J. Appl. Prob. **36,** 585–592 (1999)
Printed in Israel

SETUPS IN POLLING MODELS: DOES IT MAKE SENSE TO SET UP IF NO WORK IS WAITING?

ROBERT B. COOPER,* *Florida Atlantic University*
SHUN-CHEN NIU,** *University of Texas at Dallas*
MANDYAM M. SRINIVASAN,*** *University of Tennessee*

Abstract

We compare two versions of a symmetric two-queue polling model with switchover times and setup times. The SI version has *State-Independent setups*, according to which the server sets up at the polled queue whether or not work is waiting there; and the SD version has *State-Dependent setups*, according to which the server sets up only when work is waiting at the polled queue. Naive intuition would lead one to believe that the SD version should perform better than the SI version. We characterize the difference in the expected waiting times of these two versions, and we uncover some surprising facts. In particular, we show that, regardless of the server utilization or the service-time distribution, the SD version performs (i) the same as, (ii) worse than, or (iii) better than its SI counterpart if the switchover and setup times are, respectively, (i) both constants, (ii) variable (i.e. non-deterministic) and constant, or (iii) constant and variable. Only (iii) is consistent with naive intuition.

Keywords: Polling models; state-dependent setup times; waiting times; switchover times; vacation models; decomposition

AMS 1991 Subject Classification: Primary 60K25
Secondary 60K30; 90B22

1. Introduction

The following classical polling model has been the object of recent studies: N queues are served in cyclic order by a single server that travels from queue to queue, switching from a queue only when that queue is empty (that is, exhaustive service). In isolation, each queue would be an ordinary $M/G/1$ queue; but linked together through the sharing of a single server, they interact in complicated ways, and their performance measures are strongly dependent on each other.

Received 14 August 1997; revision received 11 March 1998.
Research supported in part by the National Science Foundation under grants DMI-9500216, 9500040, 9500471.
* Postal address: Department of Computer Science and Engineering, Florida Atlantic University, Boca Raton, FL 33431-0991, USA. Email address: bob@cse.fau.edu.
** Postal address: School of Management, The University of Texas at Dallas, P.O. Box 830688, Richardson, TX 75083-0688, USA. Research also supported in part by a Summer Research Grant from the School of Management, The University of Texas at Dallas.
*** Postal address: Management Science Program, College of Business Administration, The University of Tennessee, Knoxville, TN 37996-0562, USA.

In the context of modelling communications systems, the 'dead time' (time during which the server is not serving any customers even though some may be waiting) required for the server to travel from one queue to the next is called the *switchover time*. In the context of manufacturing systems, this travel time might be relatively negligible, but there can be a *setup time*, which is the dead time during which the server is being adjusted or prepared to begin work at a queue. In this latter context, it is natural to assume that the server sets up at a queue only when there is work waiting. However, because of this state dependency, polling models with *State-Dependent setups* (i.e. the SD model, where the server sets up only when it finds work waiting in the polled queue) have proven to be much harder to analyse than their counterparts with *State-Independent setups* (i.e. the SI model, where the server sets up at each queue whether or not work is waiting at that queue). Of the limited number of papers on SD models, we refer the reader to Ferguson (1986), Bradlow and Byrd (1987), Altman *et al.* (1994), Eisenberg (1995), Günalay (1995), Lennon (1995), Gupta and Srinivasan (1996), Olsen (1996), and Günalay and Gupta (1997). Most of these papers give only methods for approximation, and none gives explicit, closed-form results.

Common sense dictates that the performance of an SD model should be at least as good as that of its SI counterpart. Consequently, the SI model, which is much easier to analyse, has been used as a surrogate to give a conservative bound on the corresponding SD model (see, for example, Ferguson (1986)). Strangely, this common-sense intuition is not always correct. In fact, recent studies have shown that the apparently complicated interplay between the dead times and the serving times in polling models can lead to rather surprising properties. Sarkar and Zangwill (1991) showed by numerical examples that a reduction in setup times in an SI model can actually lead to an *increase* in work-in-process (or expected waiting times). Fuhrmann (1992), Srinivasan *et al.* (1995), and Cooper *et al.* (1996) showed that some SI polling models exhibit a *decomposition*, according to which the effects of the dead times 'separate' from the ordinary effects of queueing, and Cooper *et al.* (1998) applied this theory to characterize and 'explain' the observations of Sarkar and Zangwill.

Recently, Gupta and Srinivasan (1996) gave an algorithm for the exact calculation of the expected waiting times in an SD model with $N = 2$; and they applied their algorithm to Ferguson's model and showed, numerically, that his common-sense bounds can be incorrect. With the benefit of hindsight, this is not too surprising given Sarkar and Zangwill's observations: if reducing setup times in an SI model can increase waiting times, then eliminating some setup times (as in an SD model) might very well have the same counter-intuitive effect.

In this paper, we further investigate the problem of characterizing when the SD model outperforms its SI counterpart, and *vice versa*. Specifically, we examine the symmetric exhaustive-service polling model with $N = 2$ queues (and, for completeness, the case when $N = 1$, which is essentially a vacation model); and we derive an explicit formula that characterizes the difference in expected waiting times in the SD and the SI models. As particular consequences of this characterization, we show that, regardless of the server utilization or the service-time distribution, the SD model performs (i) the same as, (ii) worse than, or (iii) better than its SI counterpart if the switchover and setup times are, respectively, (i) both constants, (ii) variable (i.e. non-deterministic) and constant, or (iii) constant and variable. Only (iii) is consistent with naive intuition. The formal statements of these results will be given in Section 2, and the details of our proofs will be given in Section 3.

2. Results

Our results are for a two-queue symmetric polling model with exhaustive service. Let λ be the (Poisson) arrival rate at each queue, let b be the average service time, and let $\rho \equiv \lambda b$

be the server utilization at each queue. We assume that the arrival processes, service times, switchover times, and setup times are all mutually independent. To guarantee stability, we also assume that the total server utilization, 2ρ, is less than 1.

Consider a server-departure epoch, equally likely to be from either queue. Let R be the ensuing switchover time, and let δ be the indicator function of the event that the server sets up when it polls the next queue at the expiration of R. In the SI model, we always have $\delta = 1$ and hence R and δ are independent. This, however, is not the case in the SD model, where $\delta = 1$ only if at least one customer is waiting at the next queue at the expiration of R. Define the *dead time* X as the sum

$$X = R + \delta Z\,, \tag{1}$$

where Z denotes the *potential* setup time that follows R. For both models, we assume that Z is independent of R and δ. We further assume that both R and Z are not identically zero.

Let $\tilde{R}$ be a switchover time during which (i.e. given that) no customers arrive at the queue to which the server is switching; then

$$E(\tilde{R}) = \frac{\displaystyle\int_0^\infty t\mathrm{e}^{-\lambda t}\,\mathrm{d}F_R(t)}{\displaystyle\int_0^\infty \mathrm{e}^{-\lambda t}\,\mathrm{d}F_R(t)}\,, \tag{2}$$

where $F_R(\cdot)$ is the distribution function of R. Let W^{D} denote the waiting time in the two-queue symmetric SD polling model, and let W^{I} be its counterpart in the corresponding SI model. In Section 3, we derive a formula for $E(W^{\mathrm{D}})$ and we combine that with a parallel formula for $E(W^{\mathrm{I}})$ to yield the following theorem (where $V(\cdot)$ denotes variance).

Theorem.

$$E(W^{\mathrm{D}}) - E(W^{\mathrm{I}}) = c\left\{[E(R) + E(Z)]\left[1 - \frac{E(\tilde{R})}{E(R)}\right] + \left[\frac{V(R)}{E(R)} - \frac{V(Z)}{E(Z)}\right]\right\}, \tag{3}$$

where c is a positive number (given explicitly in equation (24) in Section 3). We note that when $N = 1$, it can be shown (see Section 3) that formula (3) remains true with $E(\tilde{R})$ replaced by 0.

The theorem completely characterizes (for $N = 2$) the sign of $E(W^{\mathrm{D}}) - E(W^{\mathrm{I}})$ in terms of only the distribution of R and the first two moments of Z. In particular, it has the following easily verified consequences.

Corollary 1. *If the switchover times are constant and the setup times are constant, then*

$$E(W^{\mathrm{D}}) = E(W^{\mathrm{I}}). \tag{4}$$

The equality (4) follows from (3) because when R is a constant, then $V(R) = 0$ and $E(\tilde{R}) = E(R)$; and when Z is a constant, then $V(Z) = 0$.

Corollary 2. *If the switchover times are variable and the setup times are constant, then*

$$E(W^{\mathrm{D}}) > E(W^{\mathrm{I}}). \tag{5}$$

The inequality (5) follows from (3) because $c > 0$, $V(R) > 0$, $V(Z) = 0$, and, as is intuitively clear (and will be proved in Section 3),

$$E(\tilde{R}) < E(R). \tag{6}$$

Corollary 3. *If the switchover times are constant and the setup times are variable, then*

$$E(W^{\mathrm{D}}) < E(W^{\mathrm{I}}). \tag{7}$$

The inequality (7) follows from (3) because $c > 0$, $V(R) = 0$, $V(Z) > 0$, and $E(\tilde{R}) = E(R)$.

It is remarkable that Corollaries 1, 2, and 3 depend only on whether the switchover times and setup times are constant or variable, and do not depend in any way on the server utilization or the service-time distribution. Furthermore, only the inequality (7) agrees with common-sense intuition. The explicit formula (3), together with these three corollaries, clearly demonstrates that the relative variance-to-mean ratios of the switchover times and of the setup times, that is, the difference

$$\frac{V(R)}{E(R)} - \frac{V(Z)}{E(Z)},$$

is the primary determinant of the sign of $E(W^{\mathrm{D}}) - E(W^{\mathrm{I}})$. We believe that this observation offers a useful qualitative insight for more general polling models as well.

3. Proofs

A very insightful analysis of the symmetric SI model (without setup times) is given in Fuhrmann (1985). The key idea in that paper is to define *every* switchover time following a server-departure epoch as a server 'vacation' (in contrast with the standard approach of defining the time interval during which the server is away from a fixed 'reference queue' as a vacation). Since waiting customers at different queues are stochastically indistinguishable, this definition of vacation allows one first to aggregate the customer counts at individual queues and then to analyse the resulting *total* number of customers in the system, using a decomposition result for generalized vacation models obtained by Fuhrmann and Cooper (1985), Proposition 2, p. 1123. Our analysis of the symmetric SD model will also be based on this idea; that is, we will consider each dead time, as defined in (1), as a vacation.

We note that in the symmetric SI model, the duration of a vacation is independent of the queue length in the system immediately after the server departs from a queue. The main complication in adapting Fuhrmann's argument to the SD model is that this independence does not hold in the latter model (since the term δ in (1) is dependent on the queue length after the preceding server-departure epoch).

We focus first on the SD model, and let L^{D} be the total number of customers in the system left behind by a randomly selected departing customer in this model. Consider a randomly selected customer who arrives during a vacation, and call this customer the 'tagged customer'. Let K^{D} be the total number of customers in the system as seen by the tagged customer. Then, from Proposition 3 of Fuhrmann and Cooper (1985), equation (4), p. 1125, we have ($\stackrel{\mathrm{d}}{=}$ denotes equality in distribution)

$$L^{\mathrm{D}} \stackrel{\mathrm{d}}{=} K^{\mathrm{D}} + L^{\mathrm{M}}, \tag{8}$$

where L^{M} is independent of K^{D} and is distributed as the total number of customers left behind by a randomly selected departing customer in a (corresponding) standard $M/G/1$ queue.

It is well known (Burke (1968) and Wolff (1982); see also Cooper (1981), p. 186) that $E(L^{\mathrm{D}})$, an average taken over departure epochs, also equals the time-average total number of customers in the system. Therefore, an application of Little's law yields (we now assume $N = 2$, unless explicitly stated otherwise)

$$E(L^{\mathrm{D}}) = 2\lambda\, E(W^{\mathrm{D}}). \tag{9}$$

Since relations parallel to (8) and (9) also hold for the SI model, we have (with similar notation for the SI model)

$$E(W^{\mathrm{D}}) - E(W^{\mathrm{I}}) = \frac{1}{2\lambda}\{E(K^{\mathrm{D}}) - E(K^{\mathrm{I}})\}. \tag{10}$$

We now show how $E(K^{\mathrm{D}})$ is computed; results for $E(K^{\mathrm{I}})$ will follow similarly.

Consider a server-departure epoch. Let T be the total number of waiting customers at both queues at this epoch, and let U be the total number of customers that arrive to the entire system during the ensuing vacation. The decomposition (8) was proved (see Fuhrmann and Cooper (1985)) under the assumption that the queue discipline for the *entire* system is non-preemptive LIFO (last-in first-out); therefore the total customer count T can, depending on the choice of queue discipline, be split between the two queues. With exhaustive service, however, there are by definition no customers remaining in the queue from which the server is switching; hence, T is distributed as the number of waiting customers at the next queue. We also note that, in light of (1), T and U are dependent random variables.

Next, observe that $E(K^{\mathrm{D}})$ equals the sum of (i) the expected number of waiting customers in the system at the preceding server-departure epoch as 'observed' by the tagged customer, and (ii) the expected number of customers who arrive at *both* queues prior to the tagged customer but within the same vacation. In Wolff (1989), equation (76), p. 460, it is shown that

$$E(K^{\mathrm{D}}) = \frac{E(TU)}{E(U)} + \frac{E[U(U-1)]}{2\,E(U)}, \tag{11}$$

where the two right-hand-side terms correspond, respectively, to (i) and (ii), and both terms have been corrected for length bias. To understand the first term in (11), one can apply a standard 'renewal-reward' argument as follows. Interpret U as the size of a 'customer batch', and T as the 'reward' given to every customer in the batch. Then, TU equals the total reward in a batch; and therefore, $E(TU)/E(U)$ gives the expected reward received by a customer (the tagged customer) that is selected randomly from an infinite sequence (e.g. stationary and ergodic) of customer batches. A similar argument yields the second term in (11) (see, for example, Section 2–5, p. 68, of Wolff (1989)).

Let A_R and A_Z be the *total* numbers of Poisson arrivals (at rate 2λ) during an R and a Z, respectively. Then, in parallel with (1), we have

$$U = A_R + \delta A_Z\,, \tag{12}$$

where A_Z is independent of both A_R and δ (whereas A_R and δ are, of course, dependent).

Upon substitution of (12), it is easily shown, using the identities $T\delta = T$ (with $N = 2$, we have $\delta = 1$ whenever $T > 0$) and $\delta A_Z(\delta A_Z - 1) = \delta A_Z(A_Z - 1)$, that

$$\frac{E(TU)}{E(U)} = \frac{E(T)[E(A_R) + E(A_Z)]}{E(A_R) + E(\delta)E(A_Z)}, \tag{13}$$

and

$$\frac{E[U(U-1)]}{2\,E(U)} = \frac{E[A_R(A_R-1)] + E(\delta)E[A_Z(A_Z-1)] + 2E(\delta A_R)E(A_Z)}{2[E(A_R) + E(\delta)E(A_Z)]}. \tag{14}$$

We next consider $E(T)$. Denote by C^{D} the duration of a cycle, defined as the time elapsed between two successive server-departure epochs at a queue. It is well known (Kuehn (1979)) that, in general and for arbitrary N,

$$E(C^{\mathrm{D}}) = \frac{\sum_{i=1}^{N} E(R_i + \delta_i Z_i)}{1 - \sum_{i=1}^{N} \rho_i},$$

where the subscript i in ρ_i, R_i, δ_i, and Z_i refers to queue i. Hence, for the symmetric two-queue model, we have

$$E(C^{\mathrm{D}}) = \frac{E(A_R) + (1 - \alpha_0)E(A_Z)}{\lambda(1 - 2\rho)},$$

where $\alpha_0 \equiv 1 - E(\delta)$, the probability for the server to find the next queue empty at the expiration of a switchover time. Now, observe that at any server-departure epoch, T is distributed as the number of arrivals at the next queue during the preceding 'half' cycle. Therefore,

$$E(T) = \lambda \frac{E(C^{\mathrm{D}})}{2} = \frac{E(A_R) + (1 - \alpha_0)E(A_Z)}{2(1 - 2\rho)}. \tag{15}$$

To evaluate $E(\delta A_R)$ in (14), note that the variable A_R can be broken down as the sum of the arrival counts at the two queues. That is, let $\hat{A}_R$ be the number of arrivals at the next queue during an R; then $A_R = (A_R - \hat{A}_R) + \hat{A}_R$. It is important to note that $A_R - \hat{A}_R$ and $\hat{A}_R$ are dependent in general, since they correspond to arrivals during the *same* R; they are independent if and only if R is a constant. Let δ_T be the indicator function of the event $\{T > 0\}$; and let δ_R be the indicator function of the event $\{\hat{A}_R > 0\}$. Since $\delta = 1$ if and only if $T + \hat{A}_R > 0$, we have

$$\delta = 1 - (1 - \delta_T)(1 - \delta_R). \tag{16}$$

Upon substitution of (16), we have

$$E(\delta A_R) = E(A_R) - E[(1 - \delta_T)(1 - \delta_R)A_R]. \tag{17}$$

Observe that: (i) $(1-\delta_T)(1-\delta_R) = 1$ if and only if $\delta_T = 0$ and $\delta_R = 0$; (ii) the event $\{\delta_T = 0\}$ is independent of A_R; and (iii) the event $\{\delta_R = 0\}$ is equivalent to the event $\{\hat{A}_R = 0\}$. It follows that

$$\begin{aligned} E[(1 - \delta_T)(1 - \delta_R)A_R] &= P\{(1 - \delta_T)(1 - \delta_R) = 1\}E[A_R \mid (1 - \delta_T)(1 - \delta_R) = 1] \\ &= \alpha_0 \, E(A_R \mid \hat{A}_R = 0). \end{aligned} \tag{18}$$

Since A_R and $\hat{A}_R$ relate to the same R, therefore

$$E(A_R \mid \hat{A}_R = 0) = \frac{E(\lambda R \,\mathrm{e}^{-\lambda R})}{E(\mathrm{e}^{-\lambda R})} = \lambda E(\tilde{R}), \tag{19}$$

where $\tilde{R}$ is a switchover time during which no customers arrive at the target queue (see (2)). Hence, by combining (17), (18), and (19), we obtain

$$E(\delta A_R) = E(A_R) - \alpha_0 \lambda E(\tilde{R}). \tag{20}$$

With (13), (14), (15), and (20) in (11), we finally obtain

$$E(K^{\mathrm{D}}) = \frac{E(A_R) + E(A_Z)}{2(1-2\rho)} + \frac{E[A_R(A_R-1)] + (1-\alpha_0)E[A_Z(A_Z-1)] + 2[E(A_R) - \alpha_0\lambda E(\tilde{R})]E(A_Z)}{2[E(A_R) + (1-\alpha_0)E(A_Z)]}. \quad (21)$$

Furthermore, a similar calculation starting with $\delta \equiv 1$ in (13) and (14) also yields, in the corresponding SI model,

$$E(K^{\mathrm{I}}) = \frac{E(A_R) + E(A_Z)}{2(1-2\rho)} + \frac{E[A_R(A_R-1)] + E[A_Z(A_Z-1)] + 2E(A_R)E(A_Z)}{2[E(A_R) + E(A_Z)]}. \quad (22)$$

After some algebra, (21) and (22) lead to

$$E(K^{\mathrm{D}}) - E(K^{\mathrm{I}}) = \frac{\alpha_0 E(A_R)E(A_Z)}{2[E(A_R) + (1-\alpha_0)E(A_Z)][E(A_R) + E(A_Z)]} \times \left\{[E(A_R) + E(A_Z)]\left[1 - \frac{2\lambda E(\tilde{R})}{E(A_R)}\right] + \left[\frac{V(A_R)}{E(A_R)} - \frac{V(A_Z)}{E(A_Z)}\right]\right\}. \quad (23)$$

Next, note that $E(A_R) = 2\lambda E(R)$ and $V(A_R) = 2\lambda E(R) + 4\lambda^2 V(R)$. Upon substitution of these and similar formulas for $E(A_Z)$ and $V(A_Z)$, it is easily shown that (10) and (23) together lead to (3), with

$$c \equiv \frac{\alpha_0 E(R)E(Z)}{2[E(R) + (1-\alpha_0)E(Z)][E(R) + E(Z)]}; \quad (24)$$

this completes our proof of the Theorem.

Our proof can also be specialized to yield similar results if $N = 1$. First, replace 2λ by λ in (9) (and hence (10)). Next, since $N = 1$, we have $T \equiv 0$. It follows that the first term in (11) equals 0. Furthermore, an ensuing setup time occurs only if there is at least one arrival during a switchover time; that is, $\delta = 1$ only if $A_R > 0$. Therefore, $E(\delta A_R) = E(A_R)$, and we have (from (11) and (14))

$$E(K^{\mathrm{D}}) = \frac{E[A_R(A_R-1)] + E(\delta)E[A_Z(A_Z-1)] + 2E(A_R)E(A_Z)}{2[E(A_R) + E(\delta)E(A_Z)]}.$$

Since $\delta \equiv 1$ for the SI model, we also have

$$E(K^{\mathrm{I}}) = \frac{E[A_R(A_R-1)] + E[A_Z(A_Z-1)] + 2E(A_R)E(A_Z)}{2[E(A_R) + E(A_Z)]}.$$

It is now easily shown that the Theorem remains valid with $E(\tilde{R})$ replaced by 0. Note that the probability α_0 in (24) is, in this case, explicitly given by $E(\mathrm{e}^{-\lambda R})$ (the evaluation of α_0 is difficult if $N = 2$).

Finally, we consider (6). Intuitively, the fact that there is no arrival at the next queue during a randomly selected R (i.e. $\hat{A}_R = 0$) implies that we must have sampled a 'short' R. Indeed, we will now show that, for $t \geq 0$,

$$P\{R > t \mid \hat{A}_R = 0\} \leq P\{R > t\} \quad (25)$$

(that is, $(R \mid \hat{A}_R = 0)$ is stochastically smaller than R), which immediately yields $E(\tilde{R}) \leq E(R)$. To prove (25), let Y be a random variable, independent of R; and consider the inequality

$$P\{R > t,\ R > Y\} \geq P\{R > t\}P\{R > Y\}, \quad t \geq 0. \tag{26}$$

This clearly is true if Y is deterministic; and therefore, the general case follows immediately by conditioning on Y. If Y is exponentially distributed with rate λ, it is easily seen that the events $\{R > t,\ R < Y\}$ and $\{R > t,\ \hat{A}_R = 0\}$ are equivalent. Since

$$P\{R > t,\ R < Y\} = P\{R > t\} - P\{R > t,\ R > Y\},$$

we see that (25) is an easy consequence of (26). The strict inequality (6) now follows since (26) holds as an equality if and only if R is a constant.

References

[1] ALTMAN, E., BLANC, H., KHAMISY, A. AND YECHIALI, U. (1994). Gated-type polling systems with walking and switch-in times. *Stochastic Models* **10**, 741–763.

[2] BRADLOW, H. S. AND BYRD, H. F. (1987). Mean waiting time evaluation of packet switches for centrally controlled PBX's. *Perf. Eval.* **7**, 309–327.

[3] BURKE, P. J. (1968). A recurrence relation for stationary birth-and-death processes with an application to finite-source traffic systems. Bell Telephone Laboratories Technical Memorandum.

[4] COOPER, R. B. (1981). *Introduction to Queueing Theory*, 2nd edn. North-Holland, New York. Available from University Microfilms International, Ann Arbor, MI.

[5] COOPER, R. B., NIU, S.-C. AND SRINIVASAN, M. M. (1996). A decomposition theorem for polling models: the switchover times are effectively additive. *Operat. Res.* **44**, 629–633.

[6] COOPER, R. B., NIU, S.-C. AND SRINIVASAN, M. M. (1998). When does forced idle time improve performance in polling models? *Management Sci.* **44**, 1079–1086.

[7] EISENBERG, M. (1995). The polling system with changeover times that depend on the occupancy of the visited queue. *Proc. 3rd INFORMS Telecommunications Conference*. INFORMS, Baltimore, MD.

[8] FERGUSON, M. J. (1986). Mean waiting time for a token ring with station-dependent overheads. In *Local Area and Multiple Access Networks*, ed. R. L. Pickholtz. Computer Science Press, Rockville, Md, pp. 43–67.

[9] FUHRMANN, S. W. (1985). Symmetric queues served in cyclic order. *Operat. Res. Lett.* **4**, 139–144.

[10] FUHRMANN, S. W. (1992). A decomposition result for a class of polling models. *Queueing Systems* **11**, 109–120.

[11] FUHRMANN, S. W. AND COOPER, R. B. (1985). Stochastic decompositions in the $M/G/1$ queue with generalized vacations. *Operat. Res.* **33**, 1117–1129.

[12] GÜNALAY, Y. (1995). State dependent server scheduling rules in polling systems. Ph.D. Dissertation, McMaster University, Hamilton, CA.

[13] GÜNALAY, Y. AND GUPTA, D. (1997). Polling systems with a patient server and state-dependent setup times. *IIE Transactions* **29**, 469–480.

[14] GUPTA, D. AND SRINIVASAN, M. M. (1996). Polling systems with state-dependent setup times. *Queueing Systems* **22**, 403–423.

[15] KUEHN, P. J. (1979). Multiqueue systems with nonexhaustive cyclic service. *The Bell System Technical Journal* **58**, 671–698.

[16] LENNON, T. (1995). Response-time approximations for multi-server, multi-class production systems with significant setups. Technical Report, The University of Michigan, Ann Arbor, MI. Revised, 1995.

[17] OLSEN, T. L. (1996). Approximations for the waiting time distribution in polling models with and without state-dependent setups. Technical Report, The University of Michigan, Ann Arbor, MI.

[18] SARKAR, D. AND ZANGWILL, W. I. (1991). Variance effects in cyclic production systems. *Management Sci.* **37**, 443–453.

[19] SRINIVASAN, M. M., NIU, S.-C. AND COOPER, R. B. (1995). Relating polling models with zero and nonzero switchover times. *Queueing Systems* **19**, 149–168.

[20] WOLFF, R. W. (1982). Poisson arrivals see time averages. *Operat. Res.*, **30**, 223–231.

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