Polling Models with and without Switchover Times

S.C. Borst, O.J. Boxma

CWI

P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

Abstract

We consider two different single-server cyclic polling models: (i) a model with zero switchover times, and (ii) a model with non-zero switchover times, in which the server keeps cycling when the system is empty. For both models we relate the steady-state queue length distribution at a queue to the queue length distributions at server visit beginning and visit completion instants at that queue; as a by-product we obtain a short proof of the Fuhrmann-Cooper decomposition. For the large class of polling systems that allow a multitype branching process interpretation, we expose a strong relation between the queue length, as well as waiting-time, distributions in the two models. The results enable a very efficient numerical computation of the waiting-time moments under different switchover time scenarios.

Consider a single-server cyclic polling system: a single server S visits n queues $Q_1, \ldots, Q_n$ in cyclic order, serving customers at Q_i according to some service discipline D_i, $i = 1, \ldots, n$. Such polling models have many applications in, e.g., computer-communications, manufacturing and traffic. The abundant literature on polling systems (cf. Takagi 1990) contains the exact analysis of polling systems for a large number of service disciplines, like the exhaustive and gated disciplines.

Polling models are closely related to queueing models with vacations of the server; the intervisit time of the server w.r.t. some Q_i can be viewed as a vacation. Accordingly, the concept of queue length decomposition for queues with vacations (cf. Fuhrmann and Cooper 1985) has proven to be very fruitful for the analysis of polling models. It has also led to the concept of work decomposition in polling models (cf. Boxma 1989), which relates the amount of work in a polling system *with* switchover times to the amount of work in the same polling system but *without* switchover times. Hitherto, polling models with switchover times and polling models without switchover times had usually been treated separately, often via different approaches. The problem with simply letting the switchover times tend to zero in a polling model with non-zero switchover times is that the number of polling epochs in an idle period tends to infinity, leading to degenerate distributions at such epochs (cf. Eisenberg 1994, Levy and Kleinrock 1991).

The relation between polling models with and without switchover times has further been exposed in some recent papers of Cooper, Niu and Srinivasan (1992), Fuhrmann (1992) and Srinivasan, Niu and Cooper (1994). The authors of the latter paper obtain the marginal queue length distributions and waiting-time distributions for the case of zero switchover times, using the so-called descendant set approach (cf. Konheim, Levy and Srinivasan (1994)). Quoting results from Konheim et al. (1994) for nonzero switchover times, they observe the similarity between both cases and they indicate that the set of 'descendant customers' at certain reference points provides the underlying link between those cases. Further Srinivasan et al. (1994) show how the similarity can be exploited to derive simple relations that express the waiting-time moments for non-zero switchover times in terms of those for zero switchover times.

The present paper was inspired by a preliminary version of Srinivasan et al. (1994), and its main goal is to interpret, unify, and expand their results. The latter paper mainly focuses on exhaustive and gated service, but indicates that a generalization to a larger class of service disciplines should be possible. Throughout the present paper we work with the general class of service disciplines with the so-called 'branching property' (see below), thus exposing the general structure as well as the limitations of the theory. Using a branching process approach, cf. Resing (1993) and Fuhrmann (1981), we demonstrate why, where, and how the similarities and differences between the cases with and without switchover times arise. In particular we relate the *similarity* between both cases to the similarity in the *offspring* generating functions of the underlying branching processes. Furthermore, we express the *differences* into the differences between the underlying *immigration* functions. The present paper also shows that the fundamental relationship between both cases occurs at the level of the *joint* queue length distribution, and that the relationship at the level of marginal queue lengths and waiting times, as obtained in Srinivasan et al. (1994), is a derivative.

The remainder of the paper is organized as follows. Firstly, in Section 1, we use a beautiful relation of Eisenberg (1972) (see also Eisenberg 1994), which has received too little attention in the literature, to relate the Generating Functions (GF) of queue lengths at various instants in the polling system (visit beginnings and completions, service beginnings and completions). We observe that this relation, which was presented by Eisenberg (1972) for the case of non-zero switchover times, also holds for the case of zero switchover times; as a by-product this yields a short proof of the above-mentioned Fuhrmann-Cooper decomposition for the queue lengths at the various queues of a polling system.

Eisenberg's relation leads to an expression for the joint queue length GF at service completion instants at some queue into the joint queue length GF's at the beginning and end of a visit of the server to that queue. The latter GF's can be easily related, and determined, for the important class of polling models in which each queue satisfies the following 'branching property' (cf. Fuhrmann 1992):

Property 0.1

If S arrives at Q_i to find k_i customers there, then during the course of the server's visit, each of these k_i customers will effectively be replaced in an i.i.d. manner by a random population having probability generating function $h_i(z_1, \ldots, z_n)$, which can be any n-dimensional probability generating function.

Resing (1993) (see also Fuhrmann 1981) has shown that, for the class of polling systems that satisfy this property, the joint queue length process at visit instants of a fixed queue is a so-called *multitype branching process* (MTBP) with immigration. The theory of multitype branching processes (cf. Athreya and Ney 1972, Resing 1990) subsequently leads to an expression for the generating function of the joint queue length process at visit beginning (polling) instants.

In Section 2, for models that satisfy Property 0.1 (note that this class includes for example exhaustive and gated but not 1-limited), we use a slightly adapted version of the results of Resing (1993) to relate the joint queue length GF's at visit beginning and visit completion instants, and then to obtain those GF's. The results expose a close similarity between the cases with and without switchover times. In Section 3 we determine the steady-state marginal queue length GF at Q_i, both for the model with and the model without switchover times, and we relate the GF's for those two cases; similarly for the waiting-time Laplace-Stieltjes Transform (LST) at Q_i. In Section 4 we describe how the results can be exploited for a very efficient numerical calculation of the waiting-time moments under different switchover time scenarios.

1 A general polling model

We study the following polling model. A single server S visits n infinite-buffer queues $Q_1, \dots, Q_n$ in cyclic order $Q_1, \dots, Q_n, Q_1, \dots$. Customers arrive at Q_i according to a Poisson process with rate λ_i; $\lambda := \sum_{i=1}^{n} \lambda_i$. A customer at Q_i, to be called a type-i customer, requires a service time $\mathbf{B}_i$ which is a random variable with distribution $B_i(\cdot)$ with mean β_i and LST $\beta_i(\cdot)$. Define $\rho_i := \lambda_i \beta_i$ as the traffic intensity at Q_i. Denote by $\rho := \sum_{i=1}^{n} \rho_i$ the total traffic intensity.

When S moves from Q_i to Q_{i+1}, this requires a switchover time $\mathbf{S}_i$ which is a random variable with distribution $S_i(\cdot)$ with mean s_i and LST $\sigma_i(\cdot)$. In this model with switchover times, S switches even when the system is empty.

We also consider the variant of this model in which all switchover times are *zero*. In the latter model, when the system becomes empty S makes a full cycle (viz. passes all the queues once) and subsequently stops right before Q_1. All this requires zero time. When the first new customer arrives, S cycles along the queues to that customer. Note that the behavior of the server when the system is empty does *not* affect the waiting-time and queue length distributions. However, it may involve a number of (zero-length) visits and hence it *does* affect the queue length distribution embedded at visit beginning and completion instants. Our convention that the server stops at some fixed queue turns out to result in cleaner expressions for that embedded queue length distribution than when the server is assumed to stop right away. Making a full cycle before stopping may be necessary if there is no central queue length information available to the server so that the only way to detect an empty system is to keep track if some fixed queue is visited twice in zero time.

We assume that all the usual independence assumptions hold between the service times, the switchover times and the interarrival times. We assume that the ergodicity conditions are fulfilled, and we restrict ourselves to results for the stationary situation.

Eisenberg (1972) studies this model for the case with switchover times and with the exhaustive service discipline at all queues (while briefly discussing the case of gated service at all queues). He makes the crucial observation that each time a visit beginning or a service completion occurs, this coincides with either a service beginning or a visit completion. Note however that this observation not only holds for the case with switchover times and exhaustive or gated service, but for any service discipline, and also for the case of zero switchover times. After some manipulations this yields (for details we refer to the report version of this paper, Borst and Boxma (1994)):

$$\gamma_i V_{b_i}(z) + S_{c_i}(z) = S_{b_i}(z) + \gamma_i V_{c_i}(z), \tag{1}$$

for $z = (z_1, \ldots, z_n)$, $|\, z_j \,| \leq 1$, $j = 1, \ldots, n$; here $V_{b_i}(z)$ and $V_{c_i}(z)$ denote the GF of the joint queue length distribution at visit beginnings and visit completions at Q_i, while $S_{b_i}(z)$ and $S_{c_i}(z)$ denote the GF of the joint queue length distribution at service beginnings and service completions at Q_i; the coefficient γ_i represents the reciprocal of the mean number of customers served at Q_i per visit.

Subsequently Eisenberg observes that $S_{c_i}(z)$ and $S_{b_i}(z)$ are further related via

$$S_{c_i}(z) = S_{b_i}(z)\beta_i(\sum_{j=1}^{n} \lambda_j(1 - z_j))/z_i, \tag{2}$$

for $|\, z_j \,| \leq 1$, $j = 1, \ldots, n$. It follows from (1) and (2) that

$$S_{c_i}(z) = \frac{\gamma_i \beta_i(\sum_{j=1}^{n} \lambda_j(1 - z_j))}{z_i - \beta_i(\sum_{j=1}^{n} \lambda_j(1 - z_j))}[V_{b_i}(z) - V_{c_i}(z)]. \tag{3}$$

Eisenberg, considering the variant with switchover times and exhaustive service, subsequently expresses $V_{b_i}(z)$ in $V_{c_{i-1}}(z)$. For the moment we refrain from that (see Section 2), but we observe that formula (3) is generally valid for the class of polling systems described in the beginning of this section (with and without switchover times).

Taking in (3) $z = (1, \ldots, 1, y, 1, \ldots, 1)$, with y as i-th argument, gives the queue length GF at Q_i at a service completion instant at Q_i. A standard up- and down-crossing argument combined with PASTA shows that the queue length distribution at Q_i at its service completion instants, at its customer arrival instants and in steady state are all the same. Hence, with $\mathbf{N}_i$ the steady-state queue length at Q_i and with $\mathbf{X}_i$ and $\mathbf{Y}_i$ the steady-state queue lengths at Q_i at the beginning and end of a visit at that queue: for $|y| \leq 1$,

$$\mathrm{E}(y^{\mathbf{N}_i}) = \frac{\gamma_i \beta_i(\lambda_i(1-y))}{y - \beta_i(\lambda_i(1-y))}[\mathrm{E}(y^{\mathbf{X}_i}) - \mathrm{E}(y^{\mathbf{Y}_i})]. \tag{4}$$

Remember that $1/\gamma_i$ equals the mean number of customers served at Q_i per visit, hence also the mean number of customers that arrive at Q_i per cycle. Because S spends on the average a fraction ρ_i of a cycle at Q_i, it follows that a fraction $(1-\rho_i)$ of the customers arrive at Q_i during the intervisit period. Now observe that $\mathbf{Y}_i$ not only denotes the queue length at Q_i at the end of a visit to that queue, but also the queue length at Q_i at the beginning of an intervisit period for that queue; similarly $\mathbf{X}_i$ denotes the queue length at Q_i at the *end* of such an intervisit period. Thus we can write:

$$\mathrm{E}\mathbf{X}_i - \mathrm{E}\mathbf{Y}_i = \frac{1-\rho_i}{\gamma_i}. \tag{5}$$

From (4) and (5), for $|y| \leq 1$:

$$\mathrm{E}(y^{\mathbf{N}_i}) = \frac{(1-\rho_i)(1-y)\beta_i(\lambda_i(1-y))}{\beta_i(\lambda_i(1-y)) - y} \frac{\mathrm{E}(y^{\mathbf{Y}_i}) - \mathrm{E}(y^{\mathbf{X}_i})}{(1-y)(\mathrm{E}\mathbf{X}_i - \mathrm{E}\mathbf{Y}_i)}. \tag{6}$$

The first term in the rhs is the GF $\mathrm{E}(y^{\mathbf{N}_{i|M/G/1}})$ of the queue length distribution in a 'corresponding' isolated $M/G/1$ queue of Q_i with arrival rate λ_i and service time distribution $B_i(\cdot)$. Now consider the second term. Introduce $\mathbf{Z}_i$, a stochastic variable with distribution the queue length distribution at an arbitrary instant in an intervisit period of Q_i. With the interpretation of $\mathbf{X}_i$ and $\mathbf{Y}_i$ as given above (5) we can then write (cf. Borst (1994), and also Wolff (1989), Section 10.5 for a similar result):

$$\mathrm{E}(y^{\mathbf{Z}_i}) = \frac{\mathrm{E}(y^{\mathbf{Y}_i}) - \mathrm{E}(y^{\mathbf{X}_i})}{(1-y)(\mathrm{E}\mathbf{X}_i - \mathrm{E}\mathbf{Y}_i)}. \tag{7}$$

This relation appears in the polling literature for various special cases (e.g., for exhaustive vacation models, where $\mathbf{Y}_i \equiv 0$). It holds also for non-Poisson arrivals, when $\mathbf{Z}_i$ is defined as the queue length observed by an arbitrary customer arriving at Q_i during an intervisit period.

Formulas (6) and (7) together yield the well-known Fuhrmann-Cooper queue length decomposition applied to a queue in a polling model with or without switchover times: for $|y| \leq 1$,

$$\mathrm{E}(y^{\mathbf{N}_i}) = \mathrm{E}(y^{\mathbf{N}_{i|M/G/1}})\mathrm{E}(y^{\mathbf{Z}_i}). \tag{8}$$

Remark 1.1 Fuhrmann and Cooper (1985) state five conditions under which their decomposition holds:

(i). Customers arrive at Q_i according to a Poisson process of rate λ_i.

(ii). All customers arriving at Q_i are eventually served.

(iii). Customers enter service in an order that is independent of their service times.

(iv). Service is non-preemptive.

(v). The rules that govern when the server begins and ends visit periods to Q_i do not anticipate future jumps of the Poisson arrival process at Q_i.

These assumptions indeed hold in our polling model, and are implicitly used in the derivation of (8). The above proof, with as key steps (1), (2) and (7), in fact also holds for general vacation models without a specific polling context. Note that the relation (5) holds generally for queues with some vacation (intervisit) mechanism. Clearly, when one were just interested in the derivation of (8), one could consider in equations (1)-(3) only the marginal queue length at Q_i rather than the joint queue length. We refer to Keilson and Servi (1990) for another short proof of the Fuhrmann-Cooper decomposition.

□

Assuming that within each queue customers are served in order of arrival, application of the distributional form of Little's law, cf. Keilson and Servi (1990), both to Q_i and the 'corresponding' isolated $M/G/1$ queue of Q_i, yields for $|y| \leq 1$:

$$\mathrm{E}(\mathrm{e}^{-\omega \mathbf{W}_i}) = \mathrm{E}(\mathrm{e}^{-\omega \mathbf{W}_{i|M/G/1}})\mathrm{E}((1-\omega/\lambda_i)^{\mathbf{Z}_i}). \tag{9}$$

In Section 3 we shall return to this relation, for the case of polling models that satisfy Property 0.1.

2 The joint queue length distribution at polling epochs

In the previous section we have seen that Eisenberg's results yield simple relations between the generating function $S_{c_i}(z)$ of the joint queue length vector at service completion epochs (or $S_{b_i}(z)$, at service beginning epochs) and the generating functions $V_{b_i}(z)$ and $V_{c_i}(z)$ of the joint queue length vector at visit beginning and visit completion epochs. We now restrict ourselves to polling models for which the service discipline at each queue satisfies Property 0.1. Property 0.1 prescribes how each of the customers present at Q_i at the visit beginning is replaced by independent families of customers at the visit completion. This enables one to relate $V_{c_i}(\cdot)$ nicely to $V_{b_i}(\cdot)$, and to finally determine each of the functions $V_{b_i}(\cdot)$ (after which the GF's $V_{c_i}(\cdot)$, $S_{c_i}(\cdot)$ and $S_{b_i}(\cdot)$ follow). In our analysis we follow Resing (1993).

First a remark about the ergodicity conditions. In the sequel we assume that $\rho < 1$ and $s_i < \infty$ for all i. Resing (1993) proves that for the subclass of so-called Bernoulli-type service disciplines these conditions together constitute sufficient ergodicity conditions. His proof is based on the observation that for these Bernoulli-type service disciplines the derivatives of $h_i(z_1, \ldots, z_n)$ take the form $\frac{\partial h_i(z)}{\partial z_j} |_{z=1} = \lambda_j \alpha_i \frac{\beta_i}{1-\rho_i}$, $i \neq j$, $\frac{\partial h_i(z)}{\partial z_i} |_{z=1} = 1 - \alpha_i$, with α_i some coefficient in $(0, 1]$ determined by the parameters of Q_i. It may be easily verified however that the latter form of the derivatives applies for *any* non-idling service discipline that satisfies Property 0.1 with $h_i(z_1, \ldots, z_n) \neq z_i$. As the proof in Resing (1993) further does not rely on the specific form of α_i, we may conclude that $\rho < 1$ and $s_i < \infty$ for all i together constitute sufficient ergodicity conditions for *any* non-idling service discipline that satisfies Property 0.1 with $h_i(z_1, \ldots, z_n) \neq z_i$.

Property 0.1 implies that

$$V_{c_i}(z) = V_{b_i}(z_1, \ldots, z_{i-1}, h_i(z), z_{i+1}, \ldots, z_n). \tag{10}$$

In the case of gated service $h_i(z)$ is simply the GF of the joint distribution of the numbers of arrivals at all queues during one service time at Q_i: $h_i(z) = \beta_i(\sum_{j=1}^{n} \lambda_j(1 - z_j))$. In the case of exhaustive service: $h_i(z) = \zeta_i(\sum_{j \neq i} \lambda_j(1 - z_j))$, with $\zeta_i(\cdot)$ the LST of the length of the busy period in a 'corresponding' isolated $M/G/1$ queue of Q_i.

Next we relate $V_{b_i}(z)$ to $V_{c_{i-1}}(z)$. In case of non-zero switchover times:

$$V_{b_i}(z) = V_{c_{i-1}}(z)\sigma_{i-1}(\sum_{j=1}^{n} \lambda_j(1 - z_j)). \tag{11}$$

In case of zero switchover times (in the sequel we add a superscript 0 for that case, to distinguish its quantities from those for non-zero switchover times):

$$V_{b_i}^0(z) = V_{c_{i-1}}^0(z), \tag{12}$$

for $i = 2, \dots, n$. The relation between $V_{b_1}^0(z)$ and $V_{c_n}^0(z)$ deserves special attention, because of our convention, stated at the beginning of Section 1, concerning the behavior of the server when the system is empty. When all queues in the model without switchover times become empty, S in our convention makes a full cycle and subsequently stops right before Q_1 (all this requires no time). When the first new customer arrives, S cycles along the queues to that customer. The consequence of this is that when the system is empty at the start of a visit to Q_1, then the next visit to Q_1 does not take place until a customer has arrived. We can write:

$$V_{b_1}^0(z) = V_{c_n}^0(z) - V_{b_1}^0(0)[1 - g^0(z)], \tag{13}$$

with

$$g^0(z) := \sum_{i=1}^{n} \frac{\lambda_i}{\lambda} z_i. \tag{14}$$

The function $g^0(\cdot)$ represents the 'immigration process' of the MTBP: it is the GF of the arrival process of customers during periods in which the system is empty.

Remark 2.1 In fact our convention slightly differs from that of Resing (1993), who assumes that when the system is empty S stops *at* Q_1 immediately, and hence takes $g^0(z) = \sum_{i=1}^{n} \frac{\lambda_i}{\lambda} f_i(z)$. His assumption would interfere with Eisenberg's approach, as a visit beginning at Q_1 in an empty system would not coincide with a service beginning or visit completion.

$\square$

Substituting (10) into (11), respectively (12) and (13), we can relate $V_{b_i}(\cdot)$ to $V_{b_{i-1}}(\cdot)$. We distinguish between the two cases of non-zero and zero switchover times. In both cases the following branching functions play a crucial role, thus establishing the link between both cases.

Define

$$f(z) := (f_1(z), \dots, f_n(z)), \tag{15}$$

with

$$f_i(z) := h_i(z_1, \dots, z_i, f_{i+1}(z), \dots, f_n(z)) \tag{16}$$

for $\mid z_j \mid \leq 1,\ j = 1, \dots, n$. This is the *offspring* GF, the GF of the joint distribution of the numbers of customers at the end of a cycle w.r.t. Q_1 that are *descendants* of a type-i customer. In this branching process setting, a descendant of some customer K is recursively defined as a customer that has arrived during the service time of K or of one of its descendants.

Define

$$f^{(0)}(z) := z,$$

$$f^{(k)}(z) := f(f^{(k-1)}(z)), \qquad k \geq 1,$$

for $|\, z_j \,| \leq 1$, $j = 1, \ldots, n$. Note that $f^{(k)}(\cdot)$ refers to the k-th generation offspring, with $f_i^{(k)}(y, 1, \ldots, 1)$ corresponding to the function $L_{i,k-1}(y)$ in Srinivasan et al. (1994).

Case I: Zero switchover times

Substituting (10) into (12),

$$V^0_{b_i}(z) = V^0_{b_{i-1}}(z_1, \ldots, z_{i-2}, h_{i-1}(z), z_i, \ldots, z_n) \tag{17}$$

for $i = 2, \ldots, n$. Starting from (13) and (10) for $i = n$, and subsequently using (17) for $i = n, n-1, \ldots, 2$, one obtains

$$V^0_{b_1}(z) = V^0_{b_1}(f(z)) - V^0_{b_1}(0)[1 - g^0(z)]. \tag{18}$$

Iterating (18) yields

$$V^0_{b_1}(z) = 1 - V^0_{b_1}(0) \sum_{k=0}^{\infty} [1 - g^0(f^{(k)}(z))], \tag{19}$$

with

$$V^0_{b_1}(0) = \left[1 + \sum_{k=0}^{\infty} [1 - g^0(f^{(k)}(0))]\right]^{-1}, \tag{20}$$

the infinite sum being convergent when the ergodicity conditions are fulfilled.

Introducing, for $|\, z_j \,| \leq 1$, $j = 1, \ldots, n$,

$$H(z) := \sum_{k=0}^{\infty} \sum_{i=1}^{n} \lambda_i (1 - f_i^{(k)}(z)), \tag{21}$$

we can write

$$V^0_{b_1}(z) = 1 - V^0_{b_1}(0) \sum_{k=0}^{\infty} \sum_{i=1}^{n} \frac{\lambda_i}{\lambda} (1 - f_i^{(k)}(z)) = 1 - V^0_{b_1}(0) \frac{H(z)}{\lambda}, \tag{22}$$

$$V_{b_1}^0(0) = \left[1 + \frac{H(0)}{\lambda}\right]^{-1}, \tag{23}$$

cf. equation (18) in Srinivasan et al. (1994) for a marginal distribution version.

Case II: Non-zero switchover times

Substituting (10) into (11),

$$V_{b_i}(z) = V_{b_{i-1}}(z_1, \ldots, z_{i-2}, h_{i-1}(z), z_i, \ldots, z_n)\sigma_{i-1}(\sum_{j=1}^{n} \lambda_j(1 - z_j)). \tag{24}$$

Applying (24) for $i = 1, \ldots, n$ (which corresponds to following the server during one full cycle w.r.t. Q_1),

$$V_{b_1}(z) = V_{b_1}(f(z))g(z), \tag{25}$$

with

$$g(z) = \prod_{i=1}^{n} \sigma_i(\sum_{j=1}^{i} \lambda_j(1 - z_j) + \sum_{j=i+1}^{n} \lambda_j(1 - f_j(z))). \tag{26}$$

The function $g(\cdot)$ represents the 'immigration process' of this MTBP: it is the GF of the vector of all customers that either have arrived in the switchover periods of the past cycle (measured w.r.t. Q_1), or are descendants of such customers. Iterating (25) yields

$$V_{b_1}(z) = \prod_{k=0}^{\infty} g(f^{(k)}(z)) = \prod_{k=0}^{\infty}\prod_{i=1}^{n} \sigma_i(\sum_{j=1}^{i} \lambda_j(1 - f_j^{(k)}(z)) + \sum_{j=i+1}^{n} \lambda_j(1 - f_j^{(k+1)}(z))), \tag{27}$$

the infinite product being convergent when the ergodicity conditions are fulfilled.

From (22) and (27) we see that $V_{b_1}(z)$ as well as $V^0_{b_1}(z)$ is determined by $\sum_{j=1}^{n} \lambda_j(1 - f_j^{(k)}(z))$. For *constant switchover times* the connection becomes even closer, as (27) in that case reduces to a simpler expression:

$$V_{b_1}(z) = \exp[-\sum_{k=0}^{\infty}\sum_{i=1}^{n} s_i\{\sum_{j=1}^{i} \lambda_j(1 - f_j^{(k)}(z)) + \sum_{j=i+1}^{n} \lambda_j(1 - f_j^{(k+1)}(z))\}]. \tag{28}$$

Using (21), and introducing the mean total switchover time $s := \sum_{i=1}^{n} s_i$, we can rewrite (28) into

$$V_{b_1}(z) = \exp[-sH(z) + \sum_{j=1}^{n} \lambda_j(1 - z_j)\sum_{i=1}^{j-1} s_i], \tag{29}$$

cf. equation (30) in Srinivasan et al. (1994) for a marginal distribution version.

3 Marginal queue lengths and waiting times

In the previous section the queue length GF's $V_{b_i}(z)$ and $V^0_{b_i}(z)$ at visit beginning instants at Q_i have been determined for the class of cyclic polling models in which Property 0.1 holds for all service disciplines. In Section 1 we already obtained a decomposition for the GF of the marginal queue length distribution at Q_i, and for the waiting-time LST at Q_i, into a corresponding $M/G/1$ term and a term involving $\mathrm{E}(y^{\mathbf{X}_i})$ and $\mathrm{E}(y^{\mathbf{Y}_i})$ (via the GF $\mathrm{E}(y^{\mathbf{Z}_i})$).

In particular, introducing

$\tilde{h}_i(y) := h_i(1, \dots, 1, y, 1, \dots, 1)$,

$\tilde{V}_{b_i}(y) := V_{b_i}(1, \dots, 1, y, 1, \dots, 1)$,

$\tilde{V}^0_{b_i}(y) := V^0_{b_i}(1, \dots, 1, y, 1, \dots, 1)$, with y as i-th argument, it follows from (7) and (10) for the case of non-zero switchover times that

$$\mathrm{E}(y^{\mathbf{Z}_i}) = \frac{\tilde{V}_{b_i}(\tilde{h}_i(y)) - \tilde{V}_{b_i}(y)}{(1-y)\tilde{V}'_{b_i}(1)(1-\tilde{h}'_i(1))}; \tag{30}$$

the same result holds for the case of zero switchover times, replacing $\tilde{V}_{b_i}(\cdot)$ by $\tilde{V}^0_{b_i}(\cdot)$ in (30). Similarly indicating queue lengths, and waiting times, by a superscript 0 in the case of zero switchover times, we find from (8) and (9):

$$\mathrm{E}(y^{\mathbf{N}_i}) = \mathrm{E}(y^{\mathbf{N}^0_i})\frac{[\tilde{V}_{b_i}(\tilde{h}_i(y)) - \tilde{V}_{b_i}(y)]\tilde{V}^{0\prime}_{b_i}(1)}{[\tilde{V}^0_{b_i}(\tilde{h}_i(y)) - \tilde{V}^0_{b_i}(y)]\tilde{V}'_{b_i}(1)}, \tag{31}$$

$$\mathrm{E}(\mathrm{e}^{-\omega\mathbf{W}_i}) = \mathrm{E}(\mathrm{e}^{-\omega\mathbf{W}^0_i})\frac{[\tilde{V}_{b_i}(\tilde{h}_i(1-\omega/\lambda_i)) - \tilde{V}_{b_i}(1-\omega/\lambda_i)]\tilde{V}^{0\prime}_{b_i}(1)}{[\tilde{V}^0_{b_i}(\tilde{h}_i(1-\omega/\lambda_i)) - \tilde{V}^0_{b_i}(1-\omega/\lambda_i)]\tilde{V}'_{b_i}(1)}. \tag{32}$$

Differentiating (9), (30), and (32) once, putting $\omega = 0$,

$$\mathrm{E}\mathbf{W}_i = \frac{\lambda_i\beta_i^{(2)}}{2(1-\lambda_i\beta_i)} + \frac{\tilde{V}''_{b_i}(1)}{2\tilde{V}'_{b_i}(1)\lambda_i}(1+\tilde{h}'_i(1)) - \frac{\tilde{h}''_i(1)}{2(1-\tilde{h}'_i(1))\lambda_i}, \tag{33}$$

and

$$\mathrm{E}\mathbf{W}_i - \mathrm{E}\mathbf{W}^0_i = [\frac{\tilde{V}''_{b_i}(1)}{2\tilde{V}'_{b_i}(1)\lambda_i} - \frac{\tilde{V}^{0\prime\prime}_{b_i}(1)}{2\tilde{V}^{0\prime}_{b_i}(1)\lambda_i}](1+\tilde{h}'_i(1)). \tag{34}$$

Let us now (without loss of generality) concentrate on $\mathbf{W}_1$ and $\mathbf{W}^0_1$. Introducing

$$\tilde{f}^{(k)}_i(y) := f^{(k)}_i(y,1,\ldots,1)$$

(the k-th generation type-1 customer offspring of a type-i customer, corresponding to the function $L_{i,k-1}(y)$ in Srinivasan et al. 1994), it follows from (22) that

$$\tilde{V}^0_{b_1}(y) = 1 - V^0_{b_1}(0)\tilde{H}(y)/\lambda, \tag{35}$$

with

$$\tilde{H}(y) := \sum_{k=0}^{\infty}\sum_{i=1}^{n}\lambda_i(1-\tilde{f}_i^{(k)}(y)). \tag{36}$$

In particular, $\tilde{V}_{b_1}^{0\,\prime}(1) = -V_{b_1}^0(0)\tilde{H}'(1)/\lambda$. It follows from (27) and (36) that $\tilde{V}_{b_1}'(1) = -s\tilde{H}'(1)$, and hence

$$\tilde{V}_{b_1}'(1) = \tilde{V}_{b_1}^{0\,\prime}(1)s\lambda/V_{b_1}^0(0). \tag{37}$$

From (32), (35), (37),

$$\mathrm{E}(\mathrm{e}^{-\omega\mathbf{W}_1}) = \mathrm{E}(\mathrm{e}^{-\omega\mathbf{W}_1^0})\frac{\tilde{V}_{b_1}(\tilde{h}_1(1-\omega/\lambda_1)) - \tilde{V}_{b_1}(1-\omega/\lambda_1)}{s[\tilde{H}(h_1(1-\omega/\lambda_1)) - \tilde{H}(1-\omega/\lambda_1)]}, \tag{38}$$

which for exhaustive service ($\tilde{h}_1(\cdot) \equiv 1$) and gated service ($\tilde{h}_1(1-\omega/\lambda_1) = \beta_1(\omega)$) corresponds to Theorems 2 and 5 in Srinivasan et al. (1994), respectively.

For constant switchover times, it follows from (29)

$$\tilde{V}_{b_1}(y) = \exp[-s\tilde{H}(y)], \tag{39}$$

and hence we have from (38)

$$\mathrm{E}(\mathrm{e}^{-\omega\mathbf{W}_1}) = \mathrm{E}(\mathrm{e}^{-\omega\mathbf{W}_1^0})\frac{\exp[-s\tilde{H}(\tilde{h}_1(1-\omega/\lambda_1))] - \exp[-s\tilde{H}(1-\omega/\lambda_1)]}{s[\tilde{H}(1-\omega/\lambda_1) - \tilde{H}(\tilde{h}_1(1-\omega/\lambda_1))]}, \tag{40}$$

which for exhaustive service ($\tilde{h}_1(\cdot) \equiv 1$) and gated service ($\tilde{h}_1(1-\omega/\lambda_1) = \beta_1(\omega)$) reduces to Theorems 3 and 6 in Srinivasan et al. (1994), respectively.

Remark 3.1 Substituting $\sigma_i(\omega) = 1 - s_i\omega + o(s_i)$ for $s_i \to 0$, $i = 1,\ldots,n$, in (27) yields that

$$\begin{aligned}\tilde{V}_{b_1}(y) &= 1 - \sum_{i=1}^{n} s_i \sum_{k=0}^{\infty}\left[\sum_{j=1}^{i}\lambda_j(1-\tilde{f}_j^{(k)}(y)) + \sum_{j=i+1}^{n}\lambda_j(1-\tilde{f}_j^{(k+1)}(y))\right] + o(s) \quad (41)\\ &= 1 - s\tilde{H}(y) + o(s).\end{aligned}$$

Hence, for $s \to 0$ in (38), $\mathbf{W}_1$ approaches $\mathbf{W}_1^0$ in distribution. Using (22) and (32), the same statement follows for other service disciplines satisfying Property 0.1.

$\square$

Remark 3.2 In case the visit order is not strictly cyclic, the joint queue length distribution may still be obtained along the lines of Section 2. However, the relationship between the models with zero and non-zero switchover times as exposed in the present section, will not be as manifest as in the case of a cyclic visit order.

$\square$

4 Computational aspects

In this last section we describe how the waiting-time results obtained can be exploited for a very efficient numerical calculation of the various moments under different switchover time scenarios. For ease of presentation we focus on the first moment, but similar observations hold for the higher moments.

To determine $\mathrm{E}\mathbf{W}_1$ and $\mathrm{E}\mathbf{W}_1^0$ according to formulas (33) and (34), we need to compute the quantities $\tilde{V}_{b_1}'(1)$, $\tilde{V}_{b_1}''(1)$, $\tilde{V}_{b_1}^{0\prime}(1)$, and $\tilde{V}_{b_1}^{0\prime\prime}(1)$. The quantities $\tilde{h}_1'(1)$ and $\tilde{h}_1''(1)$ occurring in (33) and (34) may simply be determined from the service discipline at Q_1.

We first introduce some notation. Denote $\phi_i^{(k)} := \frac{d}{dy}\tilde{f}_i^{(k)}(y)_{|y=1}$, $\psi_i^{(k)} := \frac{d^2}{dy^2}\tilde{f}_i^{(k)}(y)_{|y=1}$, $i = 1,\ldots,n$, $k \geq 0$. Define $\Phi_i := \sum_{k=0}^{\infty} \phi_i^{(k)}$, $\Psi_i := \sum_{k=0}^{\infty} \psi_i^{(k)}$, $i = 1,\ldots,n$, $\Phi := \sum_{i=1}^{n} \lambda_i \Phi_i$,

$\Psi := \sum_{i=1}^{n} \lambda_i \Psi_i$.

Differentiating (35),

$$\tilde{V}_{b_1}^{0\prime}(1) = \frac{V_{b_1}^0(0)}{\lambda} \sum_{k=0}^{\infty} \sum_{i=1}^{n} \lambda_i \phi_i^{(k)} = \frac{V_{b_1}^0(0)}{\lambda} \Phi; \tag{42}$$

$$\tilde{V}_{b_1}^{0\prime\prime}(1) = \frac{V_{b_1}^0(0)}{\lambda} \sum_{k=0}^{\infty} \sum_{i=1}^{n} \lambda_i \psi_i^{(k)} = \frac{V_{b_1}^0(0)}{\lambda} \Psi. \tag{43}$$

Taking $z = (y, 1, \ldots, 1)$ in (27), differentiating w.r.t. to y,

$$\tilde{V}_{b_1}'(1) = s \sum_{k=0}^{\infty} \sum_{i=1}^{n} \lambda_i \phi_i^{(k)} = s\Phi; \tag{44}$$

$$\begin{aligned} \tilde{V}_{b_1}''(1) &= \left(\tilde{V}_{b_1}'(1)\right)^2 + s \sum_{k=0}^{\infty} \sum_{i=1}^{n} \lambda_i \psi_i^{(k)} \\ &+ \sum_{k=0}^{\infty} \sum_{i=1}^{n} \left(s_i^{(2)} - s_i^2\right) \left(\sum_{j=1}^{i} \lambda_j \phi_j^{(k)} + \sum_{j=i+1}^{n} \lambda_j \phi_j^{(k+1)}\right)^2 \\ &= \left(\tilde{V}_{b_1}'(1)\right)^2 + s\Psi + \sum_{i=1}^{n} \left(s_i^{(2)} - s_i^2\right) \chi_i, \end{aligned} \tag{45}$$

with $\chi_i = \sum_{k=0}^{\infty} \left(\sum_{j=1}^{i} \lambda_j \phi_j^{(k)} + \sum_{j=i+1}^{n} \lambda_j \phi_j^{(k+1)} \right)^2$.

From (33), (34), and (42)-(45) we obtain

$$\mathrm{E}\mathbf{W}_1^0 = \frac{\lambda_1 \beta_1^{(2)}}{2(1 - \lambda_1 \beta_1)} + \frac{\Psi}{\Phi \lambda_1}(1 + \tilde{h}_1'(1)) - \frac{\tilde{h}_1''(1)}{2(1 - \tilde{h}_1'(1))\lambda_1}, \tag{46}$$

$$\mathrm{E}\mathbf{W}_1 - \mathrm{E}\mathbf{W}_1^0 = \frac{(s\Phi)^2 + \sum_{i=1}^{n} (s_i^{(2)} - s_i^2)\chi_i}{s\Phi\lambda_1}(1 + \tilde{h}_1'(1)). \tag{47}$$

Note that the coefficients Φ, Ψ, and χ_i, $i = 1, \ldots, n$, are completely *independent* of the switchover times. Apparently the presence of switchover times *always* increases the mean waiting times. That might not be very surprising, but it should be noted that the (perhaps also intuitively appealing) claim that the mean waiting times are increasing in the mean switchover times is known to be false (cf. Sarkar and Zangwill (1991)). We see from (47) that for given s the mean waiting times are increasing in the *variance* of each of the individual switchover times, cf. Sarkar and Zangwill (1991), Righter and Shanthikumar (1994), in particular being minimal for constant switchover times. In the latter case (47) reduces to

$$\mathrm{E}\mathbf{W}_1 - \mathrm{E}\mathbf{W}_1^0 = \frac{s\Phi}{\lambda_1}(1 + \tilde{h}_1'(1)), \tag{48}$$

which illustrates once more that the relationship between the models with and without switchover times takes a remarkably simple form for constant switchover times, cf. (40). The mean waiting times then increase linearly with s, and for given s they are insensitive to the individual s_i's, as observed before by Fuhrmann (1992) and Cooper et al. (1992) for exhaustive and gated service. In this respect it is interesting to mention some results for a two-queue model with exhaustive service at Q_1 and 1-limited service at Q_2 (cf. Groenendijk (1990); note that the branching property now is violated): for constant switchover times, the waiting time distributions are still insensitive to individual s_i, given s; but the linearity in s that is displayed in (48) no longer holds.

Finally we observe from (47) that, once the coefficients Φ, Ψ, and χ_i, $i = 1, \dots, n$, have been calculated, the mean waiting times may be determined for different switchover time scenarios with very little additional effort, cf. also Srinivasan et al. (1994). Those quantities Φ, Ψ, and χ_i, $i = 1, \dots, n$, may be computed as follows. Taking $z = (y, 1, \dots, 1)$ in (16),

$$\tilde{f}_i^{(k+1)}(y) = h_i(\tilde{f}_1^{(k)}(y), \dots, \tilde{f}_i^{(k)}(y), \tilde{f}_{i+1}^{(k+1)}(y), \dots, \tilde{f}_n^{(k+1)}(y)). \tag{49}$$

Define $\mathbf{T}_i$ as the visit time at Q_i generated by an arbitrary type-i customer present at the start of a visit to Q_i. Then $\frac{\partial}{\partial z_j} h_i(z)_{|z=1} = \lambda_j \mathrm{E}\mathbf{T}_i$, $\frac{\partial^2}{\partial z_j \partial z_l} h_i(z)_{|z=1} = \lambda_j \lambda_l \mathrm{E}(\mathbf{T}_i^2)$, $j \neq i$, $l \neq i$. Define $\mathbf{V}_i$ as the total visit time at Q_i in a cycle. From $\mathrm{E}\mathbf{V}_i = \mathrm{E}\mathbf{X}_i \mathrm{E}\mathbf{T}_i$, $\mathrm{E}\mathbf{X}_i = \mathrm{E}\mathbf{Y}_i + \lambda_i(1-\rho_i)\mathrm{E}\mathbf{C}$, $\mathrm{E}\mathbf{V}_i = \rho_i \mathrm{E}\mathbf{C}$, $\mathrm{E}\mathbf{Y}_i = \mathrm{E}\mathbf{X}_i \frac{\partial}{\partial z_i} h_i(z)_{|z=1}$, with $\mathrm{E}\mathbf{C}$ denoting the mean cycle time, it immediately follows that $\frac{\partial}{\partial z_i} h_i(z)_{|z=1} = [\lambda_i - 1/\beta_i]\, \mathrm{E}\mathbf{T}_i + 1$.

Thus, differentiating (49),

$$
\begin{aligned}
\phi_i^{(k+1)} &= \sum_{j=1}^{i} \frac{\partial}{\partial z_j} h_i(z)_{|z=1} \phi_j^{(k)} + \sum_{j=i+1}^{n} \frac{\partial}{\partial z_j} h_i(z)_{|z=1} \phi_j^{(k+1)} \\
&= \mathrm{E}\mathbf{T}_i \left[\sum_{j=1}^{i} \lambda_j \phi_j^{(k)} + \frac{\phi_i^{(k)}}{\mathrm{E}\mathbf{T}_i} - \frac{\phi_i^{(k)}}{\beta_i} + \sum_{j=i+1}^{n} \lambda_j \phi_j^{(k+1)} \right]; \qquad (50) \\
\psi_i^{(k+1)} &= \sum_{j=1}^{i} \frac{\partial}{\partial z_j} h_i(z)_{|z=1} \psi_j^{(k)} + \sum_{j=i+1}^{n} \frac{\partial}{\partial z_j} h_i(z)_{|z=1} \psi_j^{(k+1)} \\
&+ \sum_{j=1}^{i} \sum_{l=1}^{i} \frac{\partial^2}{\partial z_j \partial z_l} h_i(z)_{|z=1} \phi_j^{(k)} \phi_l^{(k)} + 2 \sum_{j=1}^{i} \sum_{l=i+1}^{n} \frac{\partial^2}{\partial z_j \partial z_l} h_i(z)_{|z=1} \phi_j^{(k)} \phi_l^{(k+1)} \\
&+ \sum_{j=i+1}^{n} \sum_{l=i+1}^{n} \frac{\partial^2}{\partial z_j \partial z_l} h_i(z)_{|z=1} \phi_j^{(k+1)} \phi_l^{(k+1)} \\
&= \mathrm{E}\mathbf{T}_i \left[\sum_{j=1}^{i} \lambda_j \psi_j^{(k)} + \frac{\psi_i^{(k)}}{\mathrm{E}\mathbf{T}_i} - \frac{\psi_i^{(k)}}{\beta_i} + \sum_{j=i+1}^{n} \lambda_j \psi_j^{(k+1)} \right] \qquad (51) \\
&+ \mathrm{E}(\mathbf{T}_i^2) \left(\sum_{j=1}^{i} \lambda_j \phi_j^{(k)} + \sum_{j=i+1}^{n} \lambda_j \phi_j^{(k+1)} \right)^2 \\
&+ \phi_i^{(k)} \left[2 \sum_{j=1}^{i-1} \tau_{ij} \phi_j^{(k)} + \tau_{ii} \phi_i^{(k)} + 2 \sum_{j=i+1}^{n} \tau_{ij} \phi_j^{(k+1)} \right],
\end{aligned}
$$

with $\tau_{ij} := \frac{\partial^2}{\partial z_i z_j} h_i(z)_{|z=1} - \lambda_i \lambda_j \mathrm{E}(\mathbf{T}_i^2)$.

Summing (50) and (51) over $k \geq 0$, we obtain

$$
\frac{\Phi_i}{\beta_i} - \frac{\phi_i^{(0)}}{\mathrm{E}\mathbf{T}_i} + \sum_{j=i+1}^{n} \lambda_j \phi_j^{(0)} = \Phi, \qquad (52)
$$

and

$$
\frac{\Psi_i}{\beta_i} - \frac{\psi_i^{(0)} + \mathrm{E}(\mathbf{T}_i^2)\chi_i + \xi_i}{\mathrm{E}\mathbf{T}_i} + \sum_{j=i+1}^{n} \lambda_j \psi_j^{(0)} = \Psi, \qquad (53)
$$

with $\xi_i = \sum_{k=0}^{\infty} \phi_i^{(k)} \left[2 \sum_{j=1}^{i-1} \tau_{ij} \phi_j^{(k)} + \tau_{ii} \phi_i^{(k)} + 2 \sum_{j=i+1}^{n} \tau_{ij} \phi_j^{(k+1)} \right]$.

Multiplying (52) and (53) by ρ_i, summing over $i = 1, \ldots, n$, using $\phi_1^{(0)} = 1$, $\phi_i^{(0)} = 0$, $i = 2, \ldots, n$, $\psi_i^{(0)} = 0$, $i = 1, \ldots, n$, we obtain

$$\Phi = \frac{\rho_1 / \mathrm{E}\mathbf{T}_1}{1 - \rho}, \tag{54}$$

and

$$\Psi = \frac{\sum_{i=1}^{n} \rho_i \left[\mathrm{E}(\mathbf{T}_i^2) \chi_i + \xi_i \right] / \mathrm{E}\mathbf{T}_i}{1 - \rho}. \tag{55}$$

For exhaustive service, i.e., $\mathrm{E}\mathbf{T}_i = \beta_i/(1-\rho_i)$, $\mathrm{E}(\mathbf{T}_i^2) = \beta_i^{(2)}/(1-\rho_i)^3$, $\tau_{ij} = -\lambda_i \lambda_j \mathrm{E}(\mathbf{T}_i^2)$, the expressions for χ_i and ξ_i, $i = 1, \ldots, n$, reduce to those in Srinivasan et al. (1994). For constant switchover times we find from (48) and (54) that $\mathrm{E}\mathbf{W}_1 - \mathrm{E}\mathbf{W}_1^0 = (1 + \tilde{h}_1'(1))\mathrm{E}\mathbf{C}\beta_1/(2\mathrm{E}\mathbf{T}_1)$, which shows that the difference then only depends on the parameters of the other queues through $\mathrm{E}\mathbf{C} = s/(1-\rho)$. In particular, for exhaustive and gated service $\mathrm{E}\mathbf{W}_1 - \mathrm{E}\mathbf{W}_1^0 = (1 - \rho_1)\mathrm{E}\mathbf{C}/2$ and $\mathrm{E}\mathbf{W}_1 - \mathrm{E}\mathbf{W}_1^0 = (1 + \rho_1)\mathrm{E}\mathbf{C}/2$, respectively, as obtained in Fuhrmann (1992) and Cooper et al. (1992) by using different techniques.

Note that the bulk of the computational effort is involved with the calculation of the quantities χ_i and ξ_i, $i = 1, \ldots, n$, which may be done completely *independent* of the switchover times, cf. also Srinivasan et al. (1994). The coefficients $\phi_i^{(k)}$, $i = 1, \ldots, n$, $k \geq 0$, needed for these calculations, may be computed recursively from (50), supplemented with $\phi_1^{(0)} = 1$, $\phi_i^{(0)} = 0$, $i = 2, \ldots, n$. The number of elementary operations (additions, multiplications) involved is $O(n \log_\rho(\epsilon))$ with ϵ the level of accuracy desired. Once the quantities χ_i and ξ_i, $i = 1, \ldots, n$, have been calculated, $\mathrm{E}\mathbf{W}_1$ may be computed for any value of the switchover times in only $O(n)$ elementary operations.

Acknowledgement The authors are indebted to J.A.C. Resing for useful discussions, and to M.M. Srinivasan for showing them a preliminary version of Srinivasan, Niu and Cooper (1994). The work of the second author was supported in part by the European Grant BRA-QMIPS of CEC DG XIII.

References

Athreya, K.B., Ney, P.E. (1972). *Branching Processes.* Springer Verlag, Berlin.

Borst, S.C. (1994). Polling systems with multiple coupled servers. *Queueing Systems*, to appear.

Borst, S.C., Boxma, O.J. (1994). Polling models with and without switchover times, CWI Report BS-R9421, May 1994.

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

Cooper, R.B., Niu, S.-C., Srinivasan, M.M. (1992). A decomposition theorem for polling models: the switchover times are effectively additive. *Oper. Res.*, to appear.

Eisenberg, M. (1972). Queues with periodic service and changeover time. *Oper. Res.* **20**, 440-451.

Eisenberg, M. (1994). The polling system with a stopping server. *Queueing Systems* **18**, 387-431.

FEDERGRUEN, A., KATALAN, Z. (1993). The impact of setup times on the performance of multi-class service and production systems, Report Graduate School of Business, Columbia University, August 1993; *Oper. Res.*, to appear.

FUHRMANN, S.W. (1981). Performance analysis of a class of cyclic schedules, Bell Laboratories technical memorandum 81-59531-1.

FUHRMANN, S.W., COOPER, R.B. (1985). Stochastic decompositions in the $M/G/1$ queue with generalized vacations. *Oper. Res.* **33**, 1117-1129.

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

GROENENDIJK, W.P. (1990). *Conservation Laws in Polling Systems.* PhD Thesis, University of Utrecht.

KEILSON, J., SERVI, L.D. (1990). The distributional form of Little's law and the Fuhrmann-Cooper decomposition. *Oper. Res. Lett.* **9**, 239-247.

KONHEIM, A.G., LEVY, H., SRINIVASAN, M.M. (1994). Descendant set: an efficient approach for the analysis of polling systems. IEEE *Trans. Comm.* **42**, 1245-1253.

LEVY, H., KLEINROCK, L. (1991). Polling systems with zero switch-over periods: a general method for analyzing the expected delay. *Perf. Eval.* **13**, 97-107.

RESING, J.A.C. (1990). *Asymptotic Results in Feedback Systems.* PhD Thesis, Technical University Delft.

RESING, J.A.C. (1993). Polling systems and multitype branching processes. *Queueing Systems* **13**, 409-426.

RIGHTER, R.L., SHANTHIKUMAR, J.G. (1994). Multi-class production systems with set-up times, submitted to *Oper. Res.*.

SARKAR, D., ZANGWILL, W.I. (1991). Variance effects in cyclic production systems. *Mgmt. Sc.* **37**, 444-453.

SRINIVASAN, M.M., NIU, S.-C., COOPER, R.B. (1994). Relating polling models with nonzero and zero switchover times. *Queueing Systems*, to appear.

TAKAGI, H. (1990). Queueing analysis of polling models: an update. In: H. Takagi (ed.), *Stochastic Analysis of Computer and Communication Systems.* North-Holland Publ. Cy., Amsterdam, pp. 267-318.

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