tu/e

Eindhoven University
of Technology

Department of Mathematics
and Computing Sciences

Memorandum COSOR 99-10

Polling systems with regularly varying service and/or switchover times

O.J. Boxma
Q. Deng
J.A.C. Resing

Eindhoven, August 1999
The Netherlands

Polling systems with regularly varying service and/or switchover times

O.J. Boxma[1], Q. Deng[2] and J.A.C. Resing

Department of Mathematics and Computing Science
Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Abstract

We consider a polling system consisting of K queues and a single server S who visits the queues in a cyclic order. The polling discipline in each queue is the gated or exhaustive service discipline. We investigate the tail behaviour of the waiting time distributions at the various queues in the case that at least one of the service time or switchover time distributions has a regularly varying tail.

1 Introduction

Cyclic polling systems are queueing systems in which a server visits several queues in cyclic order. They have a wide range of applications in, e.g., computer communications, manufacturing and traffic. The abundant literature on polling systems (see [31, 32]) contains the exact analysis of polling systems for a large number of service disciplines, like the exhaustive and gated disciplines.

Tail probabilities, which are especially helpful in understanding the performance of different polling disciplines, have received some attention in recent years. For models with Poisson arrivals, general service time and switchover time distributions, and various service disciplines, Choudhury and Whitt [14] have developed efficient iterative algorithms to compute the exact tail behaviour of, for example, the steady-state waiting time W, with the form,

$$\Pr\{W > t\} \sim \alpha t^{\beta} e^{-\eta t}, \quad t \to \infty, \tag{1.1}$$

with $\eta > 0$ and $\alpha > 0$. Here $f(t) \sim g(t)$ for $t \to \infty$ stands for $\lim_{t\to\infty} f(t)/g(t) = 1$. Such tail behaviour occurs when the service time and switchover time distributions have finite moment generating functions, i.e., there is some positive real number s such that $\beta_k(-s) < \infty$ and $\sigma_k(-s) < \infty$; here $\beta_k(\cdot)$ and $\sigma_k(\cdot)$ are the Laplace-Stieltjes transforms (LST) of the service time and switchover time distributions at the k-th queue respectively. Motivated by [14], and using analytic methods, Duffield [18] explores the relationship between the exponents β and η in (1.1) and their dependence (and sometimes independence) on the service time and switchover time distributions.

Recently it has become clear that delay and buffer content distributions in modern communication networks often do not exhibit such an exponential tail behaviour [4, 25, 34]. It appears that input distributions like file size and transmission time distributions in many cases have a $t^{-\nu}$ power tail behaviour, where ν can be less than two – i.e., the variance is infinite. An important and useful class of such heavy-tailed distributions is the class of regularly varying distributions of index $-\nu$ for $1 < \nu < 2$, i.e.,

$$1 - B(t) \sim L(t)t^{-\nu}, \quad t \to \infty, \tag{1.2}$$

[1] also: CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands.

[2] corresponding author: Tel: +31-40-2475478 Fax: +31-40-2465995 Email: deng@win.tue.nl

where $L(t)$ is a slowly varying function [6].

A key issue in present-day performance analysis is the influence of such heavy-tailed input distributions on the tail behaviour of the main performance measures, like waiting time and workload. In the case that the service time distribution has a regularly varying tail as specified in (1.2), the tail behaviour of the waiting time distribution has been investigated in queueing models with various service disciplines. For the $GI/G/1$ queue with FCFS service discipline (cf. [15]) and the $M/G/1$ queue with LCFS Nonpreemptive service discipline (cf. [10]), the waiting time distribution has been proven to have a regularly varying tail of index $1-\nu$, which is one degree higher than that of the service time. An intuitive explanation is that, in those models, customers have to wait at least a residual service time – which has a regularly varying tail of index $1-\nu$. For the $M/G/1$ queue with LCFS Preemptive Resume service discipline [10] and with processor sharing [35, 36], the sojourn time distribution is regularly varying of index $-\nu$, which *is the same* as that of the service time. In [1, 11], the $M/G/1$ queue with two classes of customers and fixed priorities was studied. If at least one of the service time distributions has a regularly varying tail, then the waiting time distribution of the low-priority customers is regularly varying of index one higher than that of the service time distribution with the heaviest tail. In the non-preemptive case, a similar result holds for the tail of the waiting time distribution of the high-priority customers (in the preemptive-resume case, high-priority customers are not affected by low-priority customers).

In view of the central role of polling in computer-communication networks, it is of importance to study the effect of heavy-tailed service and/or switchover time distributions on waiting-time tail behaviour in polling systems. That is the goal of the present paper. We consider waiting time (and, briefly, workload) tail behaviour for cyclic polling systems with Poisson arrivals, general independent service times, general independent switchover times, and the gated or exhaustive service discipline. At each queue, customers are served in the order in which they arrive at each queue. Our main result is the following. If at least one of the service time and/or switchover time distributions has a regularly varying tail of index $-\nu$ ($\nu > 1$) and the others have a lighter tail, then the waiting time distribution at each queue is regularly varying of index $1-\nu$, i.e.,

$$\Pr\{W_k > t\} \sim \alpha_k t^{1-\nu} L(t), \quad t \to \infty,$$

for $\alpha_k > 0$, where W_k is the steady-state waiting time at the k-th queue.

The rest of this paper is organized as follows. Section 2 contains the model description. Section 3 presents an explicit formula for the LST of the waiting time distribution given in [7, 8], that will be the starting-point of the tail investigation. This distribution is expressed in the generating functions of particular queue length distributions; those are discussed in Section 4. The main theorems are provided in Section 5, in which the tail behaviour of the waiting time distribution is given under the assumption that at least one of the service and/or switchover time distributions is regularly varying. Section 6 is devoted to the proof of Theorem 5.1 which describes the tail behaviour of the intervisit time distribution in the case of exhaustive service and the tail behaviour of the cycle time distribution in the case of gated service. Section 7 contains some conclusions and suggestions for further research. Finally, Appendix A provides preliminaries relating the asymptotic behaviour of a distribution function to the behaviour of its LST near the origin (the proof of our main result heavily relies on this), and Appendix B gives some results on the first moment matrix which play a key role in the proof of our main result.

2 Model description

We consider a polling system consisting of K ($K \geq 2$) queues, $Q_1, ..., Q_K$, attended by a single server S. Customers arrive at Q_k, $k = 1, ..., K$, according to a Poisson process with rate λ_k and require a generally distributed service time B_k having distribution function $B_k(\cdot)$, finite first moment β_k and LST $\beta_k(\cdot)$. In the sequel, customers arriving at Q_k are also referred to as type-k customers.

The server visits the queues in a strictly cyclic order, i.e., $Q_1, ..., Q_K, Q_1, ..., Q_K, Q_1, ...$. The service policy at each queue is either gated or exhaustive; we do allow mixtures, like gated service in Q_1 and Q_3 and exhaustive service in the other queues. In gated service, the server serves at a queue exactly those customers that are present at the start of his visit of the queue. In exhaustive service, the server continues to work at a queue until it becomes empty. In this paper we consider both the model with and without switchover times. In the model with switchover times, when moving from Q_k to Q_{k+1}, the server incurs a generally distributed switchover time S_k, having distribution function $S_k(\cdot)$ with finite first moment σ_k and LST $\sigma_k(\cdot)$. Here, we use the convention that $Q_{K+1} = Q_1$. The server continues switching even when the whole system is empty. In the model without switchover times, when the system becomes empty the server makes a full cycle (i.e. passes all the queues at least once) and subsequently stops right before Q_1. When the first new customer arrives, S cycles along the queues to that new customer.

Let us denote by $\lambda := \sum_{k=1}^{K} \lambda_k$ the total arrival intensity of customers, by $\rho_k := \lambda_k \beta_k$ the traffic load at Q_k, and by $\rho := \sum_{k=1}^{K} \rho_k$, the total traffic load. As has been stated for example by Eisenberg [19], Fricker & Jaibi [20] and Resing [27], the condition $\rho < 1$ is a necessary and sufficient condition for ergodicity of a cyclic polling system with gated or exhaustive service. From now on, we assume that this ergodicity condition is satisfied.

3 The waiting time distribution

Let W_k denote the stationary waiting time of type-k customers, with distribution function $W_k(\cdot)$ and LST $w_k(\cdot)$. In this section we give an explicit formula for $w_k(s)$, as provided by Borst and Boxma [8] (see also [30]). Let $W_{k|M/G/1}$, with LST $w_{k|M/G/1}(\cdot)$, denote the waiting time of an arbitrary customer in the 'corresponding' isolated $M/G/1$ queue of Q_k and let $N_{k|I}$, with pgf (probability generating function) $n_{k|I}(\cdot)$, denote the queue length at Q_k at an arbitrary epoch in a non-serving interval for Q_k. The formula for $w_k(s)$ is based on the following decomposition, cf. Keilson and Servi [24],

$$w_k(s) = \mathrm{E}[e^{-sW_k}] = w_{k|M/G/1}(s) n_{k|I}(1 - s/\lambda_k), \quad \mathrm{Re}\ s \geq 0. \tag{3.1}$$

By Pollaczek-Khinchine's formula, $w_{k|M/G/1}(s)$ is given by

$$w_{k|M/G/1}(s) = \mathrm{E}[e^{-sW_{k|M/G/1}}] = \frac{1 - \rho_k}{1 - \rho_k \frac{1 - \beta_k(s)}{\beta_k s}}, \quad \mathrm{Re}\ s \geq 0. \tag{3.2}$$

Introduce X_k, the queue length at Q_k at the beginning of a visit to Q_k, and Y_k, the queue length at Q_k at the end of a visit to Q_k. Borst and Boxma [8] represent $n_{k|I}(1 - s/\lambda_k)$ as

$$n_{k|I}(1 - s/\lambda_k) = \mathrm{E}[(1 - s/\lambda_k)^{N_{k|I}}] = \frac{y_k(s) - x_k(s)}{s(\mathrm{E}X_k - \mathrm{E}Y_k)/\lambda_k}, \quad \mathrm{Re}\ s \geq 0, \tag{3.3}$$

with

$$x_k(s) := \mathrm{E}[(1 - s/\lambda_k)^{X_k}], \qquad y_k(s) := \mathrm{E}[(1 - s/\lambda_k)^{Y_k}]. \tag{3.4}$$

To get a better understanding of the function $n_{k|I}(1-s/\lambda_k)$, we introduce the following random variables. For $k=1,...,K$, let C_k denote the cycle time of Q_k, i.e. the time between two successive arrivals of the server to Q_k, D_k the station time of Q_k, i.e. the time between an arrival of the server to Q_k and the next departure of the server from Q_k, and I_k the intervisit time of Q_k, i.e. the time between the departure of the server from Q_k and his next arrival to Q_k. The distribution functions of C_k, D_k and I_k are denoted by $C_k(\cdot), D_k(\cdot), I_k(\cdot)$ respectively. Furthermore, let I_k^* be the residual intervisit time, with probability density function $\frac{1-I_k(t)}{\mathrm{E}I_k}$. As has been pointed out in [2], the ergodicity condition implies the stationarity of the cycle times, the station times and the intervisit times.

In the case of exhaustive service, apparently $Y_k=0$ and thus $\mathrm{E}[y^{Y_k}]=1$. On the other hand, the customers observed by the server at a visit beginning epoch must have arrived during the previous intervisit time. Since the arrival process at Q_k is a Poisson process with rate λ_k, the generating function of the distribution of the number of arrivals during the previous intervisit time is related to the LST of the intervisit time distribution function as follows:

$$\mathrm{E}[y^{X_k}] = \mathrm{E}[e^{-\lambda_k(1-y)I_k}].$$

Thus it follows from the above relation and (3.4) that $x_k(s) = \mathrm{E}[e^{-sI_k}]$, which in combination with (3.3) implies that $n_{k|I}(1-s/\lambda_k) = \mathrm{E}[e^{-sI_k^*}]$, i.e., $n_{k|I}(1-s/\lambda_k)$ is in fact the LST of the residual intervisit time. Furthermore, the following well-known result is implied by (3.1) ($\overset{\mathrm{d}}{=}$ denoting equality in distribution):

$$W_k \overset{\mathrm{d}}{=} W_{k|M/G/1} + I_k^*, \tag{3.5}$$

where $W_{k|M/G/1}$ and I_k^* are independent.

In the case of gated service, using similar arguments as in [16], X_k and Y_k can be considered as the number of arrivals during the time interval of length C_k and D_k, respectively, and thus $x_k(s) = \mathrm{E}[e^{-sC_k}]$ and $y_k(s) = \mathrm{E}[e^{-sD_k}]$. It is readily seen that

$$n_{k|I}(1-s/\lambda_k) = \mathrm{E}[e^{-sU_k}], \tag{3.6}$$

for some non-negative random variable U_k with density function $\frac{D_k(t)-C_k(t)}{\mathrm{E}C_k-\mathrm{E}D_k}$. Thus it follows from (3.1) that

$$W_k \overset{\mathrm{d}}{=} W_{k|M/G/1} + U_k, \tag{3.7}$$

where $W_{k|M/G/1}$ and U_k are independent. The probabilistic meaning of U_k is: $U_k = D_k + I_k^*$, cf. [5] for the case of nonzero switchover times; it is the sum of the station time D_k and the subsequent excess intervisit time I_k^* (that depends on D_k).

4 Joint queue length distribution

In order to obtain an explicit expression for the LST $w_k(s)$ of the waiting time distribution, we need an expression for the generating functions of the queue lengths X_k and Y_k at the beginning and end, respectively, of a server visit to Q_k (see (3.1) and (3.3)). We first concentrate on an expression for $F_k(\mathbf{z})$ and $G_k(\mathbf{z})$, $\mathbf{z}=(z_1,...,z_K)^T$, $|z_j| \leq 1$, $j=1,...,K$, the pgf of the *joint* queue length vector at visit beginning and visit completion epochs. Here, we follow the approach of Resing [27]. In fact, in [27] a more general class of service disciplines called Bernoulli-type service is considered, which contains the gated and exhaustive service disciplines. This class of service disciplines satisfies the following property.

Property 4.1 *If the server arrives at Q_k to find n_k customers there, then during the course of the server's visit, each of these n_k customers is effectively replaced in an i.i.d. manner by a random population having pgf $h_k(\mathbf{z})$.*

The gated and exhaustive service discipline both satisfy this property. In these cases the functions $h_k(\mathbf{z})$ are, respectively, given by

$$h_k(\mathbf{z}) = \beta_k(\sum_{j=1}^{K} \lambda_j(1 - z_j)), \tag{4.8}$$

which is the pgf of the joint distribution of the numbers of arrivals at all queues during one service time at Q_k, and

$$h_k(\mathbf{z}) = \eta_k(\sum_{j \neq k} \lambda_j(1 - z_j)), \tag{4.9}$$

where $\eta_k(\cdot)$ denotes the LST of the length of a busy period in an isolated $M/G/1$ queue with arrival rate λ_k and service time distribution $B_k(\cdot)$. In the case of exhaustive service, the function $h_k(\mathbf{z})$ represents the pgf of the joint distribution of the numbers of arrivals at all other queues during a busy period of Q_k when this queue was in isolation.

In the remainder of this section, we prefer to consider the queue length pgf for the class of service disciplines that satisfy Property 4.1; it may be worthwhile to investigate whether/how the results of the present paper can be generalized to the case of that general class. For service disciplines satisfying Property 4.1, the pgf's $G_k(\mathbf{z})$ (queue length at departure epochs of the server from Q_k) can be nicely related to the pgf's $F_k(\mathbf{z})$ (queue length at arrival epochs of the server at Q_k), for $k = 1, ..., K$, by

$$G_k(\mathbf{z}) = F_k(z_1, ..., z_{k-1}, h_k(\mathbf{z}), z_{k+1}, ..., z_K). \tag{4.10}$$

Next, define for $|z_j| \leq 1, j = 1, ..., K$ the functions

$$\mathbf{f}(\mathbf{z}) := (f_1(\mathbf{z}), ..., f_K(\mathbf{z}))^T, \tag{4.11}$$

with

$$f_k(\mathbf{z}) := h_k(z_1, ..., z_k, f_{k+1}(\mathbf{z}), ..., f_K(\mathbf{z})), \tag{4.12}$$

and the iterates

$$\begin{aligned} \mathbf{f}^{(0)}(\mathbf{z}) &:= \mathbf{z}, \\ \mathbf{f}^{(i)}(\mathbf{z}) &:= \mathbf{f}(\mathbf{f}^{(i-1)}(\mathbf{z})), \quad i \geq 1. \end{aligned}$$

In the following we distinguish between the case with and the case without switchover times.

Zero switchover times

In the sequel we add a superscript 0 for the case of zero switchover times, in order to distinguish its quantities from those for the case with switchover times. The pgf's $F_k^0(\cdot)$ and $G_{k-1}^0(\cdot)$ are related by

$$F_k^0(\mathbf{z}) = G_{k-1}^0(\mathbf{z}), \quad \text{for } k = 2, ..., K, \tag{4.13}$$

$$F_1^0(\mathbf{z}) = G_K^0(\mathbf{z}) - F_1^0(\mathbf{0})(\sum_{j=1}^{K} \frac{\lambda_j}{\lambda}(1 - z_j)), \tag{4.14}$$

where $\mathbf{0}$ stands for the K-dimensional vector with all components equal to zero. Equation (4.14) is obtained by using the special convention that when the system is empty at the start of a visit to Q_1, the next visit does not take place until a customer has arrived. In fact $F_1^0(\mathbf{z})$ satisfies the functional equation

$$F_1^0(\mathbf{z}) = F_1^0(\mathbf{f}(\mathbf{z})) - F_1^0(\mathbf{0}) \sum_{k=1}^{K} \frac{\lambda_k}{\lambda}(1 - z_k), \tag{4.15}$$

the solution of which is, after iteration, given by

$$F_1^0(\mathbf{z}) = 1 - \frac{F_1^0(\mathbf{0})}{\lambda} \sum_{i=1}^{\infty} \sum_{k=1}^{K} \lambda_k (1 - f_k^{(i)}(\mathbf{z})), \tag{4.16}$$

with

$$F_1^0(\mathbf{0}) = \left[1 + \frac{1}{\lambda} \sum_{i=1}^{\infty} \sum_{k=1}^{K} \lambda_k (1 - f_k^{(i)}(\mathbf{0})) \right]^{-1}.$$

The infinite sum $\sum_{i=1}^{\infty} \sum_{k=1}^{K} \lambda_k (1 - f_k^{(i)}(\mathbf{0}))$ is convergent when the ergodicity condition is fulfilled. Once we know $F_1^0(\mathbf{z})$, we immediately get, by using (4.10) and (4.13), the pgf $F_k^0(\mathbf{z})$ of the joint queue length distribution at a visit beginning epoch,

$$F_k^0(\mathbf{z}) = F_{k-1}^0(z_1, ..., z_{k-2}, h_{k-1}(\mathbf{z}), z_k, ..., z_K), \quad \text{for } k = 2, ..., K. \tag{4.17}$$

Furthermore, by the Relations (4.13) and (4.14) we get an expression for $G_k^0(\mathbf{z})$,

$$G_k^0(\mathbf{z}) \quad = \quad F_{k+1}^0(\mathbf{z}), \quad \text{for } k = 1, ..., K-1, \tag{4.18}$$

$$G_K^0(\mathbf{z}) \quad = \quad F_1^0(\mathbf{z}) + \frac{F_1^0(\mathbf{0})}{\lambda} (\sum_{j=1}^{K} \lambda_j (1 - z_j)). \tag{4.19}$$

Non-zero switchover times
In the case of non-zero switchover times, the following equations relate $F_k(\mathbf{z})$ to $G_{k-1}(\mathbf{z})$:

$$F_k(\mathbf{z}) \quad = \quad G_{k-1}(\mathbf{z}) \sigma_{k-1} (\sum_{j=1}^{K} \lambda_j (1 - z_j)), \quad \text{for } k = 2, ..., K, \tag{4.20}$$

$$F_1(\mathbf{z}) \quad = \quad G_K(\mathbf{z}) \sigma_K (\sum_{j=1}^{K} \lambda_j (1 - z_j)). \tag{4.21}$$

Together with Equation (4.10) this leads to the functional equation

$$F_1(\mathbf{z}) = F_1(\mathbf{f}(\mathbf{z})) g(\mathbf{z}), \tag{4.22}$$

with

$$g(\mathbf{z}) = \prod_{k=1}^{K} \sigma_k (\sum_{j=1}^{k} \lambda_j (1 - z_j) + \sum_{j=k+1}^{K} \lambda_j (1 - f_j(\mathbf{z})).$$

The solution of (4.22) is given by

$$
\begin{aligned}
F_1(\mathbf{z}) &= \prod_{i=1}^{\infty} g(\mathbf{f}^{(i)}(\mathbf{z})) \\
&= \prod_{i=1}^{\infty}\prod_{k=1}^{K} \sigma_k(\sum_{j=1}^{k} \lambda_j(1-f_j^{(i)}(\mathbf{z})) + \sum_{j=k+1}^{K} \lambda_j(1-f_j^{(i+1)}(\mathbf{z}))). \qquad (4.23)
\end{aligned}
$$

Again, the infinite product is convergent when the ergodicity condition is fulfilled.

To make the obtained queue length pgf expressions suitable for the analysis of waiting time tail behaviour, we have to slightly rewrite them (we want to move from pgf asymptotics near 1 to LST asymptotics near 0).

Put $\mathbf{r} := (r_1, ..., r_K)^T$, where $0 \leq r_k \leq \lambda_k$, and relate $\mathbf{z}$ to $\mathbf{r}$ by $\mathbf{z} = (1 - r_1/\lambda_1, ..., 1 - r_K/\lambda_K)^T$. If we define $\tilde{F}_k(\mathbf{r}) := F_k(\mathbf{z})$ and $\tilde{G}_k(\mathbf{r}) := G_k(\mathbf{z})$, then it follows from (4.10) that

$$
\tilde{G}_k(\mathbf{r}) = \tilde{F}_k(r_1, ..., r_{k-1}, \tilde{h}_k(\mathbf{r}), r_{k+1}, ..., r_K), \qquad (4.24)
$$

with

$$
\begin{aligned}
\tilde{h}_k(\mathbf{r}) &:= \lambda_k(1 - h_k(\mathbf{z})) \\
&= \begin{cases} \lambda_k(1 - \beta_k(\sum_{j=1}^{K} r_j)), & \text{for gated service,} \\ \lambda_k(1 - \eta_k(\sum_{j \neq k} r_j)), & \text{for exhaustive service.} \end{cases} \qquad (4.25)
\end{aligned}
$$

Furthermore, we define similarly as in (4.11) and (4.12) the functions

$$
\tilde{\mathbf{f}}(\mathbf{r}) := (\tilde{f}_1(\mathbf{r}), ..., \tilde{f}_K(\mathbf{r}))^T,
$$

with

$$
\tilde{f}_k(\mathbf{r}) := \tilde{h}_k(r_1, ..., r_k, \tilde{f}_{k+1}(\mathbf{r}), ..., \tilde{f}_K(\mathbf{r})), \qquad (4.26)
$$

and the iterates

$$
\begin{aligned}
\tilde{\mathbf{f}}^{(0)}(\mathbf{r}) &:= \mathbf{r}, \\
\tilde{\mathbf{f}}^{(i)}(\mathbf{r}) &:= \tilde{\mathbf{f}}(\tilde{\mathbf{f}}^{(i-1)}(\mathbf{r})), \ i \geq 1.
\end{aligned}
$$

In the following we distinguish again between the cases of zero switchover times and non-zero switchover times.

Zero switchover times

The following equations relate $\tilde{F}_k^0(\mathbf{r})$ to $\tilde{G}_{k-1}^0(\mathbf{r})$. It follows from (4.13) and (4.14) that

$$
\tilde{F}_k^0(\mathbf{r}) = \tilde{G}_{k-1}^0(\mathbf{r}), \quad \text{for } i = 2, ..., K, \qquad (4.27)
$$

$$
\tilde{F}_1^0(\mathbf{r}) = \tilde{G}_K^0(\mathbf{r}) - \frac{\tilde{F}_1^0(\mathbf{\Lambda})}{\lambda} \sum_{j=1}^{K} r_j, \qquad (4.28)
$$

where $\mathbf{\Lambda} = (\lambda_1, ..., \lambda_K)$. Introduce, for $0 < r_k < \lambda_k$, $k = 1, ..., K$,

$$
H(\mathbf{r}) = \sum_{i=1}^{\infty}\sum_{k=1}^{K} \tilde{f}_k^{(i)}(\mathbf{r}), \qquad (4.29)
$$

which is well-defined if the ergodicity condition is fulfilled. Then, by (4.16), we can write

$$\tilde{F}_1^0(\mathbf{r}) := F_1^0(\mathbf{z}) = 1 - \tilde{F}_1^0(\mathbf{\Lambda})H(\mathbf{r})/\lambda. \tag{4.30}$$

By using (4.24), (4.27) and (4.28), one can derive expressions for $\tilde{F}_k^0(\mathbf{r})$ and $\tilde{G}_k^0(\mathbf{r})$, $k = 1, ..., K$.

Non-zero switchover times
It follows from (4.20) and (4.21) that

$$\tilde{F}_k(\mathbf{r}) = \tilde{G}_{k-1}(\mathbf{r})\sigma_{k-1}(\sum_{j=1}^{K} r_j), \quad \text{for } k = 2, ..., K, \tag{4.31}$$

$$\tilde{F}_1(\mathbf{r}) = \tilde{G}_K(\mathbf{r})\sigma_K(\sum_{j=1}^{K} r_j). \tag{4.32}$$

Put

$$\tilde{g}(\mathbf{r}) := \prod_{k=1}^{K} \sigma_k(\sum_{j=1}^{k} r_j + \sum_{j=k+1}^{K} \tilde{f}_j(\mathbf{r})).$$

Replacing $\mathbf{z} = (1 - r_1/\lambda_1, ..., 1 - r_K/\lambda_K)$ into (4.23), we obtain

$$\begin{aligned}\tilde{F}_1(\mathbf{r}) &= \prod_{i=1}^{\infty} \tilde{g}(\tilde{\mathbf{f}}^{(i)}(\mathbf{r})) \\ &= \prod_{i=1}^{\infty} \prod_{k=1}^{K} \sigma_k(\sum_{j=1}^{k} \tilde{f}_j^{(i)}(\mathbf{r}) + \sum_{j=k+1}^{K} \tilde{f}_j^{(i+1)}(\mathbf{r})),\end{aligned} \tag{4.33}$$

the infinite product being convergent for $0 \leq r_k \leq \lambda_k$, $k = 1, ..., K$, when the ergodicity condition is fulfilled. By using (4.24), (4.31) and (4.32), one can derive expressions for $\tilde{G}_1(\mathbf{r}), \tilde{G}_k(\mathbf{r}), \tilde{F}_k(\mathbf{r})$ $(k = 2, ..., K)$ immediately.

Marginal queue length pgf
As a final step towards the waiting time LST (see (3.1) and (3.3)), we now obtain the pgf of the marginal distributions of the queue lengths X_k and Y_k at the beginning and end of a server visit to Q_k. For simplicity we define $\mathbf{e} = (1, ..., 1)^T$ and for $k = 1, ..., K$, $\mathbf{e}_k = (0, ..., 0, 1, 0, ..., 0)^T$ with the k-th component being 1. Taking $\mathbf{r} = \mathbf{e}_k s$ in (4.30) for the case of zero switchover times (add a superscript "0") or in (4.33) for the case of non-zero switchover times, we get

$$x_k(s) := \mathrm{E}[(1 - s/\lambda_k)^{X_k}] = \tilde{F}_k(\mathbf{e}_k s), \tag{4.34}$$

$$y_k(s) := \mathrm{E}[(1 - s/\lambda_k)^{Y_k}] = \tilde{G}_k(\mathbf{e}_k s). \tag{4.35}$$

5 The main result

In this section we shall present our main result: *If at least one of the service and/or switchover times is regularly varying of index $-\nu$ $(\nu > 1)$ and the other service and/or switchover times have a lighter tail, then the waiting time distribution is regularly varying of index* $1 - \nu$. As a by-product, we also show that the intervisit time distribution at Q_k $(k = 1, ..., K)$ in the case of exhaustive service and the cycle time and station time distributions at Q_k $(k = 1, ..., K)$ in the case of gated service are all regularly varying of index $-\nu$.

As pointed out in Section 3, W_k can be represented as the sum of two independent random variables $W_{k|M/G/1}$ and V_k where $V_k = I_k^*$ is the residual intervisit time in the case of exhaustive service and $V_k = U_k$ where the LST of U_k is given by (3.6) in the case of gated service. The relation between $1 - W_{k|M/G/1}(t)$ and $1 - B_k(t)$ for $t \to \infty$ is already well known if the residual service time has a subexponential tail (this contains the case of a regularly varying tail),

$$1 - W_{k|M/G/1}(t) \sim \frac{\lambda}{1-\rho} \int_{x=t}^{\infty} (1 - B_k(x))\mathrm{d}x, \quad t \to \infty, \tag{5.1}$$

cf. [26]. In the following we first investigate the tail behaviour of the distribution of V_k by analyzing the asymptotic behaviour of its LST $n_{k|I}(1 - s/\lambda_k)$ for $s \downarrow 0$, and we subsequently derive the tail behaviour $1 - W_k(t)$ of the waiting time distribution for $t \to \infty$.

Without loss of generality, we shall only analyze the explicit expression (3.3) of $n_{k|I}(1-s/\lambda_k)$ for $k = 1$. Combining (4.24), (4.34) and (4.35) yields

$$y_1(s) = x_1(\tilde{h}_1(\mathrm{e}_1 s)). \tag{5.2}$$

If the asymptotic behaviour of $x_1(s)$ for $s \downarrow 0$ is known, we can obtain the asymptotic behaviour of $y_1(s)$ for $s \downarrow 0$ immediately by using Lemma 8.3 in Appendix A.

Assume that the tail behaviour of the service time and switchover time distributions is such that, for $k = 1, \ldots, K$:

$$1 - B_k(t) = [b_k + \mathrm{o}(1)]t^{-\nu}L(t), \quad t \to \infty, \tag{5.3}$$

$$1 - S_k(t) = [s_k + \mathrm{o}(1)]t^{-\nu}L(t), \quad t \to \infty, \tag{5.4}$$

where b_k, $s_k \geq 0$ and $L(\cdot)$ is a slowly varying function; such a distribution is called regularly varying in case the constant (b_k or s_k) is positive. For ease of presentation, we take the same function $L(\cdot)$ for all distributions, but one can easily change this into different slowly varying functions for different distributions. Note that the possibility that $b_k = 0$ or $s_k = 0$ implies that we *do allow* the possibility that some of the service and/or switchover time distributions have an exponential tail, *or* a regularly varying tail that is less heavy than $t^{-\nu}$. According to Lemma 8.1, the tail behaviour of the service time and switchover time distributions as given in (5.3) and (5.4) is equivalent with the following behaviour of their LSTs $\beta_1(s), ..., \beta_K(s)$ (of the service time distributions) and $\sigma_1(s)$,..., $\sigma_K(s)$ (of the switchover time distributions):

$$1 - \beta_k(s) = \sum_{j=1}^{m}(-1)^{j+1}\beta_{k,j}s^j + (-1)^m\beta_{k,\nu}s^{\nu}L(1/s) + \mathrm{o}(s^{\nu}L(1/s)), \ k = 1, ..., K, \tag{5.5}$$

$$1 - \sigma_k(s) = \sum_{j=1}^{m}(-1)^{j+1}\sigma_{k,j}s^j + (-1)^m\sigma_{k,\nu}s^{\nu}L(1/s) + \mathrm{o}(s^{\nu}L(1/s)), \quad k = 1, ..., K, \tag{5.6}$$

where $m < \nu < m+1$, $\beta_{k,j} > 0$, $\beta_{k,\nu} \geq 0$, $\sigma_{k,j} > 0$ and $\sigma_{k,\nu} \geq 0$ for $j = 1, ..., m$, $k = 1, ..., K$. Note that $\beta_{k,1} = \beta_k$, $\beta_{k,\nu} = (-1)^m\Gamma(1-\nu)b_k$, $\sigma_{k,1} = \sigma_k$ and $\sigma_{k,\nu} = (-1)^m\Gamma(1-\nu)s_k$ for $k = 1, ..., K$, and $\sigma_k(s) \equiv 1$ if the switchover time is zero. In order to simplify the proof of Theorem 5.1 below in Section 6, we assume, without loss of generality, that $s^{\nu}L(1/s)$ is a non-decreasing function for $s > 0$.

It follows from the main result of [17], that the asymptotic behaviour of the LST $\eta_k(s)$ of the length of the busy period in the 'corresponding' isolated $M/G/1$ queue of Q_k is given by

$$1 - \eta_k(s) = \sum_{j=1}^{m}(-1)^{j+1}\eta_{k,j}s^j + (-1)^m \eta_{k,\nu}s^\nu L(1/s) + \mathrm{o}(s^\nu L(1/s)), \quad k = 1, ..., K, \tag{5.7}$$

where $\eta_{k,1} = \beta_k/(1-\rho_k)$ and $\eta_{k,\nu} = \beta_{k,\nu}/(1-\rho_k)^\nu$ and $\eta_{k,j} > 0$ for $j = 1, ..., m$.

Theorem 5.1 *If Relations (5.5) and (5.6) hold, then*

$$x_1(s) = \sum_{j=0}^{m}(-1)^j x_{1,j}s^j + (-1)^{m+1}x_{1,\nu}s^\nu L(1/s) + \mathrm{o}(s^\nu L(1/s)), \tag{5.8}$$

where $x_{1,j} \geq 0$ for $j = 1, ..., m$ and $x_{1,\nu} \geq 0$. Moreover, $x_{1,\nu} = 0$ if and only if $\sum_{k=1}^{K}(\beta_{k,\nu}+\sigma_{k,\nu}) =$ 0.

Proof. See Section 6. □

The next corollary, which follows immediately from Theorem 5.1 and Relation (5.2) by using Lemma 8.3 in Appendix A, characterizes the asymptotic behaviour of $y_1(s)$ for $s \downarrow 0$ in the gated case. Remember that $y_1(s) \equiv 1$ if the service discipline at Q_1 is exhaustive.

Corollary 5.1 *In the case of gated service at Q_1, if (5.5) and (5.6) hold, then*

$$y_1(s) = \sum_{j=0}^{m}(-1)^j y_{1,j}s^j + (-1)^{m+1}y_{1,\nu}s^\nu L(1/s) + \mathrm{o}(s^\nu L(1/s)), \tag{5.9}$$

where $y_{1,j} \geq 0$ for $j = 1, ..., m$ and $y_{1,\nu} \geq 0$. Moreover, $y_{1,\nu} = x_{1,1}\beta_{1,\nu} + x_{1,\nu}\rho_1^\nu$.

It is now easy to give the asymptotic expansion of $n_{1|I}(1 - s/\lambda_1)$ for $s \downarrow 0$.

Corollary 5.2 *If (5.5) and (5.6) hold, then*

$$n_{1|I}(1 - s/\lambda_1) = \sum_{j=1}^{m-1}(-1)^j n_{1|I,j}s^j + (-1)^m n_{1|I,\nu}s^\nu L(1/s) + \mathrm{o}(s^\nu L(1/s)), \tag{5.10}$$

where $n_{1|I,j} > 0$ for $j = 1, ..., m-1$ and $n_{1|I,\nu} \geq 0$. Moreover, if $\sum_{k=1}^{K}(\beta_{k,\nu} + \sigma_{k,\nu}) > 0$, then $n_{1|I,\nu} > 0$.

Proof. By (3.3), (5.8) and (5.9), (5.10) follows. As shown in Section 3, $n_{1|I}(1 - s/\lambda_1)$ is the LST of some non-negative random variable. Thus $n_{1|I,j} > 0$, $n_{1|I,\nu} \geq 0$ in (5.10). Again by (3.3), $n_{1|I,\nu} = \lambda_1(x_{1,\nu} - y_{1,\nu})/(\mathrm{E}X_1 - \mathrm{E}Y_1)$ where $y_{1,\nu} = x_{1,1}\beta_{1,\nu} + x_{1,\nu}\rho_1^\nu$. By using Formula (6.5) from the next section for the case of zero switchover time or Formula (6.40) for the case of non-zero switchover time, we can prove that $n_{1|I,\nu} > 0$ if $\sum_{k=1}^{K}(\beta_{k,\nu} + \sigma_{k,\nu}) > 0$. □

Applying Lemma 8.1 in Appendix A, we get the following theorem on the relation between the tail behaviour of the service time distribution and that of the intervisit time, cycle time, station time and waiting time distributions from the above results.

Theorem 5.2 *Suppose the tail behaviour of the service time and switchover time distributions is such that*

$$1 - B_k(t) = [b_k + \mathrm{o}(1)]t^{-\nu}L(t), \quad t \to \infty, \tag{5.11}$$

$$1 - S_k(t) = [s_k + \mathrm{o}(1)]t^{-\nu}L(t), \quad t \to \infty, \tag{5.12}$$

where $b_k, s_k \geq 0$, $L(\cdot)$ is a slowly varying function and $k = 1, ..., K$. Then in the case of exhaustive service at Q_1, the tail behaviour of the intervisit time and waiting time at Q_1 satisfies the following relations:

$$1 - I_1(t) = [c_1 + \mathrm{o}(1)]t^{-\nu}L(t), \quad t \to \infty, \tag{5.13}$$

$$1 - W_1(t) = [c_2 + \mathrm{o}(1)]t^{1-\nu}L(t), \quad t \to \infty; \tag{5.14}$$

in the case of gated service at Q_1, the tail behaviour of the cycle time, station time, U_1 with LST given by (3.6) and waiting time at Q_1 is given by

$$1 - C_1(t) = [c_3 + \mathrm{o}(1)]t^{-\nu}L(t), \quad t \to \infty, \tag{5.15}$$

$$1 - D_1(t) = [c_4 + \mathrm{o}(1)]t^{-\nu}L(t), \quad t \to \infty, \tag{5.16}$$

$$1 - U_1(t) = [c_5 + \mathrm{o}(1)]t^{1-\nu}L(t), \quad t \to \infty, \tag{5.17}$$

$$1 - W_1(t) = [c_6 + \mathrm{o}(1)]t^{1-\nu}L(t), \quad t \to \infty, \tag{5.18}$$

where c_r are nonnegative constants for $r = 1, ..., 6$. Moreover, if $\sum_{k=1}^{K}(b_k + s_k) = 0$, then $c_r = 0$ for $r = 1, ..., 6$; if $\sum_{k=1}^{K}(b_k + s_k) > 0$, then $c_r > 0$ for $r = 1, ..., 6$.

Proof. In the case of exhaustive service at Q_1, applying Theorem 5.1 and Lemma 8.1 in Appendix A, (5.13) follows immediately. Combining (3.5), (5.1), (5.13) and using Lemma 7.7 in [12] yields (5.14) where

$$c_2 = \frac{\lambda_1 b_1}{(1-\rho_1)(\nu-1)} + \frac{c_1}{\mathrm{E}I_1(\nu-1)}.$$

In the case of gated service at Q_1, Relations (5.15) and (5.16) follow immediately from Theorem 5.1, Corollary 5.1 and Lemma 8.1 in Appendix A. Relation (5.17) follows from Corollary 5.2. Combining (3.5), (5.1), (5.13) and using Lemma 7.7 in [12] yields (5.14) where

$$c_6 = \frac{\lambda_1 b_1}{(1-\rho_1)(\nu-1)} + \frac{c_5}{\nu-1}.$$

It is easy to see that if $\sum_{k=1}^{K}(b_k + s_k) = 0$, then $c_r = 0$ for $r = 1, ..., 6$; if $\sum_{k=1}^{K}(b_k + s_k) > 0$, then $c_r > 0$ for $r = 1, ..., 6$. □

By symmetry, Theorem 5.1 and Corollaries 5.1, 5.2 hold not just for Q_1 but for each queue. Thus Theorem 5.2 also holds for each queue.

Remark 5.1. In order to get explicit expressions for c_r in the above theorem in terms of b_k and s_k for $k = 1, ..., K$, one can refer to Relation (6.5) in the case of zero switchover time or

Relation (6.40) in the case of non-zero switchover time.

Remark 5.2. Consider the $M/G/1$ queue with repeated vacations. The server continues serving until the system has become empty, and then takes a vacation V. If the system is still empty after this vacation, then he takes another vacation, etc.; successive vacations are independent and identically distributed. Fuhrmann and Cooper [21] have proven the following decomposition result:

$$W_{with} = W_{M/G/1} + V^*, \tag{5.19}$$

where W_{with} ($W_{M/G/1}$) denotes waiting time in the model with (without) vacations and V^* has the equilibrium (residual lifetime) distribution of V; $W_{M/G/1}$ and V^* are independent.
This vacation queue is a special case of the polling model with switchover times and exhaustive service; take $N = 1$. Asmussen et al. [3] have proven the following result, for the $M/G/1$ vacation queue with residual vacation or residual service time distributions that belong to the class $\mathcal{S}$ of subexponential distributions (which contains the class of regularly varying distributions):
(i) If the equilibrium service time $S^* \in \mathcal{S}$ and if $\mathrm{P}(V^* > x) = \mathrm{o}(\mathrm{P}(S^* > x))$ as $x \to \infty$, then

$$\mathrm{P}(W_{with} > x) \sim \frac{\rho}{1-\rho}\mathrm{P}(S^* > x), \quad x \to \infty;$$

(ii) If the equilibrium vacation time $V^* \in \mathcal{S}$ and if $\mathrm{P}(S^* > x) = \mathrm{o}(\mathrm{P}(V^* > x))$ as $x \to \infty$, then

$$\mathrm{P}(W_{with} > x) \sim \mathrm{P}(V^* > x), \quad x \to \infty;$$

(iii) If the equilibrium service time $S^* \in \mathcal{S}$ and if $\mathrm{P}(V^* > x) \sim c\mathrm{P}(S^* > x)$ as $x \to \infty$ for some $c \geq 0$, then

$$\mathrm{P}(W_{with} > x) \sim (c + \frac{\rho}{1-\rho})\mathrm{P}(S^* > x), \quad x \to \infty.$$

In the polling model one might also try to prove that the waiting time distribution is *subexponential* in the case of subexponential service and/or switchover time distributions. However, at least in the case of exhaustive service at some queue Q_k, this requires solution of the following open problem, that seems quite hard (cf. [3]): Is the busy period distribution of $M/G/1$ queue Q_k in isolation subexponential, when its service time distribution is subexponential? (Notice that the busy period distribution of Q_k appears prominently in $h_k(\mathbf{z})$ and $\tilde{f}_k(\mathbf{r})$, cf. (4.9) and (4.26)).

Remark 5.3. In the present paper we have concentrated on the tails of the waiting time distributions. It is slightly easier to study the tail behaviour of the total workload distribution in a polling system. Boxma and Groenendijk [13] (cf. also Boxma [9] for generalizations) have proven the following workload decomposition for a broad category of multiclass queueing systems with Poisson arrivals and server vacations – a category that includes cyclic polling systems with switchover times:

$$U = U_{M/G/1} + Z, \tag{5.20}$$

$U_{M/G/1}$ and Z being independent. Here U is the steady-state workload in the system, $U_{M/G/1}$ is the steady-state workload in the corresponding $M/G/1$ queue to which the multiclass system reduces when there are no switchovers, and Z is the steady-state workload at an arbitrary time during a vacation. Takagi et al. [33] provide an expression for the LST of the distribution of Z, in the case of either exhaustive or gated service at all queues. Using that expression and the above decomposition result, one can apply the technique in Section 6 of this paper to obtain similar tail behaviour results for the workload as for the individual waiting times.

Remark 5.4. In the $M/G/1$ FCFS queue, if the service time distribution is regularly varying of index $-\nu$ ($\nu > 1$), then the waiting time distribution is regularly varying of index $1-\nu$. However, the $M/G/1$ queue with the LCFS preemptive resume discipline has the attractive feature that the waiting time distribution is regularly varying of index $-\nu$ only [10]. In the polling system with a LCFS preemptive resume discipline *within a queue visit of the server*, customers may have to wait a residual cycle time (in the case of gated service) or a residual intervisit time (in the case of exhaustive service), and these are regularly varying of index $1-\nu$. Thus one cannot expect to get a 'better' index than $1-\nu$ by providing LCFS preemptive resume service within a queue visit of the server.

6 Proof of Theorem 5.1

We shall prove the cases of zero and non-zero switchover times separately. We restrict ourselves mainly to the cases in which all queues are served according to the *same* discipline (gated, or exhaustive); the proofs require only minor adaptations in the case of mixtures of these disciplines.

1. The case of zero switchover times
Using (4.30) and (4.34) we have

$$x_1(s) = \tilde{F}_1^0(\mathbf{e}_1 s) = 1 - \tilde{F}_1^0(\Lambda)H(\mathbf{e}_1 s)/\lambda. \tag{6.1}$$

In the following we concentrate on determining the asymptotic behaviour of $H(\mathbf{e}_1 s)$ for $s \downarrow 0$. We shall prove that

$$H(\mathbf{e}_1 s) = \sum_{j=1}^{m} H_{1,j}s^j + (-1)^m H_{1,\nu}s^\nu L(1/s) + \mathrm{o}(s^\nu L(1/s)), \quad s \downarrow 0, \tag{6.2}$$

for some constants $H_{1,j}$ where $j = 1, ..., m$ and $H_{1,\nu} \geq 0$. The proof of Relation (6.2) is divided into three steps. In the first step, we construct a new function $P(\cdot)$ which has a similar structure as $H(\cdot)$. In the second step, we shall show that the asymptotic expansion of this function is given by

$$P(\mathbf{e}_1 s) = \sum_{j=1}^{m} P_{1,j}s^j + \mathrm{O}(s^{m+1}), \quad \textit{for } s \downarrow 0. \tag{6.3}$$

Finally, in the third step we will show that

$$\lim_{s\downarrow 0} \frac{H(\mathbf{e}_1 s) - P(\mathbf{e}_1 s)}{s^\nu L(1/s)} = (-1)^m H_{1,\nu}, \tag{6.4}$$

for some non-negative constant $H_{1,\nu}$. Clearly, Relation (6.2) follows by combining (6.3) and (6.4). Once we have proven (6.2), the proof of Theorem 5.1 is almost completed. Substituting (6.2) into (6.1) and noting that $x_1(s)$ is the LST of some non-negative random variable (cycle time if the service discipline at Q_1 is gated service or intervisit time if the service discipline at Q_1 is exhaustive service) yields Formula (5.8) of Theorem 5.1, where

$$x_{1,1} = \tilde{F}_1^0(\Lambda)P_{1,1}/\lambda, \qquad x_{1,\nu} = \tilde{F}_1^0(\Lambda)H_{1,\nu}/\lambda. \tag{6.5}$$

Step 1: Similarly as we constructed the function $H(\cdot)$ in Section 4, we now construct the function $P(\cdot)$. So, define (cf. (4.25), and notice that we take λ_k times the first m terms in the righthand sides of (5.5) and (5.7)):

$$\xi_k(\mathbf{r}) := \begin{cases} \lambda_k \sum_{j=1}^{m}(-1)^{j+1}\beta_{k,j}(\sum_{i=1}^{k} r_i)^j, & \text{for gated service,} \\ \lambda_k \sum_{j=1}^{m}(-1)^{j+1}\eta_{k,j}(\sum_{i\neq k} r_i)^j, & \text{for exhaustive service,} \end{cases} \tag{6.6}$$

and

$$\mathbf{p}(\mathbf{r}) := (p_1(\mathbf{r}), ..., p_K(\mathbf{r}))^T, \tag{6.7}$$

with

$$p_k(\mathbf{r}) := \xi_k(r_1, \ldots, r_k, p_{k+1}(\mathbf{r}), \ldots, p_K(\mathbf{r})),$$

and the iterates

$$\begin{aligned} \mathbf{p}^{(0)}(\mathbf{r}) &:= \mathbf{r}, \\ \mathbf{p}^{(i)}(\mathbf{r}) &:= \mathbf{p}(\mathbf{p}^{(i-1)}(\mathbf{r})), \quad i \geq 1. \end{aligned}$$

The function $P(\cdot)$ is defined by

$$P(\mathbf{r}) := \sum_{i=1}^{\infty}\sum_{k=1}^{K} p_k^{(i)}(\mathbf{r}). \tag{6.8}$$

In Lemma 6.2 we shall prove that the infinite sum in (6.8) is well-defined. Before we can do that we first need to prove Lemma 6.1. In the following we make the convention that $|\mathbf{v}| = (|v_1|, ..., |v_n|)^T$ where $\mathbf{v}$ is an n-dimensional vector and $\mathbf{v} \leq \mathbf{u}$ if and only if $v_k \leq u_k$ for all $k = 1, ..., n$. For the definition of $\tilde{\mathbf{M}}$, we refer to Appendix B.

Lemma 6.1 *There exists a* $\delta_1 > 0$ *such that* $\mathbf{p}(\mathbf{r}) \leq \tilde{\mathbf{M}}\mathbf{r}$ *for* $\mathbf{0} \leq \mathbf{r} \leq \delta_1\mathbf{e}$.

Proof. For $k = 1, ..., K$, it is easy to check that

$$\frac{\mathrm{d}}{\mathrm{d}s}\xi_k(\mathbf{e}_1 s) \leq \begin{cases} \left[\frac{\mathrm{d}}{\mathrm{d}s}\lambda_k(1-\beta_k(s))\right]_{s=0} = \rho_k, & \text{for gated service,} \\ \left[\frac{\mathrm{d}}{\mathrm{d}s}\lambda_k(1-\eta_k(s))\right]_{s=0} = \dfrac{\rho_k}{1-\rho_k}, & \text{for exhaustive service,} \end{cases}$$

for $0 < s < \delta_1$ where δ_1 is some positive constant. Therefore, we have

$$\begin{aligned} p_k(\mathbf{r}) &= \xi_k(r_1, ..., r_k, p_{k+1}(\mathbf{r}), , ..., p_K(\mathbf{r})) \\ &\leq \begin{cases} \rho_k[r_1 + ... + r_k + p_{k+1}(\mathbf{r}) + ... + p_K(\mathbf{r})], & \text{for gated service,} \\ \dfrac{\rho_k}{1-\rho_k}[r_1 + ... + r_{k-1} + p_{k+1}(\mathbf{r}) + ... + p_K(\mathbf{r})], & \text{for exhaustive service.} \end{cases} \end{aligned}$$

Rewriting the above inequalities in terms of matrices,

$$\mathbf{p}(\mathbf{r}) \leq \mathbf{B}\mathbf{r} + \mathbf{A}\mathbf{p}(\mathbf{r}), \tag{6.9}$$

where the matrices $\mathbf{A}$ and $\mathbf{B}$ are given by (8.8) and (8.9) respectively in Appendix B, it follows from the fact that $(\mathbf{I}-\mathbf{A})^{-1}$ is a nonnegative matrix that

$$\mathbf{p}(\mathbf{r}) \leq (\mathbf{I}-\mathbf{A})^{-1}\mathbf{B}\mathbf{r} = \tilde{\mathbf{M}}\mathbf{r}.$$

□

Now we are able to prove that the infinite sum in (6.8) is well-defined.

Lemma 6.2 *There exists a $\delta_1 > 0$ such that $P(\mathbf{e}_1 s) < \infty$ for $0 < s < \delta_1$.*

Proof. It follows from Lemma 6.1 that there exists $\delta_1 > 0$ such that for $0 < s < \delta_1$,

$$\mathbf{p}(\mathbf{e}_1 s) \leq \tilde{\mathbf{M}}\mathbf{e}_1 s. \tag{6.10}$$

Iterating (6.10) leads to

$$\mathbf{p}^{(i)}(\mathbf{e}_1 s) \leq \tilde{\mathbf{M}}^i \mathbf{e}_1 s, \quad i = 1, 2, \ldots.$$

Summing up the above relations, we get

$$\sum_{i=1}^{\infty} \mathbf{p}^{(i)}(\mathbf{e}_1 s) \leq (\mathbf{I}-\tilde{\mathbf{M}})^{-1}\tilde{\mathbf{M}}\mathbf{e}_1 s,$$

which implies that

$$P(\mathbf{e}_1 s) = \sum_{k=1}^{K}\sum_{i=1}^{\infty} p_k^{(i)}(\mathbf{e}_1 s) \leq \mathbf{e}^T(\mathbf{I}-\tilde{\mathbf{M}})^{-1}\tilde{\mathbf{M}}\mathbf{e}_1 s < \infty. \tag{6.11}$$

□

Actually, dividing by s in (6.10) and taking the limit for $s \downarrow 0$, we obtain

$$\left[\frac{\mathrm{d}}{s}\mathbf{p}(\mathbf{e}_1 s)\right]_{s=0} = \tilde{\mathbf{M}}\mathbf{e}_1.$$

Equality is seen to hold because the first inequality in the proof of Lemma 6.1 also reduces to an equality for $s \downarrow 0$. By using similar arguments as in the proof of Lemma 6.2, it is easy to derive from (6.11) that

$$P_{1,1} = \mathbf{e}^T(\mathbf{I}-\tilde{\mathbf{M}})^{-1}\tilde{\mathbf{M}}\mathbf{e}_1. \tag{6.12}$$

This relation is used in the proof of Corollary 5.2.

Step 2: The asymptotic expansion (6.3) is proved in the following lemma.

Lemma 6.3 *The function $P(\mathbf{e}_1 s)$ defined by (6.8) has the following expansion in the neighbourhood of the origin,*

$$P(\mathbf{e}_1 s) = \sum_{j=1}^{m} P_{1,j} s^j + \mathrm{O}(s^{m+1}), \quad \textit{for } s \downarrow 0. \tag{6.13}$$

Proof. First, we observe that for all $k = 1, ..., K$ and all $i = 1, 2, ...$, the functions $p_k^{(i)}(\mathbf{e}_1 s)$ are polynomials in s, i.e.,

$$p_k^{(i)}(\mathbf{e}_1 s) = \sum_{j=1}^{n_k^{(i)}} p_{k,j}^{(i)} s^j,$$

where $n_k^{(i)} = m^{Ki-k+1}$. It remains to prove that

$$\sum_{i=1}^{\infty} \sum_{k=1}^{K} \sum_{j=1}^{n_k^{(i)}} |p_{k,j}^{(i)}| s^j < \infty, \tag{6.14}$$

for $0 \leq s \leq \delta_2$. Because, if equation (6.14) holds, we can interchange the order of summation below,

$$P(\mathbf{e}_1 s) = \sum_{i=1}^{\infty} \sum_{k=1}^{K} \sum_{j=1}^{n_k^{(i)}} p_{k,j}^{(i)} s^j = \sum_{j=1}^{\infty} \left(\sum_{k=1}^{K} \sum_{\{i:n_k^{(i)} \geq j\}} p_{k,j}^{(i)} \right) s^j = \sum_{j=1}^{\infty} P_{1,j} s^j, \tag{6.15}$$

for $0 \leq s \leq \delta_2$. Therefore, the expansion (6.13) follows from (6.15) immediately.
In order to prove (6.14), we define a function $\mathbf{q} : \mathbf{R}_K \mapsto \mathbf{R}_K$,

$$\begin{aligned} \mathbf{q}(\mathbf{r}) &:= (q_1(\mathbf{r}), ..., q_K(\mathbf{r}))^T, \\ q_k(\mathbf{r}) &:= -\xi_k(-r_1, ..., -r_k, -q_{k+1}(\mathbf{r}), ..., -q_K(\mathbf{r})), \end{aligned} \tag{6.16}$$

and its iterates

$$\begin{aligned} \mathbf{q}^{(0)}(\mathbf{r}) &:= \mathbf{r}, \\ \mathbf{q}^{(i)}(\mathbf{r}) &:= \mathbf{q}(\mathbf{q}^{(i-1)}(\mathbf{r})), \quad i \geq 1. \end{aligned}$$

Next we show that the infinite sum $\sum_{i=1}^{\infty} \sum_{k=1}^{K} q_k^{(i)}(\mathbf{r})$ converges in a neighbourhood of the origin. By the definition of $\mathbf{q}(\mathbf{r})$, using similar arguments as in the proof of Lemma 6.1, it follows that for any $\epsilon > 0$, there exists $\delta_1 > 0$ such that, for $\mathbf{0} \leq \mathbf{r} \leq \delta_1 \mathbf{e}$,

$$\mathbf{0} \leq \mathbf{q}(\mathbf{r}) \leq (1 + \epsilon)\tilde{\mathbf{M}}\mathbf{r},$$

where the entries of $\tilde{\mathbf{M}}$ are given by (8.5). Let $a_{\max} < 1$ be the maximal eigenvalue of $\tilde{\mathbf{M}}$. Taking $\epsilon = (1/a_{\max} - 1)/2$, then the maximal eigenvalue of $(1 + \epsilon)\tilde{\mathbf{M}}$ is also less than 1. Thus, applying similar arguments as in the proof of Lemma 6.2, it follows that

$$\sum_{i=1}^{\infty} \sum_{k=1}^{K} q_k^{(i)}(\mathbf{r}) < \infty, \tag{6.17}$$

for $\mathbf{0} \leq \mathbf{r} \leq \delta_2 \mathbf{e}$ for some $\delta_2 > 0$. Similar as was observed for $p_k^{(i)}(\mathbf{e}_1 s)$, we see that also the functions $q_k^{(i)}(\mathbf{e}_1 s)$, for all k and i, are polynomials in s, i.e.,

$$q_k^{(i)}(\mathbf{e}_1 s) = \sum_{j=1}^{n_k^{(i)}} q_{k,j}^{(i)} s^j.$$

Furthermore, from the definition of $q_k^{(i)}(\mathbf{e}_1 s)$, it is easy to see that

$$|p_{k,j}^{(i)}| \le q_{k,j}^{(i)}, \tag{6.18}$$

for $k = 1, ..., K$, $i = 1, 2, ...$ and $j = 1, ..., n_k^{(i)}$. So, finally, (6.14) follows from (6.17) and (6.18). □

Step 3: Having proven (6.3), we must now prove (6.4). For this we need the following lemma.

Lemma 6.4 *For any $\epsilon > 0$, there exists $\delta_1 > 0$ such that for $\mathbf{0} \le \mathbf{r}, \mathbf{u} \le \delta_1 \mathbf{e}$,*

$$|\tilde{\mathbf{f}}(\mathbf{u}) - \mathbf{p}(\mathbf{r})| \le (\mathbf{I} - \mathbf{A})^{-1}(\mathbf{D} + \epsilon \mathbf{I})\mathbf{d} + \tilde{\mathbf{M}}|\mathbf{u} - \mathbf{r}|, \tag{6.19}$$

where $\mathbf{A}$ is given by (8.8) below, and

$$\mathbf{D} = \mathrm{diag}(\frac{\lambda_1 \beta_{1,\nu}}{\rho_1^\nu}, ..., \frac{\lambda_K \beta_{K,\nu}}{\rho_K^\nu}), \tag{6.20}$$

$$\mathbf{d} = ((\phi_1)^\nu L(1/\phi_1), ..., (\phi_K)^\nu L(1/\phi_K))^T, \tag{6.21}$$

with ϕ_k being the k-th component of $\tilde{\mathbf{M}}\mathbf{u}$ ($k = 1, ..., K$).

Proof. We only prove the case of gated service. By similar arguments, one can obtain the result for exhaustive service. For $k = 1, ..., K$, recall that $\tilde{h}_k(\cdot)$ and $\xi_k(\cdot)$ are defined by (4.25) and (6.6) respectively. Then we have, for $\mathbf{0} < \mathbf{u}, \mathbf{r} < \delta_1 \mathbf{e}$ where δ_1 is some positive constant,

$$\begin{aligned}
& |\tilde{f}_k(\mathbf{u}) - p_k(\mathbf{r})| \\
\le\ & |\tilde{h}_k(u_1, ..., u_k, \tilde{f}_{k+1}(\mathbf{u}), ..., \tilde{f}_K(\mathbf{u})) - \xi_k(u_1, ..., u_k, \tilde{f}_{k+1}(\mathbf{u}), ..., \tilde{f}_K(\mathbf{u}))| \\
& + |\xi_k(u_1, ..., u_k, \tilde{f}_{k+1}(\mathbf{u}), ..., \tilde{f}_K(\mathbf{u})) - \xi_k(r_1, ..., r_k, p_{k+1}(\mathbf{r}), ..., p_K(\mathbf{r}))| \\
\le\ & (\frac{\lambda_k \beta_{k,\nu}}{\rho_k^\nu} + \epsilon)(\rho_k u_1 + ... + \rho_k u_k + \rho_k \tilde{f}_{k+1}(\mathbf{u}) + ... + \rho_k \tilde{f}_K(\mathbf{u}))^\nu \\
& L(1/(\rho_k u_1 + ... + \rho_k u_k + \rho_k \tilde{f}_{k+1}(\mathbf{u}) + ... + \rho_k \tilde{f}_K(\mathbf{u}))) + \rho_k |u_1 - r_1| \\
& + ... + \rho_k |u_k - r_k| + \rho_k |\tilde{f}_{k+1}(\mathbf{u}) - p_{k+1}(\mathbf{r})| + ... + \rho_k |\tilde{f}_K(\mathbf{u}) - p_K(\mathbf{r})|,
\end{aligned} \tag{6.22}$$

where the last inequality in (6.22) follows from the fact that, cf. (5.5),

$$|1 - \beta_k(s) - \sum_{j=1}^{m} (-1)^{j+1} \beta_{k,j} s^j| \le \beta_{k,\nu} s^\nu L(1/s), \quad 0 < s < \delta,$$

δ being a positive constant. By similar arguments as in the proof of Lemma 6.1, one can easily prove that, for $\mathbf{0} < \mathbf{u} < \delta \mathbf{e}$,

$$\tilde{\mathbf{f}}(\mathbf{u}) \le \tilde{\mathbf{M}}\mathbf{u}.$$

Thus, it follows that

$$\mathbf{B}\mathbf{u} + \mathbf{A}\tilde{\mathbf{f}}(\mathbf{u}) \le \mathbf{B}\mathbf{u} + \mathbf{A}\tilde{\mathbf{M}}\mathbf{u} = (\mathbf{B} + \mathbf{A}(\mathbf{I} - \mathbf{A})^{-1}\mathbf{B})\mathbf{u} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\mathbf{u} = \tilde{\mathbf{M}}\mathbf{u},$$

which implies that

$$\rho_k u_1 + \ldots + \rho_k u_k + \rho_k \tilde{f}_{k+1}(\mathbf{u}) + \ldots + \rho_k \tilde{f}_K(\mathbf{u}) \leq \phi_k. \tag{6.23}$$

Rewriting the inequality (6.22) in terms of matrices and combining with (6.23), we obtain

$$|\tilde{\mathbf{f}}(\mathbf{u}) - \mathbf{p}(\mathbf{r})| \leq (\mathbf{D} + \epsilon\mathbf{I})\mathbf{d} + \mathbf{B}|\mathbf{u} - \mathbf{r}| + \mathbf{A}|\tilde{\mathbf{f}}(\mathbf{u}) - \mathbf{p}(\mathbf{r})|, \tag{6.24}$$

$\mathbf{D}$ and $\mathbf{d}$ being given by (6.20) and (6.21) respectively. Since $(\mathbf{I} - \mathbf{A})^{-1}$ is a nonnegative matrix, (6.19) follows from (6.24) immediately. □

Lemma 6.5 *There exists a nonnegative constant $H_{1,\nu}$ such that*

$$\lim_{s \downarrow 0} \frac{H(\mathbf{e}_1 s) - P(\mathbf{e}_1 s)}{s^\nu L(1/s)} = (-1)^m H_{1,\nu}. \tag{6.25}$$

The constant $H_{1,\nu} = 0$ if and only if $\sum_{k=1}^K \beta_{k,\nu} = 0$.

Proof. To simplify the notation, denote by r_{ik} the k-th component of the vector $\tilde{\mathbf{M}}^i \mathbf{e}_1$, put

$$\mathbf{v}_i(s) := ((r_{i1}s)^\nu L(1/r_{i1}s), \ldots, (r_{iK}s)^\nu L(1/r_{iK}s))^T, \tag{6.26}$$

and let $v_{ik}(s)$ denote the k-th component of $\mathbf{v}_i(s)$ where $k = 1, \ldots, K$, $i = 1, 2, \ldots$. By Lemma 6.4 it follows that

$$|\tilde{\mathbf{f}}^{(i)}(\mathbf{e}_1 s) - \mathbf{p}^{(\mathbf{i})}(\mathbf{e}_1 s)| \leq (\mathbf{I} - \mathbf{A})^{-1}(\mathbf{D} + \epsilon\mathbf{I})\mathbf{v}_{\mathbf{i}}(s) + \tilde{\mathbf{M}}|\tilde{\mathbf{f}}^{(i-1)}(\mathbf{e}_1 s) - \mathbf{p}^{(i-1)}(\mathbf{e}_1 s)|, \tag{6.27}$$

for $i = 1, 2, \ldots$. Iterating the above relations, we get for $i = 1, 2, \ldots,$

$$|\tilde{\mathbf{f}}^{(i)}(\mathbf{e}_1 s) - \mathbf{p}^{(i)}(\mathbf{e}_1 s)| \leq \sum_{j=1}^{i} \tilde{\mathbf{M}}^{i-j}(\mathbf{I} - \mathbf{A})^{-1}(\mathbf{D} + \epsilon\mathbf{I})\mathbf{v}_j(s). \tag{6.28}$$

Summing up the above inequalities yields

$$\begin{aligned} \frac{\sum_{i=1}^\infty |\tilde{\mathbf{f}}^{(i)}(\mathbf{e}_1 s) - \mathbf{p}^{(i)}(\mathbf{e}_1 s)|}{s^\nu L(1/s)} &\leq \sum_{i=1}^\infty \sum_{j=1}^{i} \tilde{\mathbf{M}}^{i-j}(\mathbf{I} - \mathbf{A})^{-1}(\mathbf{D} + \epsilon\mathbf{I})\frac{\mathbf{v}_j(s)}{s^\nu L(1/s)} \\ &= \sum_{j=1}^\infty \sum_{i=j}^{\infty} \tilde{\mathbf{M}}^{i-j}(\mathbf{I} - \mathbf{A})^{-1}(\mathbf{D} + \epsilon\mathbf{I})\frac{\mathbf{v}_j(s)}{s^\nu L(1/s)} \\ &= (\mathbf{I} - \tilde{\mathbf{M}})^{-1}(\mathbf{I} - \mathbf{A})^{-1}(\mathbf{D} + \epsilon\mathbf{I})\sum_{j=1}^\infty \frac{\mathbf{v}_j(s)}{s^\nu L(1/s)}, \end{aligned} \tag{6.29}$$

where the last identity follows from (8.10) in Appendix B.

Next we prove that the infinite sum $\sum_{i=1}^\infty \mathbf{v}_i(s)/(s^\nu L(1/s))$ converges. By using Potter's theorem (cf. Theorem 1.5.6 in [6]), it follows from the fact that $\lim_{i \to \infty} r_{ik} = 0$ for $k = 1, \ldots, K$ that $\frac{r_{ik}^{\nu-1} L(1/r_{ik}s)}{L(1/s)}$ converges to 0 uniformly in s for $s > 0$ as $i \to \infty$. Thus there exists N_0 such that for $i \geq N_0$, $k = 1, \ldots, K$,

$$\frac{v_{ik}(s)}{s^\nu L(1/s)} \leq r_{ik}.$$

By the definition of $v_{ik}(s)$ and r_{ik}, we have for $k = 1, ..., K$, $0 < s < \delta$ where δ is some positive constant,

$$\sum_{i=N_0}^{\infty}\sum_{k=1}^{K}\frac{v_{ik}(s)}{s^{\nu}L(1/s)} \leq \sum_{i=N_0}^{\infty}\sum_{k=1}^{K} r_{ik} \leq \sum_{i=1}^{\infty}\sum_{k=1}^{K} r_{ik} = \sum_{i=1}^{\infty}\mathbf{e}^T\tilde{\mathbf{M}}^k\mathbf{e}_1 = \mathbf{e}^T(\mathbf{I}-\tilde{\mathbf{M}})^{-1}\tilde{\mathbf{M}}\mathbf{e}_1 < \infty.$$

Hence, applying the Dominated Convergence Theorem, it follows that

$$\begin{aligned}\lim_{s\downarrow 0}\frac{H(\mathbf{e}_1 s)-P(\mathbf{e}_1 s)}{s^{\nu}L(1/s)} &= \sum_{i=1}^{\infty}\lim_{s\downarrow 0}\frac{\mathbf{e}^T(\tilde{\mathbf{f}}^{(i)}(\mathbf{e}_1 s)-\mathbf{p}^{(i)}(\mathbf{e}_1 s))}{s^{\nu}L(1/s)}\\ &= (-1)^m\mathbf{e}^T(\mathbf{I}-\tilde{\mathbf{M}})^{-1}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{D}\sum_{i=1}^{\infty}\lim_{s\downarrow 0}\frac{\mathbf{v}_i(s)}{s^{\nu}L(1/s)}\\ &= (-1)^m\mathbf{e}^T(\mathbf{I}-\tilde{\mathbf{M}})^{-1}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{D}\sum_{i=1}^{\infty}\mathbf{u}_i < \infty, \qquad (6.30)\end{aligned}$$

where

$$\mathbf{u}_i := (r_{i1}^{\nu}, ..., r_{iK}^{\nu}), \qquad (6.31)$$

and the last identity follows from Lemma 8.2. Put

$$H_{1,\nu} = \mathbf{e}^T(\mathbf{I}-\tilde{\mathbf{M}})^{-1}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{D}\sum_{i=1}^{\infty}\mathbf{u}_i, \qquad (6.32)$$

and subsequently (6.25) follows. Noticing that $\mathbf{D} = \mathbf{0}$ if and only if $\sum_{k=1}^{K}\beta_{k,\nu} = 0$, we conclude that $H_{1,\nu} = 0$ if and only if $\sum_{k=1}^{K}\beta_{k,\nu} = 0$. □

2. The case of non-zero switchover time

Again we wish to prove (5.8) for $x_1(s)$. As shown in Section 4, $x_1(s) = \tilde{F}_1(\mathbf{e}_1 s)$ where $\tilde{F}_1(\cdot)$ is given by (4.33). Put

$$\mathbf{C} := \begin{pmatrix} 1 & 0 & ... & 0 & 0\\ 1 & 1 & ... & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ 1 & 1 & ... & 1 & 0\\ 1 & 1 & ... & 1 & 1\end{pmatrix}, \quad \mathbf{G} := \begin{pmatrix} 0 & 1 & ... & 1 & 1\\ 0 & 0 & ... & 1 & 1\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & ... & 0 & 1\\ 0 & 0 & ... & 0 & 0\end{pmatrix}, \qquad (6.33)$$

and subsequently we define

$$\begin{aligned}\hat{\mathbf{f}}^{(i)}(\mathbf{r}) &:= \mathbf{C}\tilde{\mathbf{f}}^{(i)}(\mathbf{r}) + \mathbf{G}\tilde{\mathbf{f}}^{(i+1)}(\mathbf{r}), \quad i = 1, 2, ...,\\ \hat{F}(\mathbf{r}) &:= \sum_{i=1}^{\infty}\sum_{k=1}^{K}\ln(\sigma_k(\hat{f}_k^{(i)}(\mathbf{r}))), \qquad (6.34)\end{aligned}$$

with $\hat{f}_k^{(i)}(\mathbf{r})$ being the k-th component of $\hat{\mathbf{f}}^{(i)}(\mathbf{r})$ for $k = 1, ..., K$. So we may rewrite (4.33) as

$$\tilde{F}(\mathbf{r}) = \exp\{\hat{F}(\mathbf{r})\}, \qquad (6.35)$$

where $\hat{F}(\mathbf{r})$ is given by (6.34). To prove (5.8), it is sufficient to show that

$$\hat{F}(\mathbf{e}_1 s) = \sum_{j=1}^{m}\hat{F}_{1,j}s^j + (-1)^{m+1}\hat{F}_{1,\nu}s^{\nu}L(1/s) + \mathrm{o}(s^{\nu}L(1/s)), \quad s \downarrow 0. \qquad (6.36)$$

We shall use similar arguments as in the proof for zero switchover times to obtain (6.36). We have again divided the proof into three steps. In the first step, we construct a new function $\hat{P}(\mathbf{r})$ which has similar structure as $\hat{F}(\mathbf{r})$. In the second step, we shall use similar arguments as in the proof of zero switchover time to show that

$$\hat{P}(\mathbf{e}_1 s) = \sum_{j=1}^{m} \hat{P}_{1,j} s^j + (-1)^{m+1} \hat{P}_{1,\nu} s^\nu L(1/s) + \mathrm{o}(s^\nu L(1/s)), \tag{6.37}$$

where $\hat{P}_{1,j}$ are some constants for $j = 0, ..., m$ and $\hat{P}_{1,\nu} \geq 0$. Moreover, $\hat{P}_{1,\nu} = 0$ if and only if $\sum_{k=1}^{K} \sigma_{k,\nu} = 0$. In the third step, we shall verify that

$$\lim_{s \downarrow 0} \frac{\hat{F}(\mathbf{e}_1 s) - \hat{P}(\mathbf{e}_1 s)}{s^\nu L(1/s)} = (-1)^{m+1} g_{1,\nu}, \tag{6.38}$$

where $g_{1,\nu} \geq 0$. Moreover, $g_{1,\nu} = 0$ if and only if $\sum_{k=1}^{K} (\beta_{k,\nu} + \sigma_{k,\nu}) = 0$. Obviously, combining (6.37) and (6.38) yields (6.36) where

$$\hat{F}_{1,\nu} = \hat{P}_{1,\nu} + g_{1,\nu} \tag{6.39}$$

with $\hat{P}_{1,\nu}$ and $g_{1,\nu}$ being given by (6.48) and (6.51). Then applying Lemma 8.3 in Appendix A, and noting that $x_1(s) = \tilde{F}_1(\mathbf{e}_1 s)$ is the LST of some non-negative random variable, Formula (5.8) of Theorem 5.1 follows from (6.35) and (6.36) with

$$x_{1,1} = \hat{P}_{1,1}, \quad x_{1,\nu} = \hat{F}_{1,\nu}, \tag{6.40}$$

$\hat{P}_{1,1}$ and $\hat{F}_{1,\nu}$ being given by (6.44) and (6.39) respectively.
Step 1: Define:

$$\hat{\mathbf{p}}^{(i)}(\mathbf{r}) \;:=\; \mathbf{C}\mathbf{p}^{(i)}(\mathbf{r}) + \mathbf{G}\mathbf{p}^{(i+1)}(\mathbf{r}), \quad i = 1, 2, ..., \tag{6.41}$$

$$\hat{P}(\mathbf{r}) \;:=\; \sum_{i=1}^{\infty} \sum_{k=1}^{K} \ln(\sigma_k(\hat{p}_k^{(i)}(\mathbf{r}))), \tag{6.42}$$

$\mathbf{p}^{(i)}(\cdot)$ being given by (6.7) and $\hat{p}_k^{(i)}(\cdot)$ denoting the k-th component of $\hat{\mathbf{p}}^{(i)}(\cdot)$. By Lemma 6.1, we have

$$\begin{aligned} \hat{\mathbf{p}}^{(i)}(\mathbf{e}_1 s) &= \mathbf{C}\mathbf{p}^{(i)}(\mathbf{e}_1 s) + \mathbf{G}\mathbf{p}^{(i+1)}(\mathbf{e}_1 s) \\ &\leq \mathbf{C}\tilde{\mathbf{M}}^i \mathbf{e}_1 s + \mathbf{G}\tilde{\mathbf{M}}^{i+1} \mathbf{e}_1 s \\ &= (\mathbf{C} + \mathbf{G}\tilde{\mathbf{M}})\tilde{\mathbf{M}}^i \mathbf{e}_1 s, \end{aligned} \tag{6.43}$$

where the above inequality follows from (6.10). Using the fact that $\ln(\sigma_k(x)) < \sigma_k x$ for small x, one can easily prove that $\hat{P}(\mathbf{r})$ is well-defined in some neighbourhood of the origin. It is not difficult to see that

$$\lim_{s \downarrow 0} \frac{\mathrm{d}}{\mathrm{d}s} \hat{\mathbf{p}}^{(i)}(\mathbf{e}_1 s) = (\mathbf{C} + \mathbf{G}\tilde{\mathbf{M}})\tilde{\mathbf{M}}^i \mathbf{e}_1.$$

It follows that

$$\hat{P}_{1,1} = \mathbf{e}^T \mathbf{H}(\mathbf{C} + \mathbf{G}\tilde{\mathbf{M}})(\mathbf{I} - \tilde{\mathbf{M}})^{-1} \tilde{\mathbf{M}} \mathbf{e}_1. \tag{6.44}$$

Step 2: We shall prove (6.37) by using similar arguments as in the proof for the case with zero switchover time. We omit some of the details here. Firstly by using Lemma 8.3, we may write

$$\ln(\sigma_k(x)) = \sum_{j=1}^{m} a_{k,j}x^j + (-1)^{m+1}a_{k,\nu}x^\nu L(1/x) + \mathrm{o}(x^\nu L(1/x)), \quad x \downarrow 0, \tag{6.45}$$

with $a_{k,\nu} = \sigma_{k,\nu}$. Applying similar arguments as in the proof of Lemma 6.3, we can easily verify that

$$A(s) := \sum_{i=1}^{\infty}\sum_{k=1}^{K}\sum_{j=1}^{m} a_{k,j}(\hat{p}_k^{(i)}(\mathbf{e}_1 s))^j = \sum_{j=1}^{\infty} A_j s^j. \tag{6.46}$$

For any $\epsilon > 0$, there exists a $\delta > 0$ such that for $0 < s < \delta$,

$$\begin{aligned} \frac{|\hat{P}(\mathbf{e}_1 s) - A(s)|}{s^\nu L(1/s)} &\leq \sum_{i=1}^{\infty}\sum_{k=1}^{K} \frac{(\sigma_{k,\nu}+\epsilon)(\hat{p}_k^{(i)}(\mathbf{e}_1 s))^\nu L(1/(\hat{p}_k^{(i)}(\mathbf{e}_1 s)))}{s^\nu L(1/s)} \\ &\leq \sum_{i=1}^{\infty}\sum_{k=1}^{K} \frac{(\sigma_{k,\nu}+\epsilon)(\alpha_{ik}s)^\nu L(1/\alpha_{ik}s)}{s^\nu L(1/s)} < \infty, \end{aligned} \tag{6.47}$$

α_{ik} denoting the k-th component of the vector $(\mathbf{C} + \mathbf{G}\tilde{\mathbf{M}})\tilde{\mathbf{M}}^i\mathbf{e}_1$. By the Dominated Convergence Theorem, it can be shown that

$$\lim_{s\downarrow 0} \frac{\hat{P}(\mathbf{e}_1 s) - A(s)}{s^\nu L(1/s)} = (-1)^{m+1}\hat{P}_{1,\nu}, \tag{6.48}$$

with

$$\hat{P}_{1,\nu} = \sum_{i=1}^{\infty}\sum_{k=1}^{K} \sigma_{k,\nu}\alpha_{ik}^\nu.$$

Notice that $\hat{P}_{1,\nu} = 0$ if and only if $\sum_{k=1}^{K}\sigma_{k,\nu} = 0$. Combining (6.48) and (6.46) leads to (6.37).

Step 3: The proof of (6.38) is similar to that of Lemma 6.5. Here we omit some of the details. For simplicity, define

$$\mathbf{H} := \mathrm{diag}(\sigma_1, ..., \sigma_K).$$

By the definitions of $\hat{F}(\mathbf{e}_1 s)$ and $\hat{P}(\mathbf{e}_1 s)$, we have

$$\begin{aligned} &|\hat{F}(\mathbf{e}_1 s) - \hat{P}(\mathbf{e}_1 s)| \\ = &\sum_{i=1}^{\infty}\sum_{k=1}^{K} |\ln(\sigma_k(\hat{p}_k^{(i)}(\mathbf{e}_1 s))) - \ln(\sigma_k(\hat{f}_k^{(i)}(\mathbf{e}_1 s)))| \\ \leq &\sum_{i=1}^{\infty}\sum_{k=1}^{K} \sigma_k|\hat{p}_k^{(i)}(\mathbf{e}_1 s) - \hat{f}_k^{(i)}(\mathbf{e}_1 s)| \\ \leq &\sum_{i=1}^{\infty}\sum_{k=1}^{K}(\sigma_k\sum_{j=1}^{k}|\tilde{f}_k^{(i)}(\mathbf{e}_1 s) - p_k^{(i)}(\mathbf{e}_1 s)| + \sigma_k\sum_{j=k+1}^{K}|\tilde{f}_k^{(i+1)}(\mathbf{e}_1 s) - p_k^{(i+1)}(\mathbf{e}_1 s)|) \\ = &\sum_{i=1}^{\infty}\mathbf{e}^T\mathbf{H}(\mathbf{C}|\tilde{\mathbf{f}}^{(i)}(\mathbf{e}_1 s) - \mathbf{p}^{(i)}(\mathbf{e}_1 s)| + \mathbf{G}|\tilde{\mathbf{f}}^{(i+1)}(\mathbf{e}_1 s) - \mathbf{p}^{(i+1)}(\mathbf{e}_1 s)|) \\ \leq &\ \mathbf{e}^T\mathbf{H}(\mathbf{C}+\mathbf{G})\sum_{i=1}^{\infty}|\tilde{\mathbf{f}}^{(i)}(\mathbf{e}_1 s) - \mathbf{p}^{(i)}(\mathbf{e}_1 s)|, \end{aligned} \tag{6.49}$$

which in combination with (6.29) yields

$$\begin{aligned} & \frac{|\hat{F}(\mathbf{e}_1 s) - \hat{P}(\mathbf{e}_1 s)|}{s^\nu L(1/s)} \\ \leq \; & \mathbf{e}^T \mathbf{H}(\mathbf{C} + \mathbf{G})(\mathbf{I} - \tilde{\mathbf{M}})^{-1}(\mathbf{I} - \mathbf{A})^{-1}(\mathbf{D} + \epsilon \mathbf{I}) \sum_{i=1}^{\infty} \frac{\mathbf{v}_j(s)}{s^\nu L(1/s)} < \infty, \end{aligned} \tag{6.50}$$

where $\mathbf{v}_i(s)$ is defined by (6.26). Again, using the Dominated Convergence Theorem, we obtain

$$\lim_{s \downarrow 0} \frac{\hat{F}(\mathbf{e}_1 s) - \hat{P}(\mathbf{e}_1 s)}{s^\nu L(1/s)} = (-1)^{m+1} g_{1,\nu},$$

with

$$g_{1,\nu} = \mathbf{e}^T \mathbf{H}(\mathbf{C} + \mathbf{G}\tilde{\mathbf{M}})(\mathbf{I} - \tilde{\mathbf{M}})^{-1}(\mathbf{I} - \mathbf{A})^{-1} \mathbf{D} \sum_{i=1}^{\infty} \mathbf{u}_i, \tag{6.51}$$

where $\mathbf{u}_i$ is given in (6.31). Notice that $\mathbf{H} = \mathbf{D} = \mathbf{0}$ if and only if $\sum_{k=1}^{K} (\beta_{k,\nu} + \sigma_{k,\nu}) = 0$, thus $g_{1,\nu} = 0$ if and only if $\sum_{k=1}^{K} (\beta_{k,\nu} + \sigma_{k,\nu}) = 0$.

7 Conclusions

In this paper we have investigated the tail behaviour of the waiting time distributions in cyclic polling systems with gated or exhaustive service. Under the assumption that at least one of the service or switchover time distributions has a regularly varying tail, the waiting time distributions at all queues are shown to be regularly varying at infinity, of index one higher than the heaviest tail of the service and switchover time distributions. This result gives important insight into the effect of heavy-tailed service or switchover time distributions on the performance of a large class of polling systems. We expect the same result to be true for non-cyclic polling systems, and for a larger class of arrival processes and service disciplines. For the class of service disciplines satisfying Property 4.1, it may be possible to prove this along similar lines as in the present paper. For almost all polling systems in which the service discipline in at least one queue does not satisfy Property 4.1, no explicit expression for the waiting time LST's is known, so that the approach via Lemma 8.1 does not work. An exception is provided by the 2-queue polling system with exhaustive service at Q_1 and 1-limited service at Q_2; for this model, a simple explicit expression for the waiting time LST's is known (see, e.g., [23]), which can be used to prove the above predicted result for the waiting time tails.

8 Appendix

Appendix A: Preliminaries

In this appendix we shall introduce some basic results relating the tail behaviour of probability distributions to the asymptotic behaviour of their LSTs near the origin. Regularly varying distributions receive special attention.

The following lemma (cf. Lemma 2.2 in [12]), which is an extension of Theorem 8.1.6 in [6], links the regularly varying tail behaviour of $\Pr\{X > t\}$ for $t \to \infty$ to the behaviour of its LST $f(s)$ for $s \downarrow 0$. It plays a key role in the proof of our main result. First introduce, for an LST

$f(s)$ of the probability distribution of a non-negative random variable with finite j-th moment γ_j, $j = 0, 1, \ldots, n$:

$$f_n(s) := (-1)^{n+1}\left(f(s) - \sum_{j=0}^{n} \gamma_j \frac{(-s)^j}{j!}\right), \quad s \geq 0.$$

Lemma 8.1 *Let X be a random variable with LST $f(s)$, $L(t)$ a slowly varying function, $\nu \in (n, n+1)$ $(n \in \mathbf{N})$ and $C \geq 0$. Then the following are equivalent:*
(i) $\Pr\{X > t\} = [C + \mathrm{o}(1)]L(t)/t^{\nu}$, $t \to \infty$.
(ii) $E\{X^n\} < \infty$ *and* $f_n(s) = (-1)^n\Gamma(1-\nu)[C + \mathrm{o}(1)]L(1/s)s^{\nu}$, $s \downarrow 0$.

The next lemma (cf. Lemma 5 in [11]) characterizes a property of slowly varying functions.

Lemma 8.2 *Let $L(x)$ be a slowly varying function, and $t(x)$ a positive function such that $\lim_{x\to\infty} t(x)/x = a$ where $0 < a < \infty$. Then for a constant ν $(\nu \in \mathbf{R})$,*

$$\lim_{x\to\infty} \frac{\{t(x)\}^{\nu} L(t(x))}{x^{\nu} L(x)} = a^{\nu}.$$

The key formulas of the present study involve iterated functions (see, e.g., (4.16), (4.23) and (5.2)). The following result is useful in this respect; it is a consequence of Lemma 8.2.

Lemma 8.3 *Suppose $\phi(\cdot)$, $\psi(\cdot)$ can be written as*

$$\phi(x) = \sum_{i=1}^{n} \phi_i x^i + \phi_\nu x^\nu L(1/x) + \mathrm{o}(x^\nu L(1/x)), \quad \textit{for } x \downarrow 0, \tag{8.1}$$

$$\psi(x) = \sum_{i=1}^{n} \psi_i x^i + \psi_\nu x^\nu L(1/x) + \mathrm{o}(x^\nu L(1/x)), \quad \textit{for } x \downarrow 0, \tag{8.2}$$

where $\phi_1, \psi_1 > 0$, $n < \nu < n+1$ and $L(\cdot)$ is a slowly varying function. Then the asymptotic expansion of the function $\phi(\psi(x))$ at point 0 is given by

$$\phi(\psi((x)) = \sum_{i=1}^{n} \theta_i x^i + (\phi_1\psi_\nu + \phi_\nu\psi_1^\nu) x^\nu L(1/x) + \mathrm{o}(x^\nu L(1/x)), \quad \textit{for } x \downarrow 0. \tag{8.3}$$

Proof. For $1 \leq i \leq n$, there exist polynomials $p_i(x)$ and $q_{i,j}(x)$ $(j = 1, ..., i)$ such that $(\psi(x))^i$ can be written as

$$(\psi(x))^i = p_i(x) + \sum_{j=1}^{i} (x^\nu L(1/x))^j q_{i,j}(x) + \mathrm{o}(x^\nu L(1/x)),$$

where

$$p_i(x) = (\sum_{j=1}^{n} \psi_j x^j)^i,$$

$$q_{i,j}(x) = \binom{i}{j} \phi_\nu^j (p_i(x))^{i-j}, \quad j = 1, ..., i.$$

Note that $q_{i,1}(x)$ are all equal to 0 if $x = 0$ for $2 \leq i \leq n$. Therefore, we have

$$\sum_{i=1}^{n} \phi_i(\psi(x))^i = \sum_{j=1}^{n} a_j x^j + \phi_1 \psi_\nu x^\nu L(1/x) + \mathrm{o}(x^\nu L(1/x)), \tag{8.4}$$

for some real numbers a_j ($j = 1, ..., n$). Since $\lim_{x \downarrow 0} \psi(x)/x = \psi_1$, it follows from Lemma 8.2 that

$$\lim_{x \downarrow 0} \frac{(\psi(x))^\nu L(1/\psi(x))}{x^\nu L(1/x)} = \psi_1^\nu,$$

which in combination with (8.1) and (8.4) leads to the conclusion. □

Remark 7.1. It should be noted that, despite the symmetry in (8.1) and (8.2), it is possible that $\phi(x)$ refers to a heavier-tailed function than $\psi(x)$ (or vice versa); for example, ψ_ν might be equal to zero.

Appendix B: On the first moment matrix

Consider the mean matrix $\mathbf{M} = (m_{kj} : k, j = 1, ..., K)$, where

$$m_{kj} := \frac{\partial f_k}{\partial z_j}(1, ..., 1),$$

is the mean number of type-j customers that are descendants of a single type-k customer. As proved in [27], $\mathbf{M}$ plays an essential role in proving that $\rho < 1$ is sufficient for ergodicity in the case of gated or exhaustive service. In this appendix we shall derive some properties of the matrix $\tilde{\mathbf{M}} = (\tilde{m}_{kj} : k, j = 1, ...K)$, where

$$\tilde{m}_{kj} := \frac{\partial \tilde{f}_k}{\partial r_j}(0, ..., 0). \tag{8.5}$$

The following lemma relates the eigenvalues and eigenvectors of $\mathbf{M}$ and $\tilde{\mathbf{M}}$.

Lemma 8.4 *The eigenvalues of* $\mathbf{M}$ *and* $\tilde{\mathbf{M}}$ *are identical. Moreover, if* $\mathbf{v} = (v_1, ..., v_K)^T$ *is a right eigenvector of* $\mathbf{M}$ *w.r.t. eigenvalue a, then* $\mathbf{u} = (\lambda_1 v_1, ..., \lambda_k v_K)^T$ *is a right eigenvector of* $\tilde{\mathbf{M}}$ *w.r.t. a.*

Proof. Elementary, using the fact that

$$\tilde{m}_{kj} = \frac{\lambda_k}{\lambda_j} m_{kj}, \tag{8.6}$$

which follows from the relation (see (4.12), (4.25) and (4.26))

$$\tilde{f}_k(r) = \lambda_k(1 - f_k(z)).$$

□

Furthermore, we can derive an explicit formula for $\tilde{\mathbf{M}}$. It follows from (4.26) that

$$\tilde{\mathbf{M}} = \mathbf{B} + \mathbf{A}\tilde{\mathbf{M}}, \tag{8.7}$$

where

$$\mathbf{A} = \begin{pmatrix} 0 & \frac{\partial \tilde{h}_1}{\partial r_2}(\mathbf{0}) & \dots & \frac{\partial \tilde{h}_1}{\partial r_{K-1}}(\mathbf{0}) & \frac{\partial \tilde{h}_1}{\partial r_K}(\mathbf{0}) \\ 0 & 0 & \dots & \frac{\partial \tilde{h}_2}{\partial r_{K-1}}(\mathbf{0}) & \frac{\partial \tilde{h}_2}{\partial r_K}(\mathbf{0}) \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \frac{\partial \tilde{h}_{K-1}}{\partial r_K}(\mathbf{0}) \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}, \tag{8.8}$$

$$\mathbf{B} = \begin{pmatrix} \frac{\partial \tilde{h}_1}{\partial r_1}(\mathbf{0}) & 0 & \dots & 0 & 0 \\ \frac{\partial \tilde{h}_2}{\partial r_1}(\mathbf{0}) & \frac{\partial \tilde{h}_2}{\partial r_2}(\mathbf{0}) & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ \frac{\partial \tilde{h}_{K-1}}{\partial r_1}(\mathbf{0}) & \frac{\partial \tilde{h}_{K-1}}{\partial r_2}(\mathbf{0}) & \dots & \frac{\partial \tilde{h}_{K-1}}{\partial r_{K-1}}(\mathbf{0}) & 0 \\ \frac{\partial \tilde{h}_K}{\partial r_1}(\mathbf{0}) & \frac{\partial \tilde{h}_K}{\partial r_2}(\mathbf{0}) & \dots & \frac{\partial \tilde{h}_K}{\partial r_{K-1}}(\mathbf{0}) & \frac{\partial \tilde{h}_K}{\partial r_K}(\mathbf{0}) \end{pmatrix}. \tag{8.9}$$

For the case of gated service at all queues, **A** and **B** are given by

$$\mathbf{A}_{\text{gat}} = \begin{pmatrix} 0 & \rho_1 & \dots & \rho_1 & \rho_1 \\ 0 & 0 & \dots & \rho_2 & \rho_2 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \rho_{K-1} \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}, \mathbf{B}_{\text{gat}} = \begin{pmatrix} \rho_1 & 0 & \dots & 0 & 0 \\ \rho_2 & \rho_2 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ \rho_{K-1} & \rho_{K-1} & \dots & \rho_{K-1} & 0 \\ \rho_K & \rho_K & \dots & \rho_K & \rho_K \end{pmatrix},$$

and for the case of exhaustive service at all queues, **A** and **B** are given by

$$\mathbf{A}_{\text{exh}} = \begin{pmatrix} 0 & \frac{\rho_1}{1-\rho_1} & \dots & \frac{\rho_1}{1-\rho_1} & \frac{\rho_1}{1-\rho_1} \\ 0 & 0 & \dots & \frac{\rho_2}{1-\rho_2} & \frac{\rho_2}{1-\rho_2} \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \frac{\rho_{K-1}}{1-\rho_{K-1}} \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}, \mathbf{B}_{\text{exh}} = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ \frac{\rho_2}{1-\rho_2} & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ \frac{\rho_{K-1}}{1-\rho_{K-1}} & \frac{\rho_{K-1}}{1-\rho_{K-1}} & \dots & 0 & 0 \\ \frac{\rho_K}{1-\rho_K} & \frac{\rho_K}{1-\rho_K} & \dots & \frac{\rho_K}{1-\rho_K} & 0 \end{pmatrix}.$$

From equation (8.7) we get that

$$\tilde{\mathbf{M}} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}.$$

If $\rho < 1$ then the largest eigenvalue $a_{\max} < 1$ (see [27]) and it can be readily shown that $\lim_{n\to\infty} \tilde{\mathbf{M}}^n = \mathbf{0}$. Thus, applying Lemma B.1 in [28], we have

$$(\mathbf{I} - \tilde{\mathbf{M}})^{-1} = \sum_{i=0}^{\infty} \tilde{\mathbf{M}}^i, \tag{8.10}$$

which is a nonnegative matrix.

Acknowledgement The authors are indebted to S.C. Borst for interesting discussions.

References

[1] J. Abate, and W. Whitt (1997). *Asymptotics for $M/G/1$ low-priority waiting-time tail probabilities.* Queueing Systems **25**, 173-233.

[2] E. Altman, P. Konstantopoulos, and Z. Liu (1992). *Stability, monotonicity and invariant quantities in general polling systems.* Queueing Systems **11**, 35-57.

[3] S. Asmussen, C. Klüppelberg, and K. Sigman (1999). *Sampling at subexponential times, with queueing applications.* Stochastic Processes and their Applications.

[4] J. Beran, R. Sherman, M.S. Taqqu, and W. Willinger (1995). *Long-range dependence in variable-bit-rate video.* IEEE Transactions on Communications **43**, 1566-1579.

[5] D. Bertsimas, and G. Mourtzinou (1997). *Decomposition results for general polling systems and their applications.* Report MIT.

[6] N.H. Bingham, C.M. Goldie, and J.L. Teugels (1987). *Regular Variation.* Cambridge University Press, Cambridge.

[7] S.C. Borst (1994). *Polling Systems.* Ph.D. thesis, Tilburg University.

[8] S.C. Borst, and O.J. Boxma (1995). *Polling models with and without switchover times.* Operations Research **45**, 536-543.

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

[10] O.J. Boxma, and J.W. Cohen (1998). *The $M/G/1$ queue: heavy tails and heavy traffic.* To appear in K. Park and W. Willinger (eds.)

[11] O.J. Boxma, J.W. Cohen, and Q. Deng (1999). *Heavy-traffic analysis of the $M/G/1$ queue with priority classes.* In: *Teletraffic Engineering in a Competitive World*, Proc. ITC-16, eds. P. Key and D. Smith (North-Holland, Amsterdam), 1157-1167.

[12] O.J. Boxma, and V. Dumas (1998). *The busy period in the fluid queue.* Performance Evaluation Review (Proceeding of ACM Sigmetrics/Performance '98) **26**, 100-110.

[13] O.J. Boxma, and W.P. Groenendijk (1987). *Pseudo-conservation laws in cyclic service systems.* J. Appl. Probab. **24**, 949-964.

[14] G.L. Choudhury, and W. Whitt (1996). *Computing distributions and moments in polling models by numerical transform inversion.* Performance Evaluation **25**, 267-292.

[15] J.W. Cohen (1973). *Some results on regular variation for distributions in queueing and fluctuation theory.* J. Appl. Probab. **10**, 343-353.

[16] R.B. Cooper (1972). *Introduction to Queueing Theory.* Macmillan, London.

[17] A. De Meyer, and J.L. Teugels (1980). *On the asymptotic behaviour of the distributions of the busy period and service-time in M/G/1.* J. Appl. Probab. **17**, 802-813.

[18] N.G. Duffield (1997). *Exponents for the tail of distributions in some polling models.* Queueing Systems **26**, 105-119.

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

[20] C. Fricker, and M.R. Jaibi (1994). *Monotonicity and stability of periodic polling models.* Queueing Systems **15**, 211-238.

[21] S.W. Fuhrmann, and R.B. Cooper (1985). *Stochastic decompositions in the M/G/1 queue with generalized vacations.* Operations Research **33**, 1117-1129.

[22] J. Grandell (1997). *Mixed Poisson Processes.* Chapman & Hall, London.

[23] W.P. Groenendijk (1990). *Conservation Laws in Polling Systems.* Ph.D. Thesis, University of Utrecht.

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

[25] W.E. Leland, M.S. Taqqu, W. Willinger, and D.V. Wilson (1994). *On the self-similar nature of Ethernet traffic* (extended version). IEEE/ACM Transactions on Networking **2**, 1-15.

[26] A.G. Pakes (1975). *On the tails of waiting-time distributions.* J. Appl. Probab. **12**, 555-564.

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

[28] E. Seneta (1981). *Non-negative Matrices and Markov Chains,* (2nd ed.). Springer-Verlag, New York.

[29] K. Sigman (1999). *A primer on heavy-tailed distributions.* To appear in Queueing Systems.

[30] M.M. Srinivasan, S.-C. Niu, and R.B. Cooper (1995). *Relating polling models with nonzero and zero switchover times.* Queueing Systems **19**, 149-168.

[31] H. Takagi (1990). *Queueing analysis of polling models: an update.* In: Stochastic Analysis of Computer and Communication Systems. H. Takagi (ed.), Elsevier, Amsterdam, 267-318.

[32] H. Takagi (1997). *Queuing analysis of polling models: Progress in 1990-1994.* In: Frontiers in Queueing. J.H. Dshalalow (ed.), CRC Press, Boca Raton, 119-146.

[33] H. Takagi, T. Takine, and O.J. Boxma (1992). *Distribution of the workload in multiclass queueing systems with server vacations.* Naval Research Logistics **39**, 41-52.

[34] W. Willinger, M.S. Taqqu, W.E. Leland, and D.V. Wilson (1995). *Self-similarity in high-speed packet traffic: analysis and modeling of Ethernet traffic measurements.* Statistical Science **10**, 67-85.

[35] A.P. Zwart, and O.J. Boxma (1998). *Sojourn time asymptotics in the M/G/1 processor sharing queue.* Technical Report PNA-R9802, CWI, Amsterdam.

[36] A.P. Zwart (1999). *Sojourn times in a multiclass processor sharing queue.* In: *Teletraffic Engineering in a Competitive World*, Proc. ITC-16, eds. P. Key and D. Smith (North-Holland, Amsterdam), 335-344.