coincides with the subsequent i-polling epoch; and hence (10) gives the PGF of the total contributions to Y_1 from waiting customers at the subsequent i-polling epoch as well. On the other hand, if the system is empty at this server-departure epoch, then the server waits at queue $i-1$ for the next arrival, say at queue j, which occurs with probability $a_j = \lambda_j/\lambda$; and as soon as the arrival takes place, the server cycles around to that queue in zero time, even if $j = i-1$, in the order i, $\cdots$, $j-1$, j, registering (simultaneously) a polling epoch at each of these queues, including in particular an i-polling epoch at queue i. At this i-polling epoch, the new arrival, a "supercustomer" (see [1]), will contribute either $L_{j,c}$ or $L_{j,c-1}$ to Y_1, depending on whether $j \geq i$ or $j < i$. Since $P_{i-1,c}(\mathbf{0})$ equals the probability that the system is empty at this server-departure epoch, and since the PGF of $\mathbf{0}$ equals 1, it follows that

$$G_{i,c}(L_{1,c-1}(z), \cdots, L_{i-1,c-1}(z), L_{i,c}(z), \cdots, L_{N,c}(z)) =$$

$$D_{i-1,c}(L_{1,c-1}(z), \cdots, L_{i-2,c-1}(z), 1, L_{i,c}(z), \cdots, L_{N,c}(z)) - P_{i-1,c}(\mathbf{0}) \cdot 1$$
$$+ P_{i-1,c}(\mathbf{0}) \left(\sum_{j=i}^{N} a_j L_{j,c}(z) + \sum_{j=1}^{i-1} a_j L_{j,c-1}(z) \right) , \tag{11}$$

with the last two terms accounting for the needed "correction" between the PGFs $G_{i,c}(\cdot)$ and $D_{i-1,c}(\cdot)$, due to the potential arrival of a supercustomer.

We next relate $D_{i-1,c}(\cdot)$ to $G_{i-1,c}(\cdot)$. After replacing z_{i-1} in $G_{i-1,c}(z_1, \cdots, z_{i-1}, \cdots, z_N)$ by $\tilde{\Theta}_{i-1}(\sum_{j\neq i-1} \lambda_j(1-z_j))$, we obtain (from a further extension of (2), with possibly multivariate K_{ij}s for each i)

$$D_{i-1,c}(z_1, \cdots, z_{i-2}, 1, z_i, \cdots, z_N) =$$

$$G_{i-1,c}\left(z_1, \cdots, z_{i-2}, \tilde{\Theta}_{i-1}\left(\sum_{j\neq i-1} \lambda_j(1-z_j) \right), z_i, \cdots, z_N \right) . \tag{12}$$

In [12] (eq. (3.6), p. 1248), using arguments similar to that used to derive (5), it is shown that $L_{i,c}(z) = \tilde{\Theta}_i \left(\sum_{j=i+1}^{N} \lambda_j [1 - L_{j,c}(z)] + \sum_{j=1}^{i-1} \lambda_j [1 - L_{j,c-1}(z)] \right)$, for $i = 1, \cdots, N$ and $c \geq 0$. Hence, by first replacing each z_k in (12) with either $L_{k,c-1}(z)$ or $L_{k,c}(z)$, depending on whether $k < i-1$ or $k > i-1$, and then using this recursion for $L_{i,c}(z)$, it follows that

$$D_{i-1,c}(L_{1,c-1}(z), \cdots, L_{i-2,c-1}(z), 1, L_{i,c}(z), \cdots, L_{N,c}(z)) =$$

$$G_{i-1,c}(L_{1,c-1}(z), \cdots, L_{i-2,c-1}(z), L_{i-1,c}(z), \cdots, L_{N,c}(z)) \,. \tag{13}$$

More directly, (13) can also be seen as a consequence of the simple observation that the two PGFs of the total contributions to Y_1 made by waiting customers at the i-polling epoch in the cth cycle and at the subsequent server-departure epoch from queue i must be the same, because the system stays nonempty *between* these two epochs. Finally, substitution of (13) into (11), with the PGF of $\mathbf{0}$ in (11) rewritten as $\sum_{j=1}^{N} a_j$ (which can be interpreted as the PGF of the contribution to Y_1 from a "null" customer, located at queue j with probability a_j), yields (9); and this proves Lemma 1.

Now, with $i = 1$ in (9), we have

$$\begin{aligned} G_{1,c}(L_{1,c}(z), \cdots, L_{N,c}(z)) &= G_{N,c+1}(L_{1,c}(z), \cdots, L_{N-1,c}(z), L_{N,c+1}(z)) \\ &\quad - P_{N,c+1}(\mathbf{0}) \sum_{j=1}^{N} a_j \left[1 - L_{j,c}(z)\right]; \end{aligned} \tag{14}$$

and recursively applying (9) N times, we obtain

$$\begin{aligned} G_{1,c}\left(L_{1,c}(z), \cdots, L_{N,c}(z)\right) &= G_{1,c+1}\left(L_{1,c+1}(z), \cdots, L_{N,c+1}(z)\right) \\ - \sum_{i=1}^{N} P_{i,c+1}(\mathbf{0}) &\left(\sum_{j=i+1}^{N} a_j \left[1 - L_{j,c+1}(z)\right] + \sum_{j=1}^{i} a_j \left[1 - L_{j,c}(z)\right] \right) . \end{aligned} \tag{15}$$

Continuing to express $G_{1,-1}(L_{1,-1}(z), \cdots, L_{N,-1}(z))$ in terms of $G_{1,0}(L_{1,0}(z), \cdots, L_{N,0}(z))$, and $G_{1,0}(L_{1,0}(z), \cdots, L_{N,0}(z))$ in terms of $G_{1,1}(L_{1,1}(z), \cdots, L_{N,1}(z))$, and so on, using (15) with $c = -1$, $c = 0$, etc., we obtain

$$\begin{aligned} G_{1,-1}\left(L_{1,-1}(z), \cdots, L_{N,-1}(z)\right) &= G_{1,\infty}\left(L_{1,\infty}(z), \cdots, L_{N,\infty}(z)\right) \\ - \sum_{c=-1}^{\infty} \sum_{i=1}^{N} P_{i,c+1}(\mathbf{0}) &\left(\sum_{j=i+1}^{N} a_j \left[1 - L_{j,c+1}(z)\right] + \sum_{j=1}^{i} a_j \left[1 - L_{j,c}(z)\right] \right) . \end{aligned} \tag{16}$$

Since $P_{i,c}(\mathbf{0})$ is independent of c, the last term on the right-hand side of (16) simplifies, after an interchange of the order of summation, as follows:

$$\begin{aligned} &\sum_{c=-1}^{\infty} \sum_{i=1}^{N} P_{i,c+1}(\mathbf{0}) \left(\sum_{j=i+1}^{N} a_j \left[1 - L_{j,c+1}(z)\right] + \sum_{j=1}^{i} a_j \left[1 - L_{j,c}(z)\right] \right) \\ &= \sum_{i=1}^{N} P_i(\mathbf{0}) \left(\sum_{c=-1}^{\infty} \sum_{j=i+1}^{N} a_j \left[1 - L_{j,c+1}(z)\right] + \sum_{c=-1}^{\infty} \sum_{j=1}^{i} a_j \left[1 - L_{j,c}(z)\right] \right) \\ &= P(\mathbf{0}) \sum_{c=-1}^{\infty} \sum_{j=1}^{N} a_j [1 - L_{j,c}(z)] \,, \end{aligned} \tag{17}$$

where $P(\mathbf{0}) \equiv \sum_{i=1}^{N} P_i(\mathbf{0}) \equiv \sum_{i=1}^{N} P_{i,c}(\mathbf{0})$, and where the last equality is true because $1 - L_{j,-1}(z) = 0$ for $j > 1$. Now, in the limit as $c \to \infty$, it can be shown from (5) (similar to [3], Section VI) that $L_{i,c}(z) \to L_{i,\infty}(z) = 1$ (i.e., the contribution to Y_1 made by a customer who is served c cycles ago tends to 0 as c tends to infinity). After substituting (17) into (16), setting $G_1(\cdot) \equiv G_{1,c}(\cdot)$ for all c (see (8)), and noting that $G_1(1, \cdots, 1) = 1$, we obtain the following key result.

Lemma 2

$$G_1(z, 1, \cdots, 1) = 1 - P(\mathbf{0})\, H_1(z)\,, \tag{18}$$

where

$$H_1(z) \equiv \sum_{c=-1}^{\infty} \sum_{i=1}^{N} a_i \left[1 - L_{i,c}(z)\right] . \tag{19}$$

From (18) and (19), we have

$$G_1'(z, 1, \cdots, 1) \mid_{z=1} = -P(\mathbf{0})\, H_1'(z) \mid_{z=1} = P(\mathbf{0}) \sum_{c=-1}^{\infty} \sum_{i=1}^{N} \ell_{i,c}\,, \tag{20}$$

where $\ell_{i,c} \equiv a_i E[L_{i,c}]$. It is easily shown from (5) that

$$\ell_{i,c} = \rho_i \left[\sum_{j=i}^{N} \ell_{j,c} + \sum_{j=1}^{i-1} \ell_{j,c-1} \right], \tag{21}$$

for $i = 1, \cdots, N$ and $c \geq 0$. After summing over c from 0 to ∞ and i from 1 to N in (21) and simplifying (similar to (17)), we obtain

$$\sum_{c=0}^{\infty} \sum_{i=1}^{N} \ell_{i,c} = \sum_{i=1}^{N} \rho_i \left[\sum_{c=0}^{\infty} \sum_{j=i}^{N} \ell_{j,c} + \sum_{c=0}^{\infty} \sum_{j=1}^{i-1} \ell_{j,c-1} \right] = \sum_{i=1}^{N} \rho_i \sum_{c=0}^{\infty} \sum_{j=1}^{N} \ell_{j,c} + \sum_{i=2}^{N} \rho_i a_1\,,$$

where we have used the fact that $\sum_{j=1}^{k} \ell_{j,-1} = a_1$ for all $k = 1, \cdots, N$ (which follows because $\ell_{1,-1} = a_1$ and $\ell_{i,-1} = 0$ for $i > 1$). Thus,

$$\sum_{c=0}^{\infty} \sum_{i=1}^{N} \ell_{i,c} = a_1 \sum_{i=2}^{N} \frac{\rho_i}{1-\rho} = a_1 \frac{\rho - \rho_1}{1 - \rho}\,.$$

Adding $\sum_{i=1}^{N} \ell_{i,-1} = a_1$ to the above equation, we obtain $\sum_{c=-1}^{\infty} \sum_{i=1}^{N} \ell_{i,c} = a_1(1-\rho_1)/(1-\rho)$, which, together with (20), leads to the following lemma.

Lemma 3

$$G_1'(z, 1, \cdots, 1) \mid_{z=1} = \lambda_1 \frac{P(\mathbf{0})}{\lambda(1-\rho)} \,(1 - \rho_1)\,. \tag{22}$$

Lemma 3 can also be argued directly, independent of (5). Since $G_1'(z,1,\cdots,1)\mid_{z=1}$ is the expected number of customers waiting at queue 1 when the server returns from a vacation V_1^0 away from that queue, it follows from an application of formula (3) in Wolff [19] (p. 226) that $G_1'(z,1,\cdots,1)\mid_{z=1}=\lambda_1 E[V_1^0]$ (note that although V_1^0 and the arrival process are dependent, Wolff's lack-of-anticipation assumption holds). To calculate $E[V_1^0]$, observe that $P(\mathbf{0})/\lambda$ equals the expected amount of time in a cycle the server is idle, and let C be the length of a cycle. Then, an application of "$H=\lambda G$" (see, e.g., Heyman and Stidham [11]), with "H" $=1-\rho$, "λ" $=1/E[C]$, and "G" $=P(\mathbf{0})/\lambda$, yields $E[C]=P(\mathbf{0})/[\lambda(1-\rho)]$; and since $E[V_1^0]=E[C](1-\rho_1)$ (similar to, e.g., (3.2) and (3.3) in [16], pp. 277–278), this again proves Lemma 3.

Finally, by combining (3), Lemma 2, and Lemma 3 (and cancelling $P(\mathbf{0})$ in (18) and (22)), we obtain the following result.

Theorem 1 *Under the exhaustive-service discipline, the LST of the waiting time in the zero-switchover-times model is*

$$\tilde{W}_1^0(s)=\left[\frac{\lambda(1-\rho)}{1-\rho_1}\,\frac{H_1(1-s/\lambda_1)}{s}\right]\tilde{W}_1^*(s)\,. \tag{23}$$

Formula (23), which we believe is new, complements formula (27), p. 407, in [1]. As noted at the beginning of this section, the latter formula was derived based on a Markov chain embedded at switch points, whereas ours is based on a Markov chain embedded at polling epochs. It can be shown, via a proper "translation" between the solutions to these two embedded chains, that these formulas are equivalent.

4 A Transform Result

The vacation-model decomposition (3) also constitutes the basis for the derivation of the waiting-time distribution in the nonzero-switchover-times model: Define the reference point as a randomly-selected 1-polling epoch, let X_1 be the number of waiting customers at the reference point, and let $F_1(z)$ be the PGF of X_1. Then (for instance, see [16], eqs. (3.17a) and (3.3)),

$$\tilde{W}_1(s)=\frac{1-F_1(1-s/\lambda_1)}{s\,E[V_1]}\,\tilde{W}_1^*(s)\,, \tag{24}$$

where

$$E[V_1]=\frac{E[R]}{1-\rho}\,(1-\rho_1)\,, \tag{25}$$

the expected duration of a vacation V_1 away from queue 1.

Again, define a cycle as the elapsed time between two successive 1-polling epochs. Then, it is shown in [12] that

$$X_1 = \sum_{c=0}^{\infty} \sum_{i=1}^{N} R_{i,c} \, , \tag{26}$$

where $R_{i,c}$ denotes the total contribution to X_1 from customers (or "originators") who arrived (a) in the cth cycle prior to the reference point and (b) during the switchover time from queue i. Noting that the variables $\{R_{i,c} : \ i = 1, \cdots, N \text{ and } c \geq 0\}$ are independent of one another, the representation (26) leads (upon taking PGFs) to

$$F_1(z) = \prod_{c=0}^{\infty} \prod_{i=1}^{N} R_{i,c}(z) \, , \tag{27}$$

where (see (3.4a) and (3.4b) in [12]), for $i = 1, \cdots, N$ and $c \geq 0$,

$$R_{i,c}(z) = \tilde{R}_i \left(\sum_{j=i+1}^{N} \lambda_j [1 - L_{j,c}(z)] + \sum_{j=1}^{i} \lambda_j [1 - L_{j,c-1}(z)] \right) , \tag{28}$$

and the PGFs $L_{j,c}(z)$ satisfy the same recursion (5) (which is equivalent to (3.6) in [12]), with the same initial condition (7) for the zero-switchover-times model. Hence, we see that the set of variables $\{L_{i,c} : \ i = 1, \cdots, N \text{ and } c \geq -1\}$ provides the underlying *link* between (i.e., serves as the common "input" to) these models; indeed, our analysis in Section 3 overcomes, with the recursion (9), the difficulty with the fact that a simple counterpart of (26) does not exist for the zero-switchover-times model.

Comparison of (23) with (24) and (25) leads to the following transform result.

Theorem 2 *Under the exhaustive-service discipline,*

$$\tilde{W}_1(s) = \frac{1 - F_1(1 - s/\lambda_1)}{H_1(1 - s/\lambda_1) \, \lambda E[R]} \, \tilde{W}_1^0(s) \, . \tag{29}$$

Theorem 2 gives a simple relation between the LSTs of the waiting times in the two models. In Section 6, on computational issues, we will demonstrate that Theorem 2 also has an important practical significance: It allows one to compute any desired waiting-time moments in the nonzero-switchover-times model for a given set of switchover times with $O(N)$ computations, once the corresponding waiting-time moments have been computed for the zero-switchover-times model.

Now consider the special case where the switchover times are all constants (not necessarily equal). In this case, we obtain a strikingly simple relation between the LSTs of the waiting times in the two models. From (27) and (28), we have

$$\begin{aligned} F_1(z) &= \prod_{c=0}^{\infty}\prod_{i=1}^{N} \exp\left[-\left(\sum_{j=i+1}^{N} \lambda_j[1-L_{j,c}(z)] + \sum_{j=1}^{i}\lambda_j[1-L_{j,c-1}(z)]\right) E[R_i]\right] \\ &= \prod_{i=1}^{N} \exp\left[-\sum_{c=0}^{\infty}\left(\sum_{j=i+1}^{N} \lambda_j[1-L_{j,c}(z)] + \sum_{j=1}^{i}\lambda_j[1-L_{j,c-1}(z)]\right) E[R_i]\right]. \end{aligned}$$

Next, an argument similar to (17) yields

$$F_1(z) = \prod_{i=1}^{N} \exp\left\{-H_1(z)\,\lambda E[R_i]\right\} = \exp\left(-H_1(z)\,\lambda E[R]\right). \tag{30}$$

From (29) and (30), we then obtain the following results.

Theorem 3 *If the switchover times are all constants, then under the exhaustive-service discipline,*

$$\tilde{W}_1(s) = \frac{1-\exp\left(-H_1(1-s/\lambda_1)\,\lambda E[R]\right)}{H_1(1-s/\lambda_1)\,\lambda E[R]}\,\tilde{W}_1^0(s)\,. \tag{31}$$

Corollary 1 *When the switchover times are all constants, the waiting-time distribution in the nonzero-switchover-times model depends only on the sum of the switchover times, and not on the individual switchover times.*

The relation (31) can be used, for example, to obtain simple relationships between these two models for the expectations and variances of the waiting times, when the switchover times are all constants. By differentiating (31) once and twice and simplifying the resulting expressions (using (25)), we obtain

$$E[W_1] = E[W_1^0] + \frac{E[V_1]}{2} \tag{32}$$

and

$$Var[W_1] = Var[W_1^0] + \frac{E[V_1]^2}{12} + E[V_1]\left(E[W_1^0] - E[W_1^*]\right). \tag{33}$$

More generally, it can be shown that when the switchover times are all constants, the nth waiting-time moment for the nonzero-switchover-times model can be expressed in terms of the nth and lower moments of the waiting time in the zero-switchover-times model; if these moments for the latter model have already been computed, then the computational effort involved is only $O(1)$.

It can be shown from Theorem 3 that a simple application of L'Hopital's rule verifies that $\lim_{E[R]\to 0} \tilde{W}_1(s) = \tilde{W}_1^0(s)$, thus establishing continuity for the case of constant switchover times. For the case of general, not necessarily constant, switchover times, such limiting procedures can similarly be provided. It is important to note, however, that our theorems render such procedures unnecessary.

5 Extension

The basis of our analysis in Section 3 is the decomposition formula (3), which is a particular case of decomposition (for the exhaustive-service polling model) in vacation models. To extend our analysis to cover disciplines other than exhaustive service, we begin with the more general decomposition formula in [9] (Propositions 3 and 4, pp. 1125 and 1126):

$$\tilde{W}_i^0(s) = \chi_i(1 - s/\lambda_i)\,\tilde{W}_i^*(s)\,, \tag{34}$$

where $\chi_i(\cdot)$ is the PGF of K_i, defined as the number of waiting customers at queue i as seen by a randomly-selected customer who arrives at queue i during a vacation from queue i. In Wolff [20] (eq. (75), p. 460), it is shown that

$$P\{K_i = k\} = \frac{P\{T_i \le k,\, T_i + U_i > k\}}{E[U_i]}\,, \tag{35}$$

where T_i is the number of waiting customers at queue i at the start of a vacation from queue i, and U_i (possibly dependent on T_i) is the number of customers that arrive at queue i during the ensuing vacation. (In fact, Wolff's argument shows that (35) is valid independent of the vacation-model context.) Upon writing $P\{T_i \le k,\, T_i + U_i > k\}$ as $P\{T_i + U_i > k\} - P\{T_i > k\}$ and taking PGFs, it is easily shown that (35) leads to

$$\chi_i(z) = E\left[z^{K_i}\right] = \frac{1}{E[U_i]}\left\{\frac{1 - E\left[z^{T_i+U_i}\right]}{1-z} - \frac{1 - E\left[z^{T_i}\right]}{1-z}\right\}. \tag{36}$$

Substituting (36) into (34) and simplifying yields the following result:

$$\tilde{W}_i^0(s) = \frac{D_i\,(1,\cdots,1,1-s/\lambda_i,1,\cdots,1) - G_i\,(1,\cdots,1,1-s/\lambda_i,1,\cdots,1)}{(s/\lambda_i)\,\left[G_i'(1,\cdots,1,z_i,1,\cdots,1) - D_i'\,(1,\cdots,1,z_i,1,\cdots,1)\right]|_{z_i=1}}\,\tilde{W}_i^*(s)\,, \tag{37}$$

where $D_i(\cdot)$ denotes the PGF of the numbers of waiting customers at the various queues at a randomly-selected server-departure epoch at queue i.

Formula (37) is a very useful generalization of (3). In the polling-models context, it allows one to reduce the calculation of waiting times to that of the pair of PGFs $G_i(\cdot)$ and $D_i(\cdot)$. Since for a wide variety of service disciplines, the PGF $D_i(\cdot)$ can be easily related to $G_i(\cdot)$, the problem again reduces, as in Section 3, to one of finding $G_i(\cdot)$. As explicit examples, we have $D_i\,(1,\cdots,1,1-s/\lambda_i,1,\cdots,1) = 1$ for the exhaustive-service case; and

$$D_i(1,\cdots,1,1-s/\lambda_i,1,\cdots,1) = G_i\left(1,\cdots,1,\tilde{B}_i\,(s)\,,1,\cdots,1\right) \tag{38}$$

for the gated-service case. Clearly, combinations of these and many other cases can be handled similarly (for related discussion, see [8], Section 5).

With (37), our analysis in Sections 3 and 4 extends easily. Note that the recursion (9) remains valid, as stated. To see this, simply replace the argument 1 in (10), (11), and (13) by $L_{i-1,c-1}(z)$ (because customers, if any, left behind at queue $i-1$ at the server-departure epoch at that queue during the cth cycle are, by definition, $\mathcal{C}_{i-1,c-1}$ customers, each having a contribution of $L_{i-1,c-1}$ to Y_1). Therefore, only (5) (the common input) needs to be revised. For instance, for the gated case, using arguments similar to that used to obtain (5), we have (also, see [12])

$$L_{i,c}(z) = \tilde{B}_i \left(\sum_{j=i+1}^{N} \lambda_j \left[1 - L_{j,c}(z)\right] + \sum_{j=1}^{i} \lambda_j \left[1 - L_{j,c-1}(z)\right] \right) , \tag{39}$$

for $i = 1, \cdots, N$ and $c \geq 0$. The corresponding analogues of Theorems 1, 2, and 3 for the gated-service discipline are:

Theorem 4 *Under the gated-service discipline, the LST of the waiting time in the zero-switchover-times model is*

$$\tilde{W}_1^0(s) = \left[\frac{\lambda(1-\rho)}{1-\rho_1} \, \frac{H_1(1 - s/\lambda_1) - H_1(\tilde{B}_1(s))}{s} \right] \tilde{W}_1^*(s) \, . \tag{40}$$

Theorem 5 *Under the gated-service discipline,*

$$\tilde{W}_1(s) = \frac{F_1(\tilde{B}_1(s)) - F_1(1 - s/\lambda_1)}{\left[H_1(1 - s/\lambda_1) - H_1(\tilde{B}_1(s))\right] \lambda E[R]} \, \tilde{W}_1^0(s) \, . \tag{41}$$

Theorem 6 *If the switchover times are all constants, then under the gated-service discipline,*

$$\tilde{W}_1(s) = \frac{\exp(-H_1(\tilde{B}_1(s)) \, \lambda E[R]) - \exp(-H_1(1 - s/\lambda_1) \, \lambda E[R])}{\left[H_1(1 - s/\lambda_1) - H_1(\tilde{B}_1(s))\right] \lambda E[R]} \, \tilde{W}_1^0(s) \, . \tag{42}$$

The functions $H_1(\cdot)$ and $F_1(\cdot)$ in the above theorems are still given by (19) and (27), respectively. Note that Corollary 1 holds for the gated-service discipline as well.

6 Computation

In this section, we first develop a technique for computing waiting-time moments in the zero-switchover-times model, and discuss its computational complexity. Next we demonstrate how waiting-time moments for the nonzero-switchover-times model with variable switchover times can be obtained with $O(N)$ computations, once the corresponding moments have been computed for the zero-switchover-times model. (As indicated in Section 4, the special case of constant switchover times leads to much simpler computations.) For ease of presentation, we restrict attention to the first two moments of the waiting time in a polling model in which all queues are served according to the exhaustive-service discipline.

Differentiating (23) twice with respect to s and evaluating the result at $s = 0$, we obtain

$$E[W_1^0] = E[W_1^*] + \frac{h_1^{(2)}}{2\lambda_1 h_1^{(1)}} \tag{43}$$

and

$$E[(W_1^0)^2] = E[(W_1^*)^2] + 2E[W_1^*](E[W_1^0] - E[W_1^*]) + \frac{h_1^{(3)}}{3\lambda_1^2 h_1^{(1)}}\,, \tag{44}$$

where from (19),

$$h_1^{(n)} \equiv \frac{d^n}{dz^n} H_1(z)\ |_{z=1} = -\sum_{c=-1}^{\infty}\sum_{i=1}^{N} a_i \frac{d^n}{dz^n} L_{i,c}(z)\ |_{z=1}\ . \tag{45}$$

Since the moments of the waiting time in the standard $M/G/1$ queue are easily obtained (see, for example, Takács [15]), computing the moments of W_1^0 involves computing the derivatives of $H_1(z)$ with respect to z. This, in turn, involves the computation of the moments of $\{L_{i,c} : i = 1, \cdots, N \text{ and } c \geq 0\}$. Let $\omega_{i,c}^{(n)}$ denote the nth factorial moment of $L_{i,c}$, that is,

$$\omega_{i,c}^{(n)} \equiv \frac{d^n}{dz^n} L_{i,c}(z)\ |_{z=1},$$

and let

$$\alpha_i \equiv \sum_{c=0}^{\infty} (\lambda_i \omega_{i,c}^{(1)})^2, \qquad \beta_i \equiv \sum_{c=0}^{\infty} (\lambda_i \omega_{i,c}^{(1)})^3, \qquad \text{and} \qquad \gamma_i \equiv \sum_{c=0}^{\infty} (\lambda_i \omega_{i,c}^{(1)})(\lambda_i \omega_{i,c}^{(2)})\,.$$

The $\omega_{i,c}^{(n)}$ terms are obtained from equation (5). Substituting these values in equation (45), we obtain the following expressions for $\{h_1^{(n)} : n = 1, 2, 3\}$.

$$h_1^{(1)} = -\frac{\lambda_1(1-\rho_1)}{\lambda(1-\rho)}\,, \tag{46}$$

$$h_1^{(2)} = -\frac{1}{\lambda(1-\rho)} \sum_{i=1}^{N} \alpha_i \frac{\lambda_i E[B_i^2]}{\rho_i^2}\,, \tag{47}$$

and

$$h_1^{(3)} = -\frac{1}{\lambda(1-\rho)} \sum_{i=1}^{N} \left\{ \beta_i \left[\frac{\lambda_i E[B_i^3]}{\rho_i^3} - 3\left(\frac{\lambda_i E[B_i^2]}{\rho_i^2}\right)^2 \right] + 3\gamma_i \frac{\lambda_i E[B_i^2]}{\rho_i^2} \right\}. \tag{48}$$

These expressions are obtained in a manner similar to that used to derive Lemma 3, and the details of the algebra are omitted.

Although α_i, β_i and γ_i are infinite sums, the iteration does not have to be carried out for a very large value of c, since $\omega_{i,c}^{(n)}$ quickly decreases to 0 as c increases. The computational effort required to obtain any moment of the waiting time, in general, can be shown to be $O(N \log_\rho \epsilon)$ where ϵ is the order of accuracy required, just as in the case of the descendant-set technique presented in [12] for the nonzero-switchover-times model.

It is also of interest to compare the computational effort required by the above approach to that required by techniques presented in the past for the zero-switchover-times model. The technique proposed in [1] requires the solution of a system of $O(N^2)$ equations (which involves $O(N^6)$ arithmetic operations) to evaluate the expected waiting times. Obtaining higher moments of the waiting time with this technique requires considerably more computational effort.

We now consider the nonzero-switchover-times model. In Section 4, we observed that if the switchover times are all constants, then the waiting-time moments for this model are readily obtained once the corresponding moments for the zero-switchover-times model have been computed. We will now demonstrate that the waiting-time moments for the nonzero-switchover-times model can be simply related to the corresponding moments for the zero-switchover-times model, even when the switchover times are not constants. Similar to the expressions for the zero-switchover-times model, the first two moments are obtained from (24) as

$$E[W_1] = E[W_1^*] + \frac{f_1^{(2)}}{2\lambda_1 f_1^{(1)}} \tag{49}$$

and

$$E[(W_1)^2] = E[(W_1^*)^2] + 2E[W_1^*](E[W_1] - E[W_1^*]) + \frac{f_1^{(3)}}{3\lambda_1^2 f_1^{(1)}}, \tag{50}$$

where $f_1^{(n)} \equiv \frac{d^n}{dz^n} F_1(z) \mid_{z=1}$.

Let $Var[R_i]$ and $\mathcal{T}[R_i]$ denote the variance and the third central moment of R_i. It is well known (Kuehn [13]) that $E[C] = E[R]/(1-\rho)$, where C denotes the length of a cycle. For notational ease, define

$$\mathcal{J}_i \equiv \frac{\lambda_i E[B_i^2]}{\rho_i^2} E[C] + \frac{Var[R_{i-1}]}{\rho_i^2}, \qquad \text{and} \qquad \mathcal{K}_i \equiv \frac{\lambda_i E[B_i^3]}{\rho_i^3} E[C] + \frac{\mathcal{T}[R_{i-1}]}{\rho_i^3}.$$

Then, from (27), we obtain the following expressions for $\{f_1^{(n)} : n = 1, 2, 3\}$.

$$f_1^{(1)} = \lambda_1 E[C] (1 - \rho_1), \tag{51}$$

$$f_1^{(2)} = f_1^2 + \lambda_1^2 Var[R_N] + \sum_{i=1}^{N} \alpha_i \mathcal{J}_i \, , \tag{52}$$

and

$$f_1^{(3)} = 3 f_1^{(1)} f_1^{(2)} - 2(f_1^{(1)})^3 + \lambda_1^3 \mathcal{T}[R_N] + \sum_{i=1}^{N} \left\{ \beta_i \left[\mathcal{K}_i - 3 \frac{\lambda_i E[B_i^2]}{\rho_i^2} \mathcal{J}_i \right] + 3 \gamma_i \mathcal{J}_i \right\} . \tag{53}$$

Observe that the bulk of the computational effort required in (51) and (53) is devoted to computing α_i, β_i, and γ_i. These terms have to be computed in order to obtain the first two moments of the waiting time for the zero-switchover-times model. Therefore, if these terms have already been determined, computing the corresponding moments in the nonzero-switchover-times model requires only $O(N)$ arithmetic operations.

References

[1] R. B. Cooper, Queues Served in Cyclic Order: Waiting Times, *The Bell System Technical Journal* **49**(1970)399–413.

[2] R. B. Cooper, *Introduction to Queueing Theory* (Second Edition, North-Holland, Amsterdam, 1981; reprinted by CEEPress of The George Washington University, 1990.).

[3] R. B. Cooper and G. Murray, Queues Served in Cyclic Order, *The Bell System Technical Journal* **48**(1969)675–689.

[4] R. B. Cooper, S.-C. Niu, and M. M. Srinivasan, A Decomposition Theorem for Polling Models: The Switchover Times are Effectively Additive, Technical Report (1992) (to appear in *Operations Research*).

[5] M. Eisenberg, Queues with Periodic Service and Changeover Time, *Operations Research* **20**(1972)440–451.

[6] M. Eisenberg, The Polling System with a Stopping Server, Technical Report, AT&T Bell Laboratories (1993) (to appear in *Queueing Systems*).

[7] S. W. Fuhrmann, Performance Analysis of a Class of Cyclic Schedules, Technical Report, AT&T Bell Laboratories (1981).

[8] S. W. Fuhrmann, A Decomposition Result for a Class of Polling Models, *Queueing Systems* **11**(1992)109–120.

[9] S. W. Fuhrmann and R. B. Cooper, Stochastic Decompositions in the *M/G/1* Queue with Generalized Vacations, *Operations Research* **33**(1985)1117–1129.

[10] G. M. Garner, Mean and Variance of Waiting Time for Nonsymmetric Cyclic Queueing Systems, Technical Report, AT&T Bell Laboratories (1990).

[11] D. P. Heyman and S. Stidham, Jr., The Relation between Customer and Time Averages in Queues, *Operations Research*, **28**(1980)983–994.

[12] 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**(1993)1245–1253.

[13] P. J. Kuehn, Multiqueue Systems with Nonexhaustive Cyclic Service, *The Bell System Technical Journal* **58**(1979)671–698.

[14] H. Levy and L. Kleinrock, Polling Systems with Zero Switch-Over Periods: A General Method for Analyzing Expected Delay, *Performance Evaluation*, **13**(1991)97–107.

[15] L. Takács, A Single-Server Queue with Poisson Input, *Operations Research* **10**(1962)388–397.

[16] H. Takagi, Queueing Analysis of Polling Models: An Update, in *Stochastic Analysis of Computer and Communication Systems*, H. Takagi (Editor), (Elsevier Science Publishers B.V., North Holland, Amsterdam, 1990), 267–318.

[17] H. Takagi, Queueing Analysis of Polling Models: Progress in 1990–93, Institute of Socio-Economic Planning, University of Tsukuba, Japan (May 1994).

[18] S. Tayur and D. Sarkar, Implementing an Effective Algorithm to Calculate Waiting Times in Nonsymmetric Cyclic Queueing System, Technical Report, AT&T Bell Laboratories (1989).

[19] R. W. Wolff, Poisson Arrivals See Time Averages, *Operations Research* **30**(1982)223–231.

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