

Queue lengths and workloads in polling systems

Citation for published version (APA):

Boxma, O. J., Kella, O., & Kosinski, K. M. (2011). *Queue lengths and workloads in polling systems*. (Report Eurandom; Vol. 2011027). Eurandom.

Document status and date:

Published: 01/01/2011

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

EURANDOM PREPRINT SERIES
2011-027

Queue lengths and workloads in polling systems

O.J. Boxma, O. Kella, K.M. Kosiński
ISSN 1389-2355

QUEUE LENGTHS AND WORKLOADS IN POLLING SYSTEMS

O.J. BOXMA, O. KELLA, AND K.M. KOSIŃSKI

ABSTRACT. We consider a polling system: a queueing system of $N \geq 1$ queues with Poisson arrivals Q_1, \dots, Q_N visited in a cyclic order (with or without switchover times) by a single server. For this system we derive the probability generating function $\mathcal{Q}(\cdot)$ of the joint queue length distribution at an arbitrary epoch in a stationary cycle, under no assumptions on service disciplines. We also derive the Laplace-Stieltjes transform $\mathcal{W}(\cdot)$ of the joint workload distribution at an arbitrary epoch. We express \mathcal{Q} and \mathcal{W} in the probability generating functions of the joint queue length distribution at visit beginnings, $\mathcal{V}_{b_i}(\cdot)$, and visit completions, $\mathcal{V}_{c_i}(\cdot)$, at Q_i , $i = 1, \dots, N$. It is well known that \mathcal{V}_{b_i} and \mathcal{V}_{c_i} can be computed in a broad variety of cases. Furthermore, we establish a workload decomposition result.

1. INTRODUCTION

Consider a queueing model consisting of multiple queues: Q_1, \dots, Q_N , $N \geq 1$, attended by a single server, visiting the queues one at a time in a cyclic order. Moving from one queue to another, the server incurs a (possibly zero) switchover time. Such single-server multiple-queue models are commonly referred to as *polling systems*. Triggered by a wide variety of applications, polling systems have been extensively studied in the literature, see [11, 14, 16] for a series of comprehensive surveys and [1, 8, 12] for extensive overviews of the applicability of polling systems.

In the polling literature much attention has been given to the determination of the probability generating function (PGF) of the joint queue length distribution of the polling system at various epochs. In particular, the PGF's of the joint queue length distribution at the epochs at which the server begins a visit, $\mathcal{V}_{b_i}(\cdot)$, and completes a visit, $\mathcal{V}_{c_i}(\cdot)$, at Q_i were extensively studied. A seminal result has been obtained by Resing [9], who found these PGF's for polling systems with service disciplines satisfying the so-called *branching property*. Service disciplines – rules which establish the server's behaviour while visiting a queue – that satisfy the branching property include the well known *exhaustive* discipline (per visit the server continues to serve all customers at a station until it empties) and *gated* discipline (per visit the server serves only those customers at a station which are found there upon its visit).

Polling systems with disciplines which do not satisfy the branching property usually defy exact analysis. A typical example of a discipline that does not satisfy the branching property is the *k-limited* discipline (per visit at most k customers are served at a station). Nevertheless, it is possible to compute \mathcal{V}_{b_i} and \mathcal{V}_{c_i} for some special cases like completely symmetrical queues or like the two-queue case. For example, for the case $N = 2$ with Q_1 served exhaustively and Q_2 according to the k -limited discipline, the formulas for \mathcal{V}_{b_i} and \mathcal{V}_{c_i} were found by Winands et al. [17]. Furthermore, for the case of $N = 2$ and Q_1, Q_2 both served according to the 1-limited discipline, the formulas for \mathcal{V}_{b_i} and \mathcal{V}_{c_i} were found by solving a boundary value problem [4]. For other examples of

Date: June 6, 2011.

2010 *Mathematics Subject Classification*. Primary: 60K25; Secondary: 90B22.

Key words and phrases. polling system; queue length; steady-state distribution.

The second author is supported by grant No. 434/09 from the Israel Science Foundation and the Vigevani Chair in Statistics.

The third author is supported by NWO grant 613.000.701.

polling systems with non-branching disciplines that admit a solution for these PGF's we refer to the surveys [13, 14].

Although the PGF's \mathcal{V}_{b_i} and \mathcal{V}_{c_i} have been found in a very broad variety of cases, the PGF $\mathcal{Q}(\cdot)$ of the steady-state joint queue length distribution at an *arbitrary* epoch, which in principle is the most important PGF, received little attention and has not been studied in full generality, one exception (that we are aware of) being Sidi et al. [10]. The first contribution of the present paper is to determine \mathcal{Q} for a quite general class of polling systems. To explain the formula for \mathcal{Q} let us introduce two additional PGF's. Let $\mathcal{S}_{b_i}(\cdot)$ and $\mathcal{S}_{c_i}(\cdot)$ be the PGF's of the joint queue length distribution at service beginnings and service completions, respectively. Eisenberg [7] established a link between the various PGF's: $\mathcal{V}_{b_i}, \mathcal{V}_{c_i}, \mathcal{S}_{b_i}, \mathcal{S}_{c_i}$. In fact, from his observation it follows that it is enough to determine, for example, $\mathcal{V}_{b_i}, i = 1, \dots, N$, to know all the remaining PGF's $\mathcal{V}_{c_i}, \mathcal{S}_{b_i}, \mathcal{S}_{c_i}, i = 1, \dots, N$. Hence, as noted before, all these PGF's are well known for a vast class of polling systems. One of our main results – **Theorem 1** – reveals that the formula for \mathcal{Q} can be expressed in terms of $\mathcal{S}_{c_i}, i = 1, \dots, N$ (or equivalently in any other aforementioned PGF's). Using a similar technique that allowed us to determine \mathcal{Q} , we were also able to find $\mathcal{W}(\cdot)$, the Laplace-Stieltjes transform (LST) of the joint workload distribution in steady state. We give a formula for \mathcal{W} in **Theorem 2** expressed in \mathcal{V}_{b_i} and \mathcal{V}_{c_i} , which is the second contribution of our paper.

\mathcal{Q} was found in Sidi et al. [10] for polling systems restricted to the case of cycles with either all disciplines being gated or all being exhaustive. \mathcal{W} was found in [5], for the class of polling systems with Lévy input, but restricted to the case of branching-type service disciplines. Both of the results from [5, 10] assume non-zero switchover times. It should be emphasized therefore that our main results – **Theorem 1** and **Theorem 2** – hold regardless of the service discipline and provide a unified expression for \mathcal{Q} and \mathcal{W} for both the zero and non-zero switchover case. These differences are stressed once more in **Remark 1** and **Remark 4**.

From **Theorem 2** we were also able to retrieve the well known decomposition of the total workload W in a polling system into an independent sum of the amount of work $W_{M/G/1}$ in a corresponding $M/G/1$ queue and W_{switch} , the total amount of work in the polling system at an arbitrary epoch in a switching period. This decomposition was found in [3], however the distribution of W_{switch} was only known in a few cases, cf. [15]. In **Remark 5** we present the LST of the distribution of W_{switch} expressed in \mathcal{V}_{b_i} and \mathcal{V}_{c_i} . This constitutes the third and final contribution of our paper.

The paper is organised as follows. In **Section 2** we give a detailed description of the polling model under consideration, introduce the various PGF's $\mathcal{V}_{b_i}, \mathcal{V}_{c_i}, \mathcal{S}_{b_i}, \mathcal{S}_{c_i}$ and relate them to each other. **Section 3** contains our first main result, **Theorem 1** – the expression for \mathcal{Q} , the PGF of the joint queue length distribution in steady state at an arbitrary epoch, in terms of \mathcal{S}_{c_i} . **Section 4** contains our second main result, **Theorem 2** – the expression for \mathcal{W} , the LST of the joint workload distribution in steady state at an arbitrary epoch, in terms of \mathcal{V}_{b_i} and \mathcal{V}_{c_i} . This section also contains **Remark 5**, which establishes the workload decomposition in polling systems.

2. MODEL DESCRIPTION

2.1. A general polling system. We consider a system of $N \geq 1$ infinite-buffer queues Q_1, \dots, Q_N and a single server S . The service times of customers in Q_i are i.i.d. (independent, identically distributed) positive random variables generically denoted by B_i . We denote the LST of B_i by $\tilde{B}_i(\cdot)$ and assume that $b_i := \mathbb{E}B_i < \infty$. The server moves among the queues in a cyclic order. When S moves from Q_i to Q_{i+1} , it incurs a switchover period. The durations of successive switchover times are i.i.d. non-negative random variables, which we generically denote by S_i . We denote the LST of S_i by $\tilde{S}_i(\cdot)$ and assume that $s_i := \mathbb{E}S_i < \infty$; let $s := \sum_{i=1}^N s_i$. Customers arrive at Q_i according to a Poisson process with rate λ_i ; let $\lambda := \sum_{i=1}^N \lambda_i$. We do not assume anything about the service disciplines at Q_i . Define $\rho_i := \lambda_i b_i$ as the traffic intensity at Q_i ; let $\rho := \sum_{i=1}^N \rho_i$. In what

follows we shall write \mathbf{z} for an N -dimensional vector in \mathbb{R}^N , $\mathbf{z} = (z_1, \dots, z_N)$, and we assume that $|z_i| < 1$ for every $i = 1, \dots, N$. Throughout the paper we implicitly use the convention that any index summation is modulo N , for example $Q_{N+1} \equiv Q_1$.

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.

We also consider the variant of this model in which *all* the switchover times are *zero*. In the latter model, S makes a full cycle (viz., passes all the queues once) when the system becomes empty 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 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 visit completion instants. The convention that the server stops at some fixed queue was used in [2] and turned 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.

2.2. Distributional identities. For details of this subsection we refer to [2].

Note that each time a visit beginning or a service completion occurs, this coincides with either a service beginning or a visit completion. All service beginning epochs in a visit to Q_i are also service completion epochs at Q_i , except the first service beginning epoch – which is also a visit beginning epoch. Similarly, all service completion epochs at Q_i are also service beginning epochs at that queue, except the last service completion epoch – which is also a visit completion epoch. This has been observed by many authors, but most notably by Eisenberg [7], who restricted himself to the cases of either exhaustive or gated service at all queues. However, its applicability is not restricted to a particular service discipline. After some manipulations this observation yields:

$$(1) \quad \gamma_i \mathcal{V}_{b_i}(\mathbf{z}) + \mathcal{S}_{c_i}(\mathbf{z}) = \mathcal{S}_{b_i}(\mathbf{z}) + \gamma_i \mathcal{V}_{c_i}(\mathbf{z}).$$

Here $\mathcal{V}_{b_i}(\mathbf{z})$ and $\mathcal{V}_{c_i}(\mathbf{z})$ denote the PGF's of the joint queue length distribution at visit beginnings and visit completions at Q_i , while $\mathcal{S}_{b_i}(\mathbf{z})$ and $\mathcal{S}_{c_i}(\mathbf{z})$ denote the PGF's 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, i.e., the ratio of visit beginnings to service beginnings. With C denoting the steady-state cycle length, $1/\gamma_i = \lambda_i \mathbb{E}C$. In the case of non-zero switchover times, we simply have $\mathbb{E}C = s/(1 - \rho)$. In the zero switchover times case

$$(2) \quad \frac{\mathcal{V}_{b_1}(0)}{\mathbb{E}C} = \lambda(1 - \rho).$$

Indeed, $\mathcal{V}_{b_1}(0)/\mathbb{E}C$ denotes the mean number of times per time unit that S arrives at Q_1 to find the whole system empty ($\mathcal{V}_{b_1}(0)$ is the probability that S sees an empty system at a Q_1 visit). Each of those epochs is followed by exactly one arrival of a customer to an empty system. By PASTA, there are on average $\lambda(1 - \rho)$ such arrivals per time unit.

We rewrite (1) into:

$$(3) \quad \gamma_i (\mathcal{V}_{b_i}(\mathbf{z}) - \mathcal{V}_{c_i}(\mathbf{z})) = \mathcal{S}_{b_i}(\mathbf{z}) - \mathcal{S}_{c_i}(\mathbf{z}).$$

Let $\Sigma(\mathbf{z}) := \sum_{j=1}^N \lambda_j (1 - z_j)$. It is easy to see that $\mathcal{S}_{c_i}(\mathbf{z})$ and $\mathcal{S}_{b_i}(\mathbf{z})$ are related via

$$(4) \quad \mathcal{S}_{c_i}(\mathbf{z}) = \mathcal{S}_{b_i}(\mathbf{z}) \frac{\tilde{B}_i(\Sigma(\mathbf{z}))}{z_i}.$$

It follows from (1) and (4) that

$$(5) \quad \mathcal{S}_{b_i}(\mathbf{z}) = \frac{\gamma_i z_i}{z_i - \tilde{B}_i(\Sigma(\mathbf{z}))} (\mathcal{V}_{b_i}(\mathbf{z}) - \mathcal{V}_{c_i}(\mathbf{z})).$$

Next we relate $\mathcal{V}_{b_{i+1}}$ to \mathcal{V}_{c_i} . In a polling system with switchover times

$$(6) \quad \mathcal{V}_{b_{i+1}}(\mathbf{z}) = \mathcal{V}_{c_i}(\mathbf{z}) \tilde{S}_i(\Sigma(\mathbf{z})), \quad i = 1, \dots, N.$$

In a polling system without switchover times

$$(7) \quad \mathcal{V}_{b_{i+1}}(\mathbf{z}) = \mathcal{V}_{c_i}(\mathbf{z}), \quad i = 1, \dots, N-1.$$

The relation between \mathcal{V}_{b_1} and \mathcal{V}_{c_N} deserves special attention in the zero switchover case, because of our convention stated in Section 2.1, 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. Therefore,

$$(8) \quad \mathcal{V}_{b_1}(\mathbf{z}) = \mathcal{V}_{c_N}(\mathbf{z}) - \mathcal{V}_{b_1}(0) \left(1 - \sum_{i=1}^N \frac{\lambda_i}{\lambda} z_i \right) = \mathcal{V}_{c_N}(\mathbf{z}) - \frac{\mathcal{V}_{b_1}(0)}{\lambda} \Sigma(\mathbf{z}).$$

3. QUEUE LENGTH DISTRIBUTION AT AN ARBITRARY TIME

Theorem 1. For a general polling system as introduced in Section 2, let $\mathcal{Q}(\cdot)$ be the probability generating function of the joint queue length distribution at an arbitrary time in steady-state. Then,

$$(9) \quad \mathcal{Q}(\mathbf{z}) = \frac{\sum_{i=1}^N \lambda_i (1 - z_i) \mathcal{S}_{c_i}(\mathbf{z})}{\sum_{i=1}^N \lambda_i (1 - z_i)}.$$

Proof We will divide the proof into three parts. First we will find a formula for \mathcal{Q} in the non-zero switchover times case. Next we will show that this formula is also valid in the zero switchover times case. Finally, we will show that the unified formula can be expressed as (9).

Let $X_i(\cdot)$ and $Y_i(\cdot)$ be the PGF of the joint queue length distribution at an arbitrary moment during a visit to Q_i and during a switchover (idle) period between Q_i and Q_{i+1} , respectively.

Non-zero switchover times: By the stochastic mean value theorem

$$(10) \quad \mathcal{Q}(\mathbf{z}) = \frac{1}{\mathbb{E}C} \sum_{i=1}^N \left(\frac{b_i}{\gamma_i} X_i(\mathbf{z}) + s_i Y_i(\mathbf{z}) \right).$$

Furthermore, for any i ,

$$(11) \quad X_i(\mathbf{z}) = \mathcal{S}_{b_i}(\mathbf{z}) \tilde{B}_i^{\text{past}}(\Sigma(\mathbf{z})) = \mathcal{S}_{b_i}(\mathbf{z}) \frac{1 - \tilde{B}_i(\Sigma(\mathbf{z}))}{\Sigma(\mathbf{z}) b_i} = \frac{\gamma_i z_i (\mathcal{V}_{b_i}(\mathbf{z}) - \mathcal{V}_{c_i}(\mathbf{z}))}{b_i} \frac{1 - \tilde{B}_i(\Sigma(\mathbf{z}))}{z_i - \tilde{B}_i(\Sigma(\mathbf{z})) \Sigma(\mathbf{z})}$$

where $\tilde{B}_i^{\text{past}}(\cdot)$ is the LST of B_i^{past} , the past part of B_i and where we used (5). Analogously,

$$(12) \quad Y_i(\mathbf{z}) = \mathcal{V}_{c_i}(\mathbf{z}) \tilde{S}_i^{\text{past}}(\Sigma(\mathbf{z})) = \mathcal{V}_{c_i}(\mathbf{z}) \frac{1 - \tilde{S}_i(\Sigma(\mathbf{z}))}{\Sigma(\mathbf{z}) s_i} = \frac{1}{s_i} \frac{\mathcal{V}_{c_i}(\mathbf{z}) - \mathcal{V}_{b_{i+1}}(\mathbf{z})}{\Sigma(\mathbf{z})},$$

where $\tilde{S}_i^{\text{past}}(\cdot)$ is the LST of S_i^{past} , the past part of S_i and where we used (6). Combining (10)-(12) leads to

$$(13) \quad \mathcal{Q}(\mathbf{z}) = \frac{1}{\mathbb{E}C} \sum_{i=1}^N \left(\frac{\mathcal{V}_{b_i}(\mathbf{z}) - \mathcal{V}_{c_i}(\mathbf{z})}{\Sigma(\mathbf{z})} \frac{z_i (1 - \tilde{B}_i(\Sigma(\mathbf{z})))}{z_i - \tilde{B}_i(\Sigma(\mathbf{z}))} + \frac{\mathcal{V}_{c_i}(\mathbf{z}) - \mathcal{V}_{b_{i+1}}(\mathbf{z})}{\Sigma(\mathbf{z})} \right).$$

Zero switchover times: Observe that in this case we can write

$$\mathcal{Q}(z) = \rho \mathcal{Q}_{\text{serv}}(z) + (1 - \rho) \mathcal{Q}_{\text{non-serv}}(z),$$

where $\mathcal{Q}_{\text{serv}}(\cdot)$ and $\mathcal{Q}_{\text{non-serv}}(\cdot)$ denote the PGF of the joint queue length distribution at an arbitrary moment when S is serving and non-serving, respectively. Trivially, $\mathcal{Q}_{\text{non-serv}}(z) \equiv 1$. The stochastic mean value theorem gives

$$\mathcal{Q}_{\text{serv}}(z) = \frac{1}{\sum_{i=1}^N b_i / \gamma_i} \sum_{i=1}^N \frac{b_i}{\gamma_i} X_i(z).$$

Note that

$$\frac{\rho}{\sum_{i=1}^N b_i / \gamma_i} = \frac{\rho}{\mathbb{E}C \sum_{i=1}^N b_i \lambda_i} = \frac{\rho}{\mathbb{E}C \sum_{i=1}^N \rho_i} = \frac{1}{\mathbb{E}C}.$$

Hence, in the zero switchover case,

$$\mathcal{Q}(z) = \frac{1}{\mathbb{E}C} \sum_{i=1}^N \left(\frac{\mathcal{V}_{b_i}(z) - \mathcal{V}_{c_i}(z)}{\Sigma(z)} \frac{z_i (1 - \tilde{B}_i(\Sigma(z)))}{z_i - \tilde{B}_i(\Sigma(z))} \right) + 1 - \rho.$$

Noting that (2), (7) and (8) give

$$\frac{1}{\mathbb{E}C} \sum_{i=1}^N \frac{\mathcal{V}_{c_i}(z) - \mathcal{V}_{b_{i+1}}(z)}{\Sigma(z)} = \frac{\mathcal{V}_{c_N}(z) - \mathcal{V}_{b_1}(z)}{\mathbb{E}C \Sigma(z)} = \frac{\mathcal{V}_{b_1}(0)}{\lambda \mathbb{E}C} = 1 - \rho,$$

we conclude that (13) is also valid in the zero switchover times case.

Simplification of (13): Observe that

$$\sum_{i=1}^N \frac{\mathcal{V}_{c_i}(z) - \mathcal{V}_{b_{i+1}}(z)}{\Sigma(z)} = \sum_{i=1}^N \frac{\mathcal{V}_{c_i}(z) - \mathcal{V}_{b_i}(z)}{\Sigma(z)}.$$

Hence, using (4), (5) and the definition of γ_i we have

$$\begin{aligned} & \sum_{i=1}^N \left(\frac{\mathcal{V}_{b_i}(z) - \mathcal{V}_{c_i}(z)}{\Sigma(z)} \frac{z_i (1 - \tilde{B}_i(\Sigma(z)))}{z_i - \tilde{B}_i(\Sigma(z))} + \frac{\mathcal{V}_{c_i}(z) - \mathcal{V}_{b_{i+1}}(z)}{\Sigma(z)} \right) \\ &= \sum_{i=1}^N \left(\frac{\mathcal{V}_{b_i}(z) - \mathcal{V}_{c_i}(z)}{\Sigma(z)} \left[\frac{z_i (1 - \tilde{B}_i(\Sigma(z)))}{z_i - \tilde{B}_i(\Sigma(z))} - 1 \right] \right) \\ &= \sum_{i=1}^N \left(\frac{\mathcal{V}_{b_i}(z) - \mathcal{V}_{c_i}(z)}{\Sigma(z)} \frac{\tilde{B}_i(\Sigma(z))(1 - z_i)}{z_i - \tilde{B}_i(\Sigma(z))} \right) \\ &\stackrel{(5)}{=} \sum_{i=1}^N \frac{1 - z_i}{\Sigma(z)} \frac{\mathcal{S}_{c_i}(z)}{\gamma_i} \stackrel{(4)}{=} \mathbb{E}C \sum_{i=1}^N \frac{\lambda_i (1 - z_i)}{\Sigma(z)} \mathcal{S}_{c_i}(z), \end{aligned}$$

which completes the proof. \square

We would like to re-emphasise that \mathcal{Q} is completely determined when the \mathcal{S}_{c_i} are known. Thanks to the relations in Section 2.2, this is equivalent to knowing, for example, \mathcal{V}_{b_i} . The latter functions were found for various polling models as mentioned in Section 1.

Remark 1. In Sidi et al. [10] the authors also focus on the joint queue length distribution at an arbitrary epoch. However, the result in [10] is restricted to the case of non-zero switchover times and cycles with either all policies being gated or all being exhaustive. The authors averaged over N visit times and N switchover times, but did not obtain our Theorem 1 nor Formula (9).

Our theorem reveals that exactly the same structure holds, regardless of the service discipline. However, they do allow a more general customer behavior; their paper is one of the few polling studies in which the system is viewed as a network, with customers moving from queue to queue and the server visiting the queues cyclically.

Remark 2. Observe that **Theorem 1** immediately gives the formula for the marginal distributions. Indeed, for a vector $\mathbf{z}_{M,i} = (1, \dots, 1, z_i, 1, \dots, 1)$, $\mathcal{Q}(\mathbf{z}_{M,i}) = \mathcal{S}_{c_i}(\mathbf{z}_{M,i})$. From the ‘step’ (level crossing) argument it follows that $\mathcal{S}_{c_i}(\mathbf{z}_{M,i})$ is also the PGF of the queue length distribution in Q_i at an *arrival* epoch at Q_i . By PASTA it is also the steady-state distribution of Q_i .

Next take $\mathbf{z}_T = (z, \dots, z)$. **Theorem 1** now states that the PGF of the distribution of the total queue length (in terms of z) equals $\sum_{i=1}^N \lambda_i \mathcal{S}_{c_i}(\mathbf{z}_T) / \sum_{j=1}^N \lambda_j$. This formula may be interpreted as follows. By PASTA, $\mathcal{Q}(\mathbf{z}_T)$ is also the PGF of the distribution of the total queue length at an arrival epoch. By a level crossing argument, it follows that this equals the PGF of the distribution of the total queue length just after a departure epoch. The result now follows from the observation that a fraction $\lambda_i / \sum_{j=1}^N \lambda_j$ of the departure epochs refers to a departure from Q_i .

Remark 3. It would be very interesting to have an interpretation of (9). Below we present a short derivation, which essentially combines several of the formulas above, but which fails to give an immediate explanation:

$$\begin{aligned}
\mathcal{Q}(\mathbf{z}) &\stackrel{(10)}{=} \sum_{i=1}^N \rho_i X_i(\mathbf{z}) + (1 - \rho) \sum_{i=1}^N \frac{s_i}{s} Y_i(\mathbf{z}) \\
&\stackrel{(11),(12)}{=} \sum_{i=1}^N \rho_i \frac{\mathcal{S}_{b_i}(\mathbf{z}) - z_i \mathcal{S}_{c_i}(\mathbf{z})}{b_i \sum_{j=1}^N \lambda_j (1 - z_j)} + \frac{1}{\mathbb{E}C} \sum_{i=1}^N \frac{\mathcal{V}_{c_i}(\mathbf{z}) - \mathcal{V}_{b_{i+1}}(\mathbf{z})}{\sum_{j=1}^N \lambda_j (1 - z_j)} \\
&= \frac{\sum_{i=1}^N \lambda_i (\mathcal{S}_{b_i}(\mathbf{z}) - z_i \mathcal{S}_{c_i}(\mathbf{z}))}{\sum_{j=1}^N \lambda_j (1 - z_j)} + \frac{1}{\mathbb{E}C} \frac{\sum_{i=1}^N (\mathcal{V}_{c_i}(\mathbf{z}) - \mathcal{V}_{b_i}(\mathbf{z}))}{\sum_{j=1}^N \lambda_j (1 - z_j)} \\
&\stackrel{(3)}{=} \frac{\sum_{i=1}^N \lambda_i (\mathcal{S}_{b_i}(\mathbf{z}) - z_i \mathcal{S}_{c_i}(\mathbf{z}))}{\sum_{j=1}^N \lambda_j (1 - z_j)} + \frac{\sum_{i=1}^N \lambda_i (\mathcal{S}_{c_i}(\mathbf{z}) - \mathcal{S}_{b_i}(\mathbf{z}))}{\sum_{j=1}^N \lambda_j (1 - z_j)} \\
&= \frac{\sum_{i=1}^N \lambda_i (1 - z_i) \mathcal{S}_{c_i}(\mathbf{z})}{\sum_{i=1}^N \lambda_i (1 - z_i)}.
\end{aligned}$$

4. WORKLOAD DISTRIBUTION AT AN ARBITRARY TIME

The idea of the proof of **Theorem 1** can be also used to determine the LST \mathcal{W} of the joint workload distribution at an arbitrary epoch. For future use, let $\boldsymbol{\omega} := (\omega_1, \dots, \omega_N)$ and $\tilde{\mathbf{B}}(\boldsymbol{\omega}) := (\tilde{B}_1(\omega_1), \dots, \tilde{B}_N(\omega_N))$. Moreover, recall that $\Sigma(\mathbf{z}) = \sum_{j=1}^N \lambda_j (1 - z_j)$.

Theorem 2. For a general polling system as introduced in **Section 2**, let $\mathcal{W}(\cdot)$ be the Laplace-Stieltjes transform of the joint workload distribution at an arbitrary time in steady-state. Then,

$$\mathcal{W}(\boldsymbol{\omega}) = \frac{1}{\mathbb{E}C} \sum_{i=1}^N \frac{\mathcal{V}_{b_i}(\tilde{\mathbf{B}}(\boldsymbol{\omega})) - \mathcal{V}_{c_i}(\tilde{\mathbf{B}}(\boldsymbol{\omega}))}{\Sigma(\tilde{\mathbf{B}}(\boldsymbol{\omega}))} \frac{\omega_i}{\Sigma(\tilde{\mathbf{B}}(\boldsymbol{\omega})) - \omega_i},$$

where $\mathbb{E}C = s/(1 - \rho)$ in the non-zero switchover case and $\mathbb{E}C = \mathcal{V}_{b_1}(0)/(\lambda(1 - \rho))$ in the zero switchover case.

Proof The proof follows the same reasoning as the proof of **Theorem 1**. In particular, we focus only on the non-zero switchover times case.

Let $\tilde{X}_i(\cdot)$ and $\tilde{Y}_i(\cdot)$ be the LST of the joint workload distribution at an arbitrary moment during a visit to Q_i and during a switchover (idle) period between Q_i and Q_{i+1} , respectively. Then by the

stochastic mean value theorem

$$\mathscr{W}(\omega) = \frac{1}{\mathbb{E}C} \sum_{i=1}^N \left(\frac{b_i}{\gamma_i} \tilde{X}_i(\omega) + s_i \tilde{Y}_i(\omega) \right).$$

Firstly, note that using (12),

$$\tilde{Y}_i(\omega) = Y_i(\tilde{\mathbf{B}}(\omega)) = \frac{1}{s_i} \frac{\mathcal{V}_{c_i}(\tilde{\mathbf{B}}(\omega)) - \mathcal{V}_{b_{i+1}}(\tilde{\mathbf{B}}(\omega))}{\Sigma(\tilde{\mathbf{B}}(\omega))}.$$

Secondly,

$$\hat{X}_i(\omega) = \frac{\mathcal{S}_{b_i}(\tilde{\mathbf{B}}(\omega))}{\tilde{B}_i(\omega_i)} \times \int_{u=0}^{\infty} \int_{t=0}^{\infty} \exp\left(-\Sigma(\tilde{\mathbf{B}}(\omega))u - \omega_i t\right) d\mathbb{P}(B_i^{past} < u, B_i^{res} < t).$$

Indeed, consider an arbitrary service time during a visit to Q_i . The term $\mathcal{S}_{b_i}(\tilde{\mathbf{B}}(\omega))/\tilde{B}_i(\omega_i)$ corresponds to the LST of the workload of all the customers in the system at the beginning of that service time, *excluding* the customer whose service is about to begin. Secondly, consider the workload that arrives at all queues during the past part B_i^{past} of that service time, *and* the residual part B_i^{res} of that same service time. Next integrate with respect to the joint distribution of B_i^{past} and B_i^{res} . Using (5) and the fact that, cf. [6, Section I.6.3],

$$\int_{u=0}^{\infty} \int_{t=0}^{\infty} \exp\left(-\Sigma(\tilde{\mathbf{B}}(\omega))u - \omega_i t\right) d\mathbb{P}(B_i^{past} < u, B_i^{res} < t) = \frac{\tilde{B}_i(\omega_i) - \tilde{B}_i(\Sigma(\tilde{\mathbf{B}}(\omega)))}{b_i \left(\Sigma(\tilde{\mathbf{B}}(\omega)) - \omega_i\right)},$$

we obtain

$$\begin{aligned} \hat{X}_i(\omega) &= \frac{\mathcal{S}_{b_i}(\tilde{\mathbf{B}}(\omega))}{\tilde{B}_i(\omega_i)} \frac{\tilde{B}_i(\omega_i) - \tilde{B}_i(\Sigma(\tilde{\mathbf{B}}(\omega)))}{b_i \left(\Sigma(\tilde{\mathbf{B}}(\omega)) - \omega_i\right)} = \frac{\gamma_i \mathcal{V}_{b_i}(\tilde{\mathbf{B}}(\omega)) - \mathcal{V}_{c_i}(\tilde{\mathbf{B}}(\omega))}{b_i \tilde{B}_i(\omega_i) - \tilde{B}_i(\Sigma(\tilde{\mathbf{B}}(\omega)))} \frac{\tilde{B}_i(\omega_i) - \tilde{B}_i(\Sigma(\tilde{\mathbf{B}}(\omega)))}{\Sigma(\tilde{\mathbf{B}}(\omega)) - \omega_i} \\ &= \frac{\gamma_i \mathcal{V}_{b_i}(\tilde{\mathbf{B}}(\omega)) - \mathcal{V}_{c_i}(\tilde{\mathbf{B}}(\omega))}{b_i \Sigma(\tilde{\mathbf{B}}(\omega)) - \omega_i}. \end{aligned}$$

This completes the proof. \square

Remark 4. \mathscr{W} was found in [5] for the more general class of polling systems with Lévy input, but restricted to the case of branching-type service disciplines at all queues and non-zero switchover times. Notice that $\Sigma(\tilde{\mathbf{B}}(\omega)) - \omega_i$ is the Laplace exponent $\phi_i^A(\omega)$ of the Lévy process

$$\mathbf{A}_i(t) = (W_1(t), \dots, W_{i-1}(t), W_i(t) - t, W_{i+1}(t), \dots, W_N(t)),$$

where W_i is a compound Poisson process with jump distribution B_i and rate λ_i . Moreover $\Sigma(\tilde{\mathbf{B}}(\omega))$ is the Laplace exponent of the Lévy process $\mathbf{W}(t) = (W_1(t), \dots, W_N(t))$. After identifying $\mathcal{V}_{b_i}(\tilde{\mathbf{B}}(\omega))$ and $\mathcal{V}_{c_i}(\tilde{\mathbf{B}}(\omega))$ of **Theorem 2** with $\tilde{B}_i(\omega)$ and $\tilde{E}_i(\omega)$ of **Theorem 3** of [5], it can now be verified that **Theorem 2** indeed coincides with [5, Theorem 3].

Remark 5. Observe that **Theorem 2** gives the LST of the distribution of W , the total workload in a polling system. With $\omega_T = (\omega, \dots, \omega)$, we have $\mathbb{E}[e^{-\omega W}] = \mathscr{W}(\omega_T)$, so that

$$\mathbb{E}[e^{-\omega W}] = \frac{1 - \rho}{s} \frac{\omega}{\omega - \sum_{j=1}^N \lambda_j (1 - \tilde{B}_j(\omega))} \sum_{i=1}^N \frac{\mathcal{V}_{c_i}(\tilde{\mathbf{B}}(\omega_T)) - \mathcal{V}_{b_i}(\tilde{\mathbf{B}}(\omega_T))}{\sum_{j=1}^N \lambda_j (1 - \tilde{B}_j(\omega))}.$$

Note that the LST of the amount of work $W_{M/G/1}$ in a corresponding $M/G/1$ queue, that is, an $M/G/1$ queue with arrival rate λ and service time LST $\sum_{j=1}^N \lambda_j \tilde{B}_j(\omega)/\lambda$, is given by

$$\mathbb{E}[e^{-\omega W_{M/G/1}}] = \frac{(1 - \rho)\omega}{\omega - \sum_{j=1}^N \lambda_j (1 - \tilde{B}_j(\omega))}.$$

Note that $W_{M/G/1}$ is also the amount of work in the polling system with zero switchover times. Finally, from the proof of [Theorem 2](#) we know that the LST of the distribution of the total amount of work W_{switch} in a polling system at an arbitrary epoch in a switching period is given by

$$(14) \quad \mathbb{E}[e^{-\omega W_{switch}}] = \frac{1}{s} \sum_{i=1}^N s_i \tilde{Y}_i(\omega_T) = \frac{1}{s} \sum_{i=1}^N \frac{\mathcal{V}_{c_i}(\tilde{\mathbf{B}}(\omega_T)) - \mathcal{V}_{b_i}(\tilde{\mathbf{B}}(\omega_T))}{\sum_{j=1}^N \lambda_j (1 - \tilde{B}_j(\omega))}.$$

Hence, we retrieve the well known decomposition (see, e.g., [\[3\]](#))

$$\mathbb{E}[e^{-\omega W}] = \mathbb{E}[e^{-\omega W_{M/G/1}}] \mathbb{E}[e^{-\omega W_{switch}}].$$

The distribution of W_{switch} was only known in a few special cases, cf. [\[15\]](#). From [Theorem 2](#) it follows that in general its LST is given by [\(14\)](#).

5. ACKNOWLEDGEMENTS

We would like to thank Marko Boon for reading the first draft of this paper, which led to the nice simplification of Formula [\(13\)](#) into [\(9\)](#).

REFERENCES

- [1] M.A.A. Boon, R.D. van der Mei, and E.M.M. Winands. Applications of polling systems. *SORMS*, 16: 67–82, 2011.
- [2] S.C. Borst and O.J. Boxma. Polling models with and without switchover times. *Oper. Res.*, 45:536–543, 1997.
- [3] O.J. Boxma and W.P. Groenendijk. Pseudo-conservation laws in cyclic-service systems. *J. Appl. Probab.*, 24:949–964, 1987.
- [4] O.J. Boxma and W.P. Groenendijk. Two queues with alternating service and switching times. In O.J. Boxma and R. Syski, editors, *Queueing Theory and its Applications*, Liber Amicorum for J.W. Cohen, pages 261–282. North-Holland, Amsterdam, 1988.
- [5] O.J. Boxma, J. Ivanovs, K.M. Kosiński, and Mandjes M. Lévy driven polling systems and continuous-state branching processes. Technical Report 2009-030, EURANDOM, 2009.
- [6] J.W. Cohen. *The Single Server Queue*. North-Holland Publishing Company, Amsterdam, 1982.
- [7] M. Eisenberg. Queues with periodic service and changeover time. *Oper. Res.*, 20:440–451, 1972.
- [8] H. Levy and M. Sidi. Polling models: applications, modeling and optimization. *IEEE Trans. Commun.*, 38:1750–1760, 1990.
- [9] J.A.C. Resing. Polling systems and multitype branching processes. *Queueing Syst.*, 13:409–426, 1993.
- [10] M. Sidi, H. Levy, and S.W. Fuhrmann. A queueing network with a single cyclically roving server. *Queueing Syst.*, 11:121–144, 1992.
- [11] H. Takagi. *Queueing Analysis*, volume 1. North-Holland, Amsterdam, 1991.
- [12] H. Takagi. Application of polling models to computer networks. *Comput. Netw. ISDN Syst.*, 22:193–211, 1991.
- [13] H. Takagi. Queueing analysis of polling models: progress in 1990-1994. In J.H. Dshalalow, editor, *Frontiers in Queueing: Models, Methods and Problems*, pages 119–146. CRC Press, Boca Raton, 1997.
- [14] H. Takagi. Analysis and application of polling models. In G. Haring, C. Lindemann, and M. Reiser, editors, *Performance Evaluation: Origins and Directions*, volume 1769 of *Lecture Notes in Computer Science*, pages 424–442. Springer, Berlin, 2000.
- [15] H. Takagi, T. Takine, and O.J. Boxma. Distribution of the workload in multiclass queueing systems with server vacations. *Naval Research Logistics*, 39:41–52, 1992.
- [16] V.M. Vishnevskii and O.V. Semenov. Mathematical methods to study the polling systems. *Autom. Remote Control*, 67:173–220, 2006.
- [17] E.M.M. Winands, I.J.B.F. Adan, G.J. Houtum, and D.G. Down. A state-dependent polling model with k -limited service. *Probab. Engrg. Inform. Sci.*, 23:385–408, 2009.

EURANDOM AND DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE, EINDHOVEN UNIVERSITY OF TECHNOLOGY, THE NETHERLANDS

E-mail address: `Boxma@win.tue.nl`

DEPARTMENT OF STATISTICS, THE HEBREW UNIVERSITY OF JERUSALEM, JERUSALEM 91905, ISRAEL

E-mail address: `Offer.Kella@huji.ac.il`

EURANDOM, EINDHOVEN UNIVERSITY OF TECHNOLOGY; KORTEWEG-DE VRIES INSTITUTE FOR MATHEMATICS, UNIVERSITY OF AMSTERDAM, THE NETHERLANDS

E-mail address: `Kosinski@eurandom.tue.nl`