

# VU Research Portal

## **A new method for deriving waiting-time approximations for polling systems with renewal arrivals**

Dorsman, J.L.; van der Mei, R.D.; Winands, E.M.M.

### ***published in***

Stochastic Models  
2012

### ***DOI (link to publisher)***

[10.1080/15326349.2011.567933](https://doi.org/10.1080/15326349.2011.567933)

### ***document version***

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

### ***citation for published version (APA)***

Dorsman, J. L., van der Mei, R. D., & Winands, E. M. M. (2012). A new method for deriving waiting-time approximations for polling systems with renewal arrivals. *Stochastic Models*, 27(2), 318-332.  
<https://doi.org/10.1080/15326349.2011.567933>

### **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 ?

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

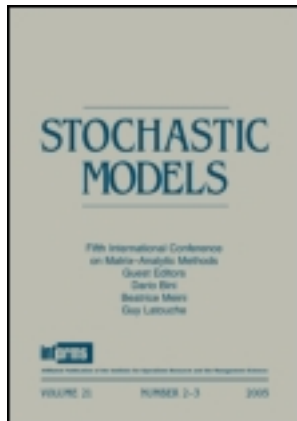
[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

This article was downloaded by: [Vrije Universiteit Amsterdam]

On: 29 October 2013, At: 02:28

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Stochastic Models

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/Istm20>

### A New Method for Deriving Waiting-Time Approximations in Polling Systems with Renewal Arrivals

J. L. Dorsman<sup>a b</sup>, R. D. van der Mei<sup>a c</sup> & E. M. M. Winands<sup>c</sup>

<sup>a</sup> CWI, Probability and Stochastic Networks, Amsterdam, The Netherlands

<sup>b</sup> Eindhoven University of Technology, EURANDOM, Department of Mathematics and Computer Science, Eindhoven, The Netherlands

<sup>c</sup> VU University Amsterdam, Department of Mathematics, Amsterdam, The Netherlands

Published online: 03 May 2011.

To cite this article: J. L. Dorsman, R. D. van der Mei & E. M. M. Winands (2011) A New Method for Deriving Waiting-Time Approximations in Polling Systems with Renewal Arrivals, *Stochastic Models*, 27:2, 318-332, DOI: [10.1080/15326349.2011.567933](http://dx.doi.org/10.1080/15326349.2011.567933)

To link to this article: <http://dx.doi.org/10.1080/15326349.2011.567933>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

## A NEW METHOD FOR DERIVING WAITING-TIME APPROXIMATIONS IN POLLING SYSTEMS WITH RENEWAL ARRIVALS

J. L. Dorsman<sup>1,2</sup>, R. D. van der Mei<sup>1,3</sup>, and E. M. M. Winands<sup>3</sup>

<sup>1</sup>*CWI, Probability and Stochastic Networks, Amsterdam, The Netherlands*

<sup>2</sup>*Eindhoven University of Technology, EURANDOM, Department of Mathematics and Computer Science, Eindhoven, The Netherlands*

<sup>3</sup>*VU University Amsterdam, Department of Mathematics, Amsterdam, The Netherlands*

□ *We study the waiting-time distributions in cyclic polling models with renewal arrivals, general service and switch-over times, and exhaustive service at each of the queues. The assumption of renewal arrivals prohibits an exact analysis and reduces the available analytic results to heavy-traffic asymptotics, limiting results for large switch-over times and large numbers of queues, and some numerical algorithms. Motivated by this, the goal of this paper is to propose a new method for deriving simple closed-form approximations for the complete waiting-time distributions that work well for arbitrary load values. Extensive simulation results show that the approximations are highly accurate over a wide range of parameter settings.*

**Keywords** Approximation; Polling systems; Renewal arrivals; Waiting-time distribution.

**Mathematics Subject Classification** Primary 68M20, 60K25; Secondary 90B22.

### 1. INTRODUCTION

Polling systems are queueing systems consisting of multiple queues, attended by a single server that visits the queues to serve waiting customers. Whenever the server proceeds from one queue to another, a switch-over time is incurred. Typically, the queues are visited in a cyclic order. Polling models occur naturally in the modelling of applications where different types of customers compete for access to a common resource. Typical application areas of polling models are computer-communication systems,

Received June 2010; Accepted October 2010

Address correspondence to J. L. Dorsman, Eindhoven University of Technology, Eindhoven, Netherlands; E-mail: j.l.dorsman@tue.nl

manufacturing systems, maintenance systems and traffic systems. We refer to Refs.<sup>[3,10]</sup> for an overview of the applicability of polling models, and to Refs.<sup>[16,18]</sup> for overviews of the available results. In the present paper, we study cyclic polling with exhaustive service at all queues and with renewal arrival processes.

In the literature, the vast majority of papers on polling models rely on the assumption of Poisson arrivals, whereas for models with non-Poisson arrivals hardly any exact results on the waiting-time distributions are known. The most general results known for renewal-driven polling models are asymptotic results, such as heavy-traffic asymptotics<sup>[14]</sup> or asymptotics for large switch-over times<sup>[20]</sup>. Closed-form approximations are available for the mean waiting time<sup>[2]</sup>, but today there is no simple approximation available for the tail probabilities of the waiting times at each of the queues. Even in case of Poisson arrivals, generally little attention is paid to the tail probabilities of the waiting-time distributions. As an exception, assuming Poisson arrivals, Choudhury and Whitt<sup>[4]</sup> propose an efficient numerical algorithm to calculate the moments and the tail probabilities of the waiting times based on numerical transform inversion, for branching-type polling models<sup>[15]</sup>. For models that violate the branching structure, more computationally intensive algorithms exist<sup>[1,9]</sup>. A common drawback of these numerical algorithms is that they only give limited insight into how the waiting-time distribution reacts to changes in the system parameters. These observations raise the need for the development of simple yet accurate approximations for the tail probabilities of the waiting times for polling models with renewal arrivals.

Motivated by this, in this paper we propose a new method for deriving closed-form approximations of the waiting-time distribution for arbitrary values of the load. The approach taken is that of combining known heavy-traffic (HT) asymptotics for the waiting-time distributions<sup>[12,14]</sup>, which work well when the system is heavily loaded, with a recently developed approximation for the mean waiting times in Ref.<sup>[2]</sup>, which works well for the whole range of load values (see also Ref.<sup>[19]</sup>). This paper presents the first *closed-form* approximation of the waiting-time *distribution* in polling systems with renewal arrivals. In this regard, the present research is an extension of Ref.<sup>[2]</sup>, which derives approximate expressions for the *mean* waiting times. This approximation is shown to be exact in the known limiting cases, and extensive experimentation with simulations shows that the approximation is highly accurate for a wide range of parameter settings. We emphasize that the strength of this combined approach lies in its striking simplicity and the fact that it leads to approximations in closed form, which opens up many possibilities for generalization of the approach to other polling models (e.g., with more general branching-type service policies, and with non-cyclic periodic server routing) and for optimization of the system performance with respect to the system parameters.

The remainder of this paper is organized as follows. In Section 2 we introduce the model and notation. In Section 3 we present the distributional approximation, which is the main result of this paper. In Section 4 we discuss properties of the approximation, and in Section 5, the accuracy of the approximation is evaluated by an extensive simulation study. Finally, in Section 6 we discuss suggestions for further research.

## 2. MODEL DESCRIPTION AND NOTATION

Consider a polling system consisting of  $N \geq 1$  queues,  $Q_1, \dots, Q_N$ , with an infinite-sized buffer. Customers arrive at  $Q_j$  according to a renewal process, with rate  $\lambda_i = \frac{1}{\mathbb{E}[A_i]}$ , where  $A_i$  is the random variable describing the interarrival times of customers at  $Q_j$ . The total arrival rate to the system is denoted by  $\Lambda = \sum_{i=1}^N \lambda_i$ . Within a queue, customers are served on a first-in-first-out (FIFO) basis. The service time of a type- $i$  customer at  $Q_j$  is denoted by the random variable  $B_i$  with  $k$ th moment  $\mathbb{E}[B_i^k]$ , and its waiting time in  $Q_j$  by the random variable  $W_i$  with  $k$ th moment  $\mathbb{E}[W_i^k]$ ,  $k > 0$ . The random variable  $B$  denotes the service time of an arbitrary customer entering the system, with  $\mathbb{E}[B^k] = \sum_{i=1}^N \frac{\lambda_i}{\Lambda} \mathbb{E}[B_i^k]$ . Queues are served according to an *exhaustive* service discipline, and as soon as  $Q_j$  becomes empty, the server proceeds to  $Q_{j+1}$ . We define a cycle at  $Q_j$  as the time between two successive departures of the server from  $Q_j$ . The time needed by the server to switch from  $Q_j$  to  $Q_{j+1}$  is denoted by the random variable  $S_j$  with  $k$ th moment  $\mathbb{E}[S_j^k]$  ( $k > 0$ ). Let the random variable  $S = \sum_{i=1}^N S_i$  denote the total switch-over time in a cycle. Throughout, it is assumed that  $\mathbb{E}[S] > 0$  and that all inter-arrival times, service times and switch-over times are mutually independent and independent of the state of the system. The load offered to  $Q_j$  is denoted by  $\rho_j = \lambda_j \mathbb{E}[B_j]$ ,  $1 \leq j \leq N$ . The total load in the system is denoted by  $\rho = \sum_{i=1}^N \rho_i > 0$ . A necessary and sufficient condition for the stability of the described system is  $\rho < 1$ <sup>[6]</sup>. The waiting time at  $Q_j$  is defined as the time between the arrival of an arbitrary customer in the system and the moment when he is taken into service.

Throughout, it may be convenient to scale a system such that a certain load is achieved. This scaling is done by keeping the service time distributions fixed and varying the rates of the renewal processes. In particular, it proves convenient to denote with  $\hat{x}$  the value of each variable  $x$  that is a function of  $\rho$  evaluated at  $\rho = 1$ .  $\hat{A}_i$  then denotes the inter-arrival time of  $Q_j$  customers evaluated at  $\rho = 1$ . Then, scaling to a load  $\rho < 1$  is done by taking the random variable  $A_i := \hat{A}_i/\rho$ .

Finally, we introduce some notation. The residual length of a random variable  $X$  is denoted by  $X^{res}$ , with  $\mathbb{E}[X^{res}] = \frac{\mathbb{E}[X^2]}{2\mathbb{E}[X]}$ . The squared coefficient of variation (SCV) of a random variable  $X$ ,  $\frac{\text{Var}[X]}{\mathbb{E}[X]^2}$  is denoted by  $c_X^2$ . We define  $\sigma^2 = \sum_{i=1}^N \hat{\lambda}_i (\text{Var}[B_i] + c_{A_i}^2 \mathbb{E}[B_i]^2)$  and  $\delta = \sum_{j=1}^N \sum_{k=j+1}^N \hat{\rho}_j \hat{\rho}_k$ ; note

that in the case of Poisson arrivals, the former can be simplified to  $\sigma^2 = \mathbb{E}[B^2]/\mathbb{E}[B]$ . The notation  $\xrightarrow{d}$  means convergence in distribution, and  $\mathbf{1}_{\{A\}}$  denotes the indicator function on the event  $A$ . Finally, when a random variable  $X$  is said to have a gamma distribution with shape parameter  $\alpha$  and inverse scale parameter  $\mu$ , its density function is given by  $f_X(x) = e^{-\mu x} \mu^\alpha x^{\alpha-1} \mathbf{1}_{\{x \geq 0\}} / \Gamma(\alpha)$ , where  $\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1} dx$ .

### 3. DERIVATION OF THE APPROXIMATION

The two key ingredients of the distributional approximation will be the HT diffusion approximation for the waiting time by Olsen and Van der Mei<sup>[14]</sup>, and the mean waiting-time approximation by Boon et al.<sup>[2]</sup> for a general value of the load  $\rho < 1$ . The HT diffusion approximation will be refined such that its mean coincides with the mean waiting-time approximation, while the diffusion approximation remains unchanged in the case of HT after refinement. The two ingredients are given first, after which the main result is derived and presented. Although not stated explicitly, it follows naturally from Ref.<sup>[14]</sup> that

$$(1 - \rho)W_i \xrightarrow{d} UI_i, \quad \rho \uparrow 1, \tag{1}$$

where  $U$  is a uniformly distributed random variable on  $[0,1]$ , and  $I_i$  a gamma distributed random variable with parameters

$$\alpha = \frac{2\mathbb{E}[S]\delta}{\sigma^2} + 1 \quad \text{and} \quad \mu_i = \frac{2\delta}{(1 - \hat{\rho}_i)\sigma^2}. \tag{2}$$

Let  $\mathbb{E}[W_i]$  denote the mean waiting time of a type- $i$  customer at  $Q_j$ ,  $1 \leq i \leq N$ . Then, Boon et al.<sup>[2]</sup> propose the following approximation  $\mathbb{E}[W_{i,Boon}]$  for the mean waiting time:

$$\mathbb{E}[W_{i,Boon}] = \frac{K_0 + K_{1,i}\rho + K_{2,i}\rho^2}{1 - \rho}, \tag{3}$$

where the constants  $K_0, K_{1,i}$  and  $K_{2,i}$  depend on several parameters of the polling system at hand:

$$\begin{aligned} K_0 &= \mathbb{E}[S^{res}], \\ K_{1,i} &= \hat{\rho}_i((c_{A_i}^2)^4 \mathbf{1}_{\{c_{A_i}^2 \leq 1\}} + 2 \frac{c_{A_i}^2}{c_{A_i}^2 + 1} \mathbf{1}_{\{c_{A_i}^2 > 1\}} - 1) \mathbb{E}[B_i^{res}] + \mathbb{E}[B^{res}] \\ &\quad + \hat{\rho}_i(\mathbb{E}[S^{res}] - \mathbb{E}[S]) - \frac{1}{\mathbb{E}[S]} \sum_{j=0}^{N-1} \sum_{k=0}^j \hat{\rho}_{i+k} \text{Var}[S_{i+j}], \\ K_{2,i} &= \frac{1 - \hat{\rho}_i}{2} \left( \frac{\sigma^2}{2\delta} + \mathbb{E}[S] \right) - K_0 - K_{1,i}. \end{aligned}$$

In fact,  $\mathbb{E}[W_{i,Boon}]$  is an interpolation between a light-traffic (LT) limit and a HT limit of the mean waiting-time;  $K_0$  and  $K_{1,i}$  stem from the LT limit, while  $K_{2,i}$  is based directly on the HT limit.

Our distributional approximation will be a refinement of the HT diffusion approximation given in (1). To this end, in the spirit of (1) we assume that the waiting-time distribution of  $Q_i$  can be well approximated by a product of a uniform random variable on  $[0,1]$  and a gamma random variable with  $\alpha_{ia}$  and  $\mu_{ia}$ , divided by  $(1 - \rho)$ . To parameterize  $\alpha_{ia}$  and  $\mu_{ia}$ , we formulate the following three requirements for the refinement of the approximation in (3):

1. In HT the refined approximation must coincide with the diffusion approximation of Ref.<sup>[14]</sup>, i.e.,

$$\frac{\alpha}{\alpha_{ia}} \rightarrow 1 \quad \text{and} \quad \frac{\mu_i}{\mu_{ia}} \rightarrow 1, \quad \text{when } \rho \uparrow 1.$$

2. The expectation of the refined approximation coincides with the mean waiting-time approximation of Ref.<sup>[2]</sup>.
3. The SCV of the refined approximating distribution matches the SCV of the HT diffusion approximation by Ref.<sup>[14]</sup>, so that the shape of the refined diffusion approximation matches the shape of the HT diffusion approximation.

These three requirements uniquely determine the parameters  $\alpha_{ia}$  and  $\mu_{ia}$ , leading to the following approximation for the waiting-time distribution in polling systems with renewal arrivals and  $\rho < 1$ :

$$\mathbb{P}[W_i < x] \approx \mathbb{P}[U I_{i,app} < (1 - \rho)x], \quad (4)$$

where  $U$  is a uniformly distributed random variable on  $[0,1]$ , and  $I_{i,app}$  a gamma distribution with parameters

$$\alpha_{ia} = \alpha_a = \frac{2\mathbb{E}[S]\delta}{\sigma^2} + 1 \quad \text{and} \quad \mu_{ia} = \frac{2\mathbb{E}[S]\delta + \sigma^2}{2\sigma^2(1 - \rho)\mathbb{E}[W_{i,Boon}]}. \quad (5)$$

It can be verified that the  $k$ th moment of the obtained distributional approximation can be expressed as follows, for  $k \geq 1$ ,

$$\mathbb{E}[W_{i,app}^k] = \frac{1}{(1 - \rho)^k} \frac{1}{k + 1} \prod_{i=0}^{k-1} \frac{\alpha_a + i}{\mu_{ia}} = \frac{2^k \mathbb{E}[W_{i,Boon}]^k}{k + 1} \prod_{i=1}^k \frac{2\mathbb{E}[S]\delta + i\sigma^2}{2\mathbb{E}[S]\delta + \sigma^2}, \quad (6)$$

with  $\alpha_a$  and  $\mu_{ia}$  as defined above.

Note that we have tried a number of adjustments, but the one presented in the third requirement turned out to be the most robust and

intuitively appealing (see also Remark 3). Other natural possibilities for adjustment are, for example, matching the variance of the HT diffusion approximation or an interpolation between the SCV in LT and HT.

We end this section with a number of remarks.

**Remark 1** (Olsen's Approximation). A refined diffusion approximation for the distribution of the waiting time in polling systems with Poisson arrivals was presented by Olsen in Ref.<sup>[13]</sup>. The HT diffusion approximation used in Ref.<sup>[13]</sup> consists of a uniformly distributed random variable on  $[0,1]$  "times" a gamma distribution with shape parameter  $\frac{\mathbb{E}[S] \sum_{i=1}^N \rho_i (\rho - \rho_i)}{\sum_{i=1}^N \lambda_i (\text{Var}[B_i] + \mathbb{E}[B_i]^2)} + 1$  and inverse scale parameter  $\frac{(1-\rho) \sum_{i=1}^N \rho_i (\rho - \rho_i)}{(1-\rho_i) \sum_{i=1}^N \lambda_i (\text{Var}[B_i] + \mathbb{E}[B_i]^2)}$ . Note that in HT these parameters coincide with the ones used in (1). Refinement is done using an approximation of the mean delay obtained by Ref.<sup>[7]</sup> for Poisson arrivals. Suggested by this mean delay approximation, Olsen adds an extra factor of  $\rho$  in the shape parameter, such that it becomes  $\frac{\mathbb{E}[S] \sum_{i=1}^N \rho_i (\rho - \rho_i)}{\rho \sum_{i=1}^N \lambda_i (\text{Var}[B_i] + \mathbb{E}[B_i]^2)} + 1 = \frac{2\mathbb{E}[S]\delta}{\sigma^2} + 1$ . The inverse scale parameter is changed accordingly such that the approximation satisfies the mean delay approximation in Ref.<sup>[7]</sup>. One can verify that in case of Poisson arrivals, the shape parameters of Olsen's approximation and our approximation coincide. Hence, the distributional approximation as given in this paper generalises Olsen's approximation to systems with renewal arrivals, and the presented derivation of the main result of this paper creates intuition and justification behind the distributional approximation of Ref.<sup>[13]</sup>.

**Remark 2** (Information Availability). The derived waiting-time distribution approximation (4) only requires the first two moments of the interarrival, service and switch-over time distributions as an input, whereas the complete waiting-time distribution generally depends on their complete distributions, even for Poisson arrivals. This makes the approximations useful for practical purposes, because in reality information about more than the first two moments is often hard to get.

**Remark 3** (Applicability). Yet another view is provided by the notion that the derived approximation gives a procedure to estimate the complete waiting-time distribution based on the mean waiting time and aggregate measures for imbalance  $\delta$  and variability  $\sigma^2$ . In this regard, it is important to note that the mean waiting time can easily be measured in real-life applications, in contrast to higher moments or tail probabilities.

**Remark 4** (Other Service Disciplines). Given that the two key ingredients are available for polling systems with other service disciplines, distributional waiting-time approximations can be derived for these



classes of polling systems as well using the method as described above. To illustrate this, consider a polling model where all queues are served according to a *gated* service discipline, i.e., where during a visiting period at a queue the server will only serve the customers which arrived before the start of this period. For this service discipline, it follows from Ref.<sup>[14]</sup> that

$$(1 - \rho)W_i \xrightarrow{d} UI_i, \quad \rho \uparrow 1, \quad (7)$$

where  $U$  is uniform on  $[\hat{\rho}_i, 1]$ , and where  $I_i$  is a gamma distribution with parameters  $\alpha$  and  $(1 - \hat{\rho}_i)\mu_i$ , with  $\alpha$  and  $\mu_i$  as given in (2). As for the second key ingredient, Ref.<sup>[2]</sup> also contains an approximation of the delay's first moment for the gated service discipline. This approximation still has the form of (3), however with different values for  $K_0, K_{1,i}$  and  $K_{2,i}$ , derived in Ref.<sup>[2]</sup>. Using these two key ingredients, one obtains for the distributional approximation

$$\mathbb{P}[W_i < x] \approx \mathbb{P}[U_{i,app} < (1 - \rho)x], \quad (8)$$

where  $U$  is uniform on  $[\hat{\rho}_i, 1]$  and where  $I_{i,app}$  has a gamma distribution with parameters  $\alpha_a$  and  $(1 + \hat{\rho}_i)\mu_{ia}$ , with  $\alpha_a$  and  $\mu_{ia}$  as given in (5). Note that  $\mathbb{E}[W_{i,Boon}]$  in the expression of  $\mu_{ia}$  in this case refers to the gated version of Boon's approximation, derived in Ref.<sup>[2]</sup>.

#### 4. ALIGNMENT WITH ASYMPTOTIC REGIMES

In Ref.<sup>[2]</sup> it is shown that the first moment of the distributional approximation is in line with several known exact results, which gives support for the quality of the approximation. For Poisson arrivals the approximation satisfies the well-known pseudo-conservation laws and is exact in symmetric systems, vacation queues and general systems in LT. Moreover, the approximation gives exact results for systems with general renewal arrivals in the asymptotic regimes of HT or infinite switch-over times, as shown below. In the present section, comparable results are shown for higher moments of the distributional approximation.

##### 4.1. Heavy-Traffic

By construction, the distributional approximation is exact in systems with general renewal arrivals in HT. This property is very desirable from a practical perspective, since the proper operation of a system is particularly critical when the system is heavily loaded.

## 4.2. Large Switch-Over Times

In case of deterministic switch-over times, the waiting time is only dependent on the total switch-over time in a cycle  $S$ , rather than the marginal switch-over times  $S_i$  (cf. Ref.<sup>[81]</sup>). A strong conjecture is presented in Ref.<sup>[20]</sup> that in this case the distribution of  $\frac{W_i}{S}$  tends to a uniform distribution on  $[0, \frac{1-\rho_i}{1-\rho}]$  as  $S \rightarrow \infty$ ; for the case of Poisson arrivals, a rigorous proof of this result was given in Ref.<sup>[11]</sup>. It turns out that the distributional approximation as presented satisfies this result. To this end, consider the  $k$ th moment of  $\frac{W_{i,app}}{S}$ ,  $k > 0$  as  $S \rightarrow \infty$ . It can easily be verified that  $\lim_{S \rightarrow \infty} \mathbb{E}[\frac{W_{i,Boon}}{S}] = \frac{1-\rho_i}{2(1-\rho)}$ , and hence,

$$\begin{aligned} \lim_{S \rightarrow \infty} \mathbb{E} \left[ \left( \frac{W_{i,app}}{S} \right)^k \right] &= \frac{1}{k+1} \left( \frac{1-\rho_i}{1-\rho} \right)^k \lim_{S \rightarrow \infty} \prod_{i=1}^k \frac{2S\delta + i\sigma^2}{2S\delta + \sigma^2} \\ &= \frac{1}{k+1} \left( \frac{1-\rho_i}{1-\rho} \right)^k. \end{aligned} \quad (9)$$

This expression exactly coincides with the finite  $k$ th moment of a uniformly distributed random variable  $Y$  on  $[0, \frac{1-\rho_i}{1-\rho}]$ . Thus, the  $k$ th moment of  $\frac{W_{i,app}}{S}$  converges to the  $k$ th moment of  $Y$  when  $S$  tends to infinity,  $k \geq 1$ . Under certain conditions (which are met here), this moment-wise convergence implies convergence in distribution (cf. Ref.<sup>[5]</sup>, Theorem 4.5.5). Therefore, the distributional approximation becomes exact in the case of deterministic switch-over times that tend to infinity.

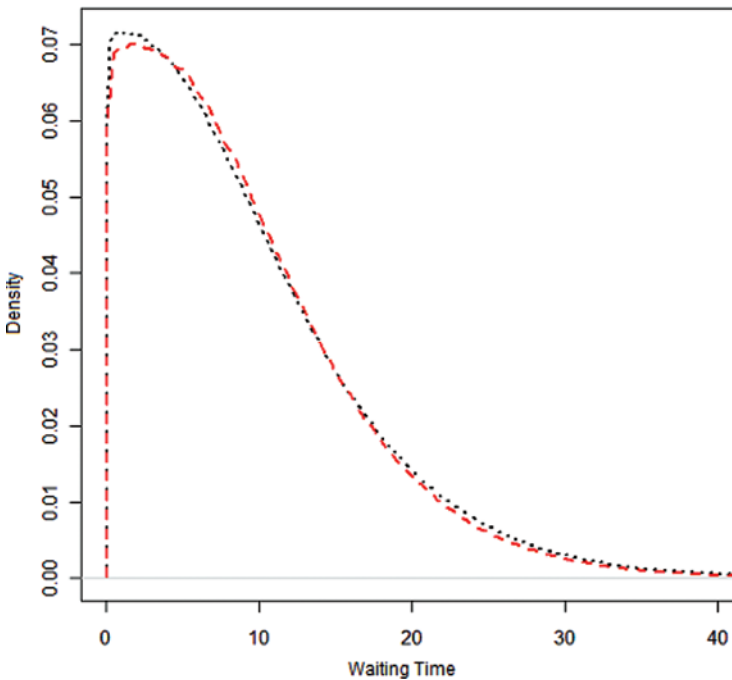
## 5. SIMULATION STUDY

In this section, we evaluate the accuracy of the approximation of the waiting-time distribution as presented in Section 3. First, we regard a rather arbitrary polling system and see how well the approximated and exact density functions coincide. The “exact” density function is determined by means of simulation. Then, we study the accuracy of the approximation in a wide range of parameter combinations by applying the approximation to a test bed containing 10368 polling systems and summarising the errors of standard deviation and percentile approximations. Again, the “exact” standard deviations and percentiles are determined by means of simulation. Also in case of Poisson arrivals, where numerical methods exist to determine the exact distribution, we opt for simulation, since the determination of the exact values using numerical methods can be very cumbersome. All simulation results presented in this section are an average taken from a variable number of simulation runs with a length of at least 1,000,000 time units, such that the width of the confidence interval of the average is less than 1% of the value of the actual average.

### 5.1. Accuracy of the Approximated Density Function

We consider a symmetric polling system with five queues. The load  $\rho$  equals 0.7, the SCV of the interarrival times at each queue is 0.25. All the service times and switch-over times are exponentially distributed with mean 1. Since there is no exact closed-form expression available for the waiting-time distribution in this case, we compare the density function of the approximated distribution with the simulated density function for an arbitrary queue. To obtain the latter, a kernel estimation was made based on a huge set of simulated waiting-time realisations. Both the approximated density function and the simulated density function are depicted in Figure 1. For the interarrival times, a gamma distribution was used with shape parameter 4 and inverse scale parameter 16.

Figure 1 shows that the shape of the exact waiting-time distribution (dashed line) closely resembles the approximation (dotted line), which suggests that the approximation is useful for approximating the waiting-time distribution. The next subsection will show that the approximation works well not only in this case, but also in a variety of other polling systems.



**FIGURE 1** Approximated and simulated density function of the waiting-time of an arbitrary queue in the example in Subsection 5.1.

### 5.2. Accuracy of Approximated Percentiles and Standard Deviation

In this subsection we assess the accuracy of the approximation by evaluating errors in the approximation of the standard deviation and several percentiles. We regard the standard deviation and several percentiles of the approximated distribution and the exact distribution of the waiting time of the first queue in a large number of polling systems with exhaustive, cyclic service. The standard deviation and percentiles of the exact distribution are determined by means of simulation. We first give a general impression of the accuracy, after which we try to explore the impact of each of the parameters on the accuracy.

The parameter values contained in the test bed can be found in Table 1. There is no difference between the queues within a particular polling system in terms of service times and switch-over times. All parameters are explained above, except for the last one displayed in the table.  $Q_2, \dots, Q_N$  take on the same amount of load each.  $p_1$  denotes what amount of load is taken on by  $Q_1$  relative to the other queues. For example, if the first queue takes half, twice or five times as much load as any other,  $p_1$  becomes 0.5, 2 or 5, respectively. In case of a symmetric system,  $p_1 = 1$ . In the test bed, several SCV's are regarded for several random variables. When the SCV equals one, an exponential distribution is fitted to the mean. For other SCV values, distributions commonly used in a two-moment fit were fitted to the first two moments; a mixture of two Erlang distributions in case of a SCV smaller than one and a  $H_2$  distribution (with balanced means) for a SCV larger than one (cf. Ref.<sup>[17]</sup>).

For each polling system, the approximation error of the standard deviation and the approximation error of the 40th, 50th, 60th, 70th, 80th, 90th, and 95th-percentiles are calculated. The errors are measured in a

TABLE 1 Parameter values of the test bed used in subsection 5.2

Notation	Parameter	Considered parameter values
$N$	Number of queues	{5, 10, 20}
$\rho$	Load	{0.5, 0.6, 0.7, 0.8, 0.9, 0.95}
$c_{A_i}^2$	SCV interarrival times	{0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2}
$E[B_i]$	Mean service times	{1}
$c_{B_i}^2$	SCV service times	{0, 1, 4}
$E[S_i]$	Mean switch-over times	{0.2, 1, 10}
$c_{S_i}^2$	SCV switch-over times	{0, 1}
$p_1$	Measure of asymmetry	{0.5, 1, 2, 5}

**TABLE 2** Mean standard deviation error of the approximation applied to the test bed, categorised in bins of 5%

$p_1$	Bins				
	0–5%	5–10%	10–15%	15–20%	20%+
0.5	69.98%	19.68%	5.02%	2.28%	3.05%
1	69.87%	19.33%	5.56%	2.39%	2.85%
2	67.67%	19.98%	6.17%	3.20%	2.97%
5	58.72%	22.07%	8.60%	4.86%	5.75%

percentual absolute relative way, i.e.,

$$\Delta\% = \frac{|a - s|}{s} \times 100\%, \quad (10)$$

where  $a$  denotes the approximated value by means of the distributional approximation as presented in Section 3 and  $s$  denotes the exact value, determined by means of simulation.

Tables 2 and 3 show the errors made in approximating the standard deviation and in approximating the percentiles respectively in bins of 5%. This table suggests that the approximation performs best in case of  $p_1 \leq 1$ . Regardless, it is shown that the majority of the standard deviation approximations, and even more so the percentile approximation have errors lower than 5%.

To assess the impact of the system parameters on the performance of the approximation, the mean absolute relative errors of the standard deviation approximation and the percentile approximation are given in Tables 4 and 5 respectively, vertically categorised in the different rates of asymmetry, horizontally categorised in each of the relevant system parameters. Tables 4(a) and 5(a) show that the distributional approximation becomes better when  $N$  increases. The same behaviour of the approximation error in  $N$  is present in the approximations of Refs.<sup>[2,13]</sup>.

**TABLE 3** Mean percentile error of the approximation applied to the test bed, categorised in bins of 5%

$p_1$	Bins				
	0–5%	5–10%	10–15%	15–20%	20%+
0.5	82.25%	10.77%	3.26%	1.07%	2.65%
1	82.14%	10.98%	3.44%	1.03%	2.41%
2	82.25%	11.17%	3.38%	1.04%	2.16%
5	77.63%	13.32%	4.76%	2.00%	2.29%

**TABLE 4** Mean standard deviation error categorised by the value of  $p_1$  vertically and the number of queues (a), the SCV of the service times (b), the SCV of the switch-over times (c), the total load (d), the mean switch-over time (e), and the SCV of the interarrival times (f) horizontally

(a)			(b)			(c)				
$N$			$c_{B_i}^2$			$c_{S_i}^2$				
$p_1$	5	10	20	$p_1$	0	1	4	$p_1$	0	1
0.5	5.97	4.02	3.17	0.5	3.91	3.25	5.99	0.5	4.49	4.28
1	5.99	4.06	3.25	1	4.15	3.33	5.81	1	4.59	4.27
2	6.41	4.31	3.29	2	4.80	3.64	5.57	2	8.80	4.46
5	9.08	5.76	3.77	5	6.97	5.49	6.15	5	6.73	5.68

(d)						(e)				
$\rho$						$\mathbb{E}[S_i]$				
$p_1$	0.5	0.6	0.7	0.8	0.9	0.95	$p_1$	0.2	1	10
0.5	8.22	6.66	5.11	3.48	1.81	1.06	0.5	7.35	3.37	2.44
1	8.37	6.77	5.12	3.49	1.79	1.06	1	7.34	3.53	2.33
2	8.82	7.14	5.42	3.65	1.87	1.12	2	7.44	4.07	2.30
5	11.02	9.30	7.36	5.19	2.75	1.60	5	8.76	6.79	3.06

(f)								
$c_{A_i}^2$								
$p_1$	0.25	0.5	0.75	1	1.25	1.50	1.75	2
0.5	4.11	4.13	3.95	4.04	4.34	4.59	4.82	5.10
1	4.10	4.14	4.00	4.06	4.40	4.67	4.91	5.18
2	4.18	4.42	4.28	4.33	4.67	4.93	5.18	5.37
5	4.77	5.20	5.28	6.00	6.48	6.94	7.30	7.65

Tables 4(b) and 5(b) show a surprising effect of the SCV of the service times on the performance of the approximation. While in case of  $c_{B_i}^2 = 0$  the standard deviation approximation error seems to grow with  $p_1$ , the same effect does not seem to happen when  $c_{B_i}^2 = 4$ . Also, if  $c_{B_i}^2 = 1$  or  $c_{B_i}^2 = 4$ , the approximation error of the standard deviation and the approximation error of the percentiles do not seem to react in the same way to changes in  $p_1$ . Tables 4(c) and 5(c) suggest that the distributional approximation becomes better when the variance of the switch-over times increases. Tables 4(d) and 5(d) suggest that approximations become better as  $\rho$  approaches 1, i.e., as the system becomes closer to HT. This is very plausible, since by construction the approximation is exact in HT, as discussed in Section 4. According to Tables 4(e) and 5(e) the approximations seem to become better as the switch-over times larger. This is in line with the observation made in Section 4 that the approximations become exact as the total switch-over time tends to infinity. Also, using

**TABLE 5** Mean percentile error categorised by the value of  $p_1$  vertically and the number of queues (a), the SCV of the service times (b), the SCV of the switch-over times (c), the total load (d), the mean switch-over time (e), and the SCV of the interarrival times (f) horizontally

(a)			(b)			(c)				
$N$			$c_{B_i}^2$			$c_{S_i}^2$				
$p_1$	5	10	20	$p_1$	0	1	4	$p_1$	0	1
0.5	5.77	3.27	1.99	0.5	2.15	2.40	6.48	0.5	3.87	3.49
1	5.45	3.23	2.05	1	2.28	2.27	6.19	1	3.77	3.39
2	5.10	3.18	2.02	2	2.44	2.31	5.56	2	3.58	3.29
5	5.87	3.50	2.15	5	3.35	3.03	5.14	5	4.14	3.54

(d)						(e)				
$\rho$						$\mathbb{E}[S_i]$				
$p_1$	0.5	0.6	0.7	0.8	0.9	0.95	$p_1$	0.2	1	10
0.5	5.99	5.79	4.46	2.89	1.37	0.79	0.5	7.50	2.19	1.34
1	5.86	5.65	4.28	2.75	1.30	0.79	1	6.98	2.31	1.45
2	5.67	5.44	4.05	2.56	1.22	0.75	2	6.43	2.42	1.46
5	6.14	5.86	4.60	3.10	1.55	0.91	5	6.14	3.64	1.74

(f)								
$c_{A_i}^2$								
$p_1$	0.25	0.5	0.75	1	1.25	1.50	1.75	2
0.5	4.20	3.59	3.25	3.31	3.48	3.65	3.86	4.09
1	3.97	3.41	3.14	3.29	3.44	3.61	3.78	3.97
2	3.53	3.13	3.04	3.30	3.45	3.56	3.67	3.80
5	3.15	3.26	3.52	3.80	4.00	4.18	4.32	4.50

(6) one can show that the moments of the distributional approximation become less dependent of  $\sigma^2$  and  $\delta$  as switch-over times become smaller, which gives support to the plausibility of the approximation becoming less reliable when the switch-over times become relatively small. Finally, both Tables 4(f) and 5(f) show that the approximations' quality is dependent on the SCV of the interarrival times, but again an interaction effect with the value of  $p_1$  is observed.

Table 6 shows the mean absolute relative error categorised per tested percentile. Generally, the 80% percentiles seem to be approximated best.

From the test-bed results we can conclude that the approximation performs well over a wide range of parameter combinations. In case of extremely variable service times, low load and negligibly small switch-over times, the relative error becomes worse. The worst-case scenarios found in the testbed in terms of absolute relative error are approximations of the 50th percentile in systems with  $N = 5$ ,  $\rho = 0.5$ ,  $c_{B_i}^2 = 4$  and  $\mathbb{E}[S_i] = 0.2$

**TABLE 6** Mean absolute relative errors categorised in the several percentiles

$p_1$	Percentile						
	40th	50th	60th	70th	80th	90th	95th
0.5	5.90	5.23	4.15	2.88	1.72	2.29	3.57
1	5.73	5.00	3.93	2.71	1.67	2.37	3.63
2	5.42	4.59	3.60	2.50	1.67	2.49	3.78
5	6.29	4.87	3.40	2.11	1.88	3.42	4.92

having errors with an order of magnitude of 100%. However, in practice these characteristics are uncommon. For example, in production systems settings like  $c_{B_i}^2 = 4$  are hardly found due to the just-in-time philosophy, which dictates to reduce variability in, e.g. service times in order to reduce in-process inventory. Also, these systems are typically utilized beyond  $\rho = 0.5$  to increase productivity, and switch-over periods are commonly longer than service periods. Moreover, in case of a low load and small switch-over times, although the relative error of the percentile approximations can be high, the absolute errors may still be rather small when compared to the order of longitude of service time durations. Therefore, the sojourn time distribution is already much better approximated in these situations.

## 6. FURTHER RESEARCH

The present paper gives birth to a variety of directions for further research. Firstly, the distributional approximation for cyclic systems with exhaustive service may be generalised to models with branching-type service policies<sup>[15]</sup>, non-cyclic periodic server routing<sup>[14]</sup> and other model variations. Secondly, the simple closed-form expression may act as a basis for design decisions within polling systems. Finally, one could attempt to improve the approximation by deriving an interpolation approximation for higher moments of the waiting-time and, subsequently, fit a phase-type distribution. Such an extension would be of particular interest if one has more information than the first two moments of the interarrival, service and switch-over time distributions (see also Remark 2). However, this would impel one to considerably extend the analysis of Ref.<sup>[2]</sup>, while potentially losing the simple form of the current distributional approximation.

## ACKNOWLEDGMENTS

The authors thank Marko Boon for placing parts of his polling system simulation program at their disposal, and for helpful comments on earlier drafts of this article.



## REFERENCES

1. Blanc, J.P.C. Performance evaluation of polling systems by means of the power-series algorithm. *Ann. Operations Res.* **1992**, *35*, 155–186.
2. Boon, M.A.A.; Winands, E.M.M.; Adan, I.J.B.F.; van Wijk, A.C.C. Closed-form waiting-time approximations for polling systems. *Performance Evaluation* **2011**, *68*, 290–306.
3. Boon, M.A.A.; van der Mei, R.D.; Winands, E.M.M. Applications of polling systems. *Surveys in Operations Research and Management Science* **2011**, *16*, 67–82.
4. Choudhury, G.; Whitt, W. Computing transient and steady state distributions in polling models by numerical transform inversion. *Perform. Eval.* **1994**, *25*, 267–292.
5. Chung, K.L. *A Course in Probability*, 2nd ed. Academic Press: New York, 1974.
6. Down, D. On the stability of polling models with multiple servers. *J. Appl. Prob.* **1998**, *35*, 925–935.
7. Everitt, D. Simple approximations for token rings. *IEEE Trans. Commun.* **1986**, *34*, 719–721.
8. Fuhrmann, S.W. A decomposition result for a class of polling models. *Queueing Sys.* **1992**, *11*, 109–120.
9. Leung, K.K. Cyclic-service systems with probabilistically-limited service. *IEEE J. Select. Areas Commun.* **1991**, *9*, 185–193.
10. Levy, H.; Sidi, M. Polling systems: applications, modeling, and optimization. *IEEE Trans. Commun.* **1990**, *38*, 1750–1760.
11. van der Mei, R.D. Delay in polling systems with large switch-over times. *J. Appl. Prob.* **1990**, *36*, 232–243.
12. van der Mei, R.D.; Winands, E.M.M. A note on polling models with renewal arrivals and nonzero switch-over times. *Oper. Res. Lett.* **2008**, *36*, 500–505.
13. Olsen, T.L. Approximations for the waiting-time distribution in polling models with and without state-dependent setups. *Oper. Res. Lett.* **2001**, *28*, 113–123.
14. Olsen, T.L.; van der Mei, R.D. Polling systems with periodic server routing in heavy-traffic: renewal arrivals. *Oper. Res. Lett.* **2005**, *33*, 17–25.
15. Resing, J.A.C. Polling systems and multitype branching processes. *Queueing Syst.* **1993**, *13*, 409–426.
16. Takagi, H. *Analysis of Polling Systems*; MIT Press: Cambridge, 1985.
17. Tijms, H.C. *Stochastic Models: An Algorithmic Approach*; Chichester: Wiley, 1994.
18. Vishnevskii, V.M.; Semenova, O.V. Mathematical methods to study the polling systems. *Automation Remote Control* **2006**, *67* (2), 173–220.
19. Whitt, W. Refining diffusion approximations for queues. *Oper. Res. Lett.* **1982**, *1*, 165–169.
20. Winands, E.M.M. Branching-type polling systems with large setups. *OR Spectrum* **2011**, *33*, 77–97.