

A note on polling models with renewal arrivals and nonzero switch-over times

Citation for published version (APA):

Mei, van der, R. D., & Winands, E. M. M. (2007). *A note on polling models with renewal arrivals and nonzero switch-over times*. (SPOR-Report : reports in statistics, probability and operations research; Vol. 200705). Technische Universiteit Eindhoven.

Document status and date:

Published: 01/01/2007

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.

A note on polling models with renewal arrivals and nonzero switch-over times

R.D. van der Mei* and E.M.M. Winands†

Abstract

We consider polling systems with general service times and switch-over times, gated service at all queues and with general renewal arrival processes. We derive closed-form expressions for the expected delay in heavy-traffic (HT). So far, rigorous proofs of HT limits have only been obtained for Poisson-type arrival processes, whereas for renewal arrivals HT limits are based on conjectures [5, 6, 8].

Keywords: polling systems, switch-over times, renewal arrivals, heavy traffic

1 Introduction

A typical polling system consists of a number of queues, attended by a single server in a fixed order. The ubiquity of polling systems can be observed in applications in computer-communication, production, transportation and maintenance systems [12]. In this note, we study polling models in which the arrival process at each of the queues follows a general renewal process. In particular, we focus on the heavy-traffic (HT) behavior of such models, i.e., when the load tends to one. In case of Poisson arrivals, rigorous proofs for HT limits can be obtained for models that possess a multi-type branching process (MTBP) structure [9] (see, e.g., [13]). In case of renewal arrivals and a general number of queues, HT limits have only been obtained on the basis of conjectures [5, 6, 8]. In this note, we study a method to derive rigorous proofs for HT asymptotics in gated polling models with a general number of queues under the assumption of general renewal arrivals. The approach in the present note has its origin in [14], where we study systems with *zero* switch-over times. At face value the extension to *nonzero* switch-over times may seem a small one, however this extension impels us to, considerably, modify and extend the analysis in [14] as done in this note.

We start our analysis by extending a result of Bertsimas and Mourtzinou [2] to general switch-over time distributions, which yields a set of linear equations for the variance of the cycle times for polling models with renewal arrivals in HT. Exploiting the similarities of this set with the corresponding set for systems with Poisson arrivals yields a closed-form expression for the asymptotic pseudo-conservation law (PCL) for systems with renewal arrivals in HT. Subsequently, by taking the proper HT limits of this set in combination with the derived PCL we obtain explicit closed-form expressions for the mean asymptotic scaled delay in HT. The latter result can be seen as the main contribution of the present note and opens up a range of challenges for further generalizations. Finally, we present - in the interest of space - detailed proofs only for the gated policy, but we want to stress that the approach is also readily applicable to the exhaustive policy.

*Centre for Mathematics and Computer Science (CWI), Department of Probability and Stochastic Networks, 1098 SJ Amsterdam, & Vrije Universiteit, Department of Mathematics, 1081 HV Amsterdam, The Netherlands, mei@cwi.nl

†Department of Mathematics and Computer Science & Department of Technology Management, Technische Universiteit Eindhoven, P.O. Box 513, 5600 MB Eindhoven, The Netherlands, e.m.m.winands@tue.nl

The remainder of this note is structured as follows. In Section 2 the model is introduced and an expression is given for the mean asymptotic scaled delay in HT, which is the main result of this note. In Section 3 a rigorous proof is presented and several asymptotic insensitivity properties are formulated. Lastly, Section 4 touches upon a very interesting challenge for further research.

2 Model description and notation

Consider a system consisting of $N \geq 2$ stations Q_1, \dots, Q_N , each with an infinite-sized buffer. A single server visits the queues in cyclic order, where he applies the *gated* service policy, i.e., when the server polls a queue, he serves all, and only, customers found at the polling instant. Type- i customers arrive at Q_i according to a renewal arrival process, defined by the distribution of the interarrival times A_i ; the arrival rate at Q_i is denoted by $\lambda_i := 1/E[A_i]$. The total arrival rate is denoted by $\Lambda = \sum_{i=1}^N \lambda_i$. The service time of a type- i customer is a random variable B_i , with finite k -th moment $b_i^{(k)}$, $k = 1, 2$. The k -th moment of the service time of an arbitrary customer is denoted by $b^{(k)} = \sum_{i=1}^N \lambda_i b_i^{(k)} / \Lambda$, $k = 1, 2$. The load offered to Q_i is $\rho_i = \lambda_i b_i^{(1)}$, and the total offered load is equal to $\rho = \sum_{i=1}^N \rho_i$.

After completing service at queue i , the server proceeds to queue $i + 1$, after incurring a switch-over period with distribution equal to that of a random variable R_{i+1} with finite k -th moment $r_{i+1}^{(k)}$, $k = 1, 2$. Denote by $r = \sum_{i=1}^N r_i^{(1)}$ the total expected switch-over period per cycle. Throughout it is assumed that $r > 0$. All interarrival times, service times and switch-over periods are assumed to be mutually independent and independent of the state of the system. A necessary and sufficient condition for the stability of the system is $\rho < 1$ [7]. Throughout, for each variable x that is a function of ρ , we denote its values *evaluated at* $\rho = 1$ by \hat{x} . Furthermore, we use the notation that $h(x) \sim g(x)$ as $x \uparrow a$ means that $\lim_{\rho \uparrow a} h(x)/g(x) = 1$. Finally, for compactness of presentation, all references to queue indices greater than N or less than 1 are implicitly assumed to be modulo N , e.g., queue $N + 1$ actually refers to queue 1.

Let W_i be the delay incurred by an arbitrary customer at Q_i , defined as the time between the arrival of a customer at a station and the moment at which he starts to receive service. Our main interest is in the behavior of the mean delay $E[W_i]$ in HT, i.e., as ρ tends to 1. It goes without saying that, in HT, all queues become unstable and, thus, $E[W_i]$ tends to infinity for all i . To be precise, $E[W_i]$ has a first-order pole at $\rho = 1$, for $i = 1, 2, \dots, N$,

$$\mathbb{E}[W_i] = \frac{\omega_i}{1 - \rho} + o((1 - \rho)^{-1}), \quad \rho \uparrow 1, \quad (1)$$

where $g(x) = o(f(x))$ means that $g(x)/f(x) \rightarrow 0$ as $x \uparrow 1$. More colloquially, we can say that ω_i , which is referred to as the *mean asymptotic scaled delay* at queue i , indicates the rate at which $E[W_i]$ tends to infinity as $\rho \uparrow 1$. For the validity of the statement that $E[W_i]$ has a first-order pole at $\rho = 1$, we refer to Remark 3.7.

The main result of the present note is the following.

Theorem 2.1. *For $i = 1, 2, \dots, N$,*

$$\omega_i = \frac{(1 + \hat{\rho}_i)}{2} \left(\frac{\sigma^2}{\sum_{j=1}^N \hat{\rho}_j (1 + \hat{\rho}_j)} + r \right), \quad (2)$$

with

$$\sigma^2 := \sum_{i=1}^N \hat{\lambda}_i \left(\text{Var}[B_i] + \hat{\rho}_i^2 \text{Var}[\hat{A}_i] \right). \quad (3)$$

Here, the limit is taken such that the arrival rates are increased, while keeping the service-time distributions fixed, and keeping the distributions of the interarrival times A_i ($i = 1, \dots, N$) fixed up to a common scaling constant ρ (i.e., $A_i =_d \rho IA_i$, where IA_i ($i = 1, \dots, N$) are the interarrival times at $\rho = 1$). Notice that in the case of Poisson arrivals we have $\sigma^2 = b^{(2)}/b^{(1)}$.

3 Analysis

In the present section we review and extend the asymptotic results of Bertsimas and Mourtzinou [2] for gated polling systems with arbitrary renewal arrival processes in HT. Starting point of our analysis is the following expression for the mean delay of each customer class as $\rho \uparrow 1$, for $i = 1, 2, \dots, N$ (cf. [2]),

$$E[W_i] \sim \frac{(1 + \rho_i)}{2} \left(\frac{\text{Var}[C_i]}{E[C_i]} + E[C_i] \right) + \frac{(c_{A_i}^2 - 1)b_i^{(1)}}{2}, \quad (4)$$

where $c_{A_i}^2$ is the squared coefficient of the interarrival time for queue i and where the i -cycle C_i is defined to be the time between two successive polling instants at Q_i . The mean cycle lengths $E[C_i]$ can be shown to be independent of the queue involved and are given by, for $i = 1, 2, \dots, N$ and $\rho < 1$ (see, e.g, [1]),

$$E[C_i] = \frac{r}{1 - \rho}, \quad (5)$$

whereas the variances of the cycle lengths $\text{Var}[C_i]$ ($i = 1, \dots, N$), can generally not be obtained in closed form and do depend on the queue involved.

3.1 Set of equations

Bertsimas and Mourtzinou [2] prove that the N unknowns $\text{Var}[C_i]$ ($i = 1, \dots, N$), satisfy the following set of N linear equations, for $i = 1, 2, \dots, N$ as $\rho \uparrow 1$,

$$\begin{aligned} & \left(\frac{1 + 2\rho_i - \rho_i^3}{2(1 + \rho_i)} - \sum_{l=1}^{i-1} F_{i,l}^{(i)} - \sum_{l=i+1}^N E_{l,i}^{(i)} \right) \text{Var}[C_i] \\ & - \left(\frac{1}{2(1 + \rho_i)} + \sum_{l=1}^{i-1} F_{i,l}^{(i+1)} + \sum_{l=i+1}^N E_{l,i}^{(i+1)} \right) \text{Var}[C_{i+1}] \\ & - \sum_{k \neq i, i+1} \left(\sum_{l=1}^{i-1} F_{i,l}^{(k)} + \sum_{l=i+1}^N E_{l,i}^{(k)} \right) \text{Var}[C_k] \sim \frac{H_i \rho_i}{1 + \rho_i} + \sum_{l=1}^{i-1} F_{i,l}^{(0)} + \sum_{l=i+1}^N E_{l,i}^{(0)}, \end{aligned} \quad (6)$$

where the constant H_i is given by, for $i = 1, 2, \dots, N$, $\rho < 1$,

$$H_i := \lambda_i E[C_i] (\text{Var}[B_i] + \rho_i^2 \text{Var}[A_i]) + \text{Var}[R_{i+1}], \quad (7)$$

and where the coefficients $E_{i,j}^{(k)}$ and $F_{i,j}^{(k)}$ are recursively defined by as $\rho \uparrow 1$,

$$E_{i,j}^{(0)} \sim (a_i - \rho_i e_j) E_{i-1,j}^{(0)} - a_i f_j E_{i-1,j+1}^{(0)} + f_j E_{i,j+1}^{(0)} + \frac{H_{i-1} \rho_i}{a_{i-1} \rho_{i-1}}, \quad \text{for } i - j = 2, \quad (8)$$

and for $k = 1, 2, \dots, N$ as $\rho \uparrow 1$,

$$E_{i,j}^{(k)} \sim (a_i - \rho_i e_j) E_{i-1,j}^{(k)} - a_i f_j E_{i-1,j+1}^{(k)} + f_j E_{i,j+1}^{(k)}, \quad \text{for } i - j = 2, \quad (9)$$

and for $k = 0, 1, \dots, N$ as $\rho \uparrow 1$,

$$E_{i,j}^{(k)} \rightsquigarrow (a_i - \rho_i e_j) E_{i-1,j}^{(k)} - a_i f_j E_{i-1,j+1}^{(k)} + f_j E_{i,j+1}^{(k)}, \quad \text{for } i - j \geq 3, \quad (10)$$

$$F_{i,j}^{(k)} \rightsquigarrow (a_i - \rho_i e_j) F_{i-1,j}^{(k)} - a_i f_j F_{i-1,j+1}^{(k)} + f_j F_{i,j+1}^{(k)}, \quad \text{for } i - j \geq 2, \quad (11)$$

with initial conditions, for $j = 1, 2, \dots, N$ as $\rho \uparrow 1$,

$$E_{j,j}^{(0)} \rightsquigarrow H_j, \quad (12)$$

$$E_{j,j}^{(k)} \rightsquigarrow \begin{cases} \rho_j^2, & k = j, \\ 0, & \text{else,} \end{cases} \quad (13)$$

and, for $j = 1, 2, \dots, N - 1$ as $\rho \uparrow 1$,

$$E_{j+1,j}^{(0)} \rightsquigarrow \frac{H_j \rho_{j+1}}{1 + \rho_j}, \quad (14)$$

$$E_{j+1,j}^{(k)} \rightsquigarrow \begin{cases} \frac{\rho_j(1+2\rho_j)\rho_{j+1}}{2(1+\rho_j)}, & k = j, \\ + \frac{\rho_j \rho_{j+1}}{2(1+\rho_j)}, & k = j + 1, \\ 0, & \text{else.} \end{cases} \quad (15)$$

Moreover, for $j = 1, 2, \dots, N$ as $\rho \uparrow 1$,

$$F_{j,j}^{(0)} \rightsquigarrow \frac{H_j \rho_j}{1 + \rho_j}, \quad (16)$$

$$F_{j,j}^{(k)} \rightsquigarrow \begin{cases} \frac{\rho_j(1+2\rho_j+2\rho_j^3)}{2(1+\rho_j)}, & k = j, \\ - \frac{\rho_j}{2(1+\rho_j)}, & k = j + 1, \\ 0, & \text{else.} \end{cases} \quad (17)$$

and, for $j = 1, 2, \dots, N - 1$ as $\rho \uparrow 1$,

$$F_{j+1,j}^{(0)} \rightsquigarrow \frac{e_j \rho_{j+1}}{1 + \rho_j} H_j + \frac{f_j \rho_{j+1}}{1 + \rho_{j+1}} H_{j+1}, \quad (18)$$

$$F_{j+1,j}^{(k)} \rightsquigarrow \begin{cases} \frac{e_j \rho_j(1+2\rho_j)\rho_{j+1}}{2(1+\rho_j)}, & k = j, \\ + \frac{e_j \rho_j \rho_{j+1}}{2(1+\rho_j)} + \frac{f_j \rho_{j+1}(1+2\rho_{j+1}+2\rho_{j+1}^3)}{2(1+\rho_{j+1})}, & k = j + 1, \\ - \frac{f_j \rho_{j+1}}{2(1+\rho_{j+1})}, & k = j + 2, \\ 0, & \text{else.} \end{cases} \quad (19)$$

Finally, the constants a_i , e_i and f_i are defined as, respectively, for $i = 1, 2, \dots, N$ as $\rho \uparrow 1$,

$$a_i \rightsquigarrow \frac{\rho_i(1 + \rho_{i-1})}{\rho_{i-1}}, \quad e_i \rightsquigarrow \frac{\rho_i}{1 + \rho_i} \quad \text{and} \quad f_i \rightsquigarrow \frac{1}{a_{i+1}}. \quad (20)$$

The complexity of the set (6) prevents us from solving it explicitly in general, but we do obtain closed-form expressions in the following cases. First, if we restrict our attention to a specific weighted sum of the solutions for $\text{Var}[C_i]$, $i = 1, \dots, N$, we obtain an explicit closed-form expression immediately leading to the PCL of the model under consideration (see Subsection 3.2). Second, we can apply asymptotic expansions to find asymptotically exact closed-form expressions for the dominating factors of $\text{Var}[C_i]$ by analyzing a scaled version of (6) in combination with the just derived PCL (see Subsection 3.3).

This subsection is closed with two remarks.

Remark 3.1. The asymptotic approach expounded in the present section is exact for Poisson processes under any traffic intensity $\rho < 1$, cf. [10] (implying that all \rightsquigarrow -signs could be replaced by =-signs), where we note that in this case, for $i = 1, 2, \dots, N$,

$$c_{A_i}^2 = 1, \quad (21)$$

and where the constant H_i reduces to, for $i = 1, 2, \dots, N$,

$$H_i = \lambda_i E[C_i] b_i^{(2)} + \text{Var}[R_{i+1}]. \quad (22)$$

Remark 3.2. In [2], the identity (4) and the set (6) are actually derived only for the special case of *deterministic* setup times; details of the derivation in case of *stochastic* setup times leading to (4) and (6) in full generality are available from the authors of the present note by request.

3.2 Asymptotic pseudo-conservation law

By working out an expression for the weighted sum of the solutions for $\text{Var}[C_i]$ of the set (6), the present subsection derives a PCL for the mean delays for the model described in Section 2 as shown in the following lemma.

Lemma 3.3. *As $\rho \uparrow 1$, we have*

$$\sum_{i=1}^N \rho_i E[W_i] \sim \frac{\rho}{2r} \sum_{i=1}^N H_i + \frac{r}{2(1-\rho)} \sum_{i=1}^N \rho_i (1 + \rho_i) + \sum_{i=1}^N \rho_i \frac{(c_{A_i}^2 - 1) b_i^{(1)}}{2}. \quad (23)$$

Proof: Starting point of our proof is the set of equations for $\text{Var}[C_i]$ given by (6) in the special case of Poisson arrivals. Recall that in this Poisson case (6) is exact under any traffic intensity $\rho < 1$ and H_i is given by (22). Notice that the coefficient matrix in the lefthand side of (6) is *independent* of H_i and that the righthand side of (6) is a *linear* function of H_i implying that the solutions for $\text{Var}[C_i]$ of (6) are linear functions of H_i as well. That is, for $i = 1, 2, \dots, N$ and $\rho < 1$,

$$\text{Var}[C_i] = g_i(H_1, H_2, \dots, H_N), \quad (24)$$

where $g_i : \mathbb{R}^N \rightarrow \mathbb{R}$ are (unknown) linear functions of H_1, \dots, H_N , i.e., there exist constants $c_{i,j}$ ($i = 1, \dots, N, j = 0, 1, \dots, N$) such that for $i = 1, \dots, N$,

$$g_i(H_1, \dots, H_N) = c_{i,0} + \sum_{j=1}^N c_{i,j} H_j. \quad (25)$$

In order to find a closed-form expression for a weighted sum of these functions, we use the PCL for gated polling systems with Poisson arrivals (cf. [4]): For $\rho < 1$,

$$\sum_{i=1}^N \rho_i E[W_i] = \frac{\rho}{2r} \sum_{i=1}^N H_i + \frac{r}{2(1-\rho)} \sum_{i=1}^N \rho_i (1 + \rho_i). \quad (26)$$

For this Poisson case, by using simple balance arguments the mean delay at Q_i can be expressed in terms of the first two moments of C_i as follows: For $i = 1, 2, \dots, N$ and $\rho < 1$ (cf. [11]),

$$E[W_i] = \frac{1 + \rho_i}{2} \left(\frac{\text{Var}[C_i]}{E[C_i]} + E[C_i] \right) = \frac{1 + \rho_i}{2(1-\rho)} \left(\frac{g_i(H_1, H_2, \dots, H_N) (1-\rho)^2}{r} + r \right), \quad (27)$$

where the last equality follows from application of (5) and (25). Subsequently, substituting (27) into (26) yields the following weighted sum of $g_i(H_1, H_2, \dots, H_N)$ in the Poisson case: For $\rho < 1$,

$$\sum_{i=1}^N \rho_i (1 + \rho_i) g_i(H_1, H_2, \dots, H_N) = \frac{\rho}{1-\rho} \sum_{i=1}^N H_i. \quad (28)$$

Returning to the general case of renewal arrivals, (6) states that asymptotically $\text{Var}[C_i]$ ($i = 1, 2, \dots, N$) satisfy the *same set of linear equations* as in the Poisson case, where the variables

H_i ($i = 1, 2, \dots, N$) are defined as in (7). Due to the fact that the coefficient matrix in the lefthand side of (6) is a linear invertible mapping in conjunction with the fact that the H_i ($i = 1, 2, \dots, N$), defined in (7), only show up at the right-hand side of (6), we have that as $\rho \uparrow 1$, $i = 1 \dots, N$,

$$\text{Var}[C_i] \rightsquigarrow g_i(H_1, \dots, H_N) = c_{i,0} + \sum_{j=1}^N c_{i,j} H_j, \quad (29)$$

where the last equality follows from (25). Note that the variables H_i ($i = 1, 2, \dots, N$) are generally not the same as in the Poisson case (see Remark 3.1). Here, the crucial observation is that the coefficients $c_{i,j}$ in (29) are the same as those in the Poisson case (25). This immediately implies that (28) remains asymptotically true for renewal arrivals, i.e., as $\rho \uparrow 1$,

$$\sum_{i=1}^N \rho_i (1 + \rho_i) g_i(H_1, H_2, \dots, H_N) \rightsquigarrow \frac{\rho}{1 - \rho} \sum_{i=1}^N H_i. \quad (30)$$

Finally, calling upon (4) in combination with (29) completes the proof. \square

The above PCL is exact for Poisson arrival processes under any traffic intensity $\rho < 1$ and, therewith, generalizes the PCL in gated polling systems with Poisson arrivals [4]. Although such a PCL does not give explicit expressions for the mean delays themselves, it appears instrumental in constructing approximations and providing tests for the accuracy of simulations, numerical calculations and approximations (see again [4]). Further, it gives a relatively simple closed-form expression for the weighted sum of the mean delays, which may be used as a first indication of overall system performance.

3.3 Mean asymptotic scaled delays

As mentioned earlier, the set (6) can in general not be solved in closed form, but the present subsection finds explicit expressions for the dominating terms of $\text{Var}[C_i]$ in HT. Thereto, we multiply both sides of (6) by $(1 - \rho)^2$ and let $\rho \uparrow 1$, which renders the corresponding scaled set, for $i = 1, 2, \dots, N$,

$$\begin{aligned} \left(\frac{1 + 2\hat{\rho}_i - \hat{\rho}_i^3}{2(1 + \hat{\rho}_i)} - \sum_{l=1}^{i-1} F_{i,l}^{(i)} - \sum_{l=i+1}^N E_{l,i}^{(i)} \right) \xi_i - \left(\frac{1}{2(1 + \hat{\rho}_i)} + \sum_{l=1}^{i-1} F_{i,l}^{(i+1)} + \sum_{l=i+1}^N E_{l,i}^{(i+1)} \right) \xi_{i+1} \\ - \sum_{k \neq i, i+1} \left(\sum_{l=1}^{i-1} F_{i,l}^{(k)} + \sum_{l=i+1}^N E_{l,i}^{(k)} \right) \xi_k = 0, \end{aligned} \quad (31)$$

where ξ_i represents the variance of the asymptotic scaled i -cycle, i.e., for $i = 1, 2, \dots, N$,

$$\xi_i = \lim_{\rho \uparrow 1} (1 - \rho)^2 \text{Var}[C_i], \quad (32)$$

where the existence of the limit is guaranteed by the fact that $E[W_i]$ has a first-order pole at $\rho = 1$ in conjunction with (4) and (5). The set (31) can be solved up to some unknown scaling factor $c \in \mathbb{R}$ as shown in the following lemma.

Lemma 3.4. *The solution of the set (31) is given by, for $i = 1, 2, \dots, N$,*

$$\xi_i = c, \quad (33)$$

with $c \in \mathbb{R}$.

Proof: One can verify that in (31), for $i = 1, 2, \dots, N$,

$$\frac{2\hat{\rho}_i - \hat{\rho}_i^3}{2(1 + \hat{\rho}_i)} - \sum_{k=1}^N \left(\sum_{l=1}^{i-1} F_{i,l}^{(k)} + \sum_{l=i+1}^N E_{l,i}^{(k)} \right) = 0, \quad (34)$$

which shows that (33) is indeed a solution of the *homogeneous* set (31). Either by elementary, but tedious, row and column operations or by quoting from [14] we observe that the rank of the coefficient matrix of (31) equals $N - 1$, which completes the proof. \square

Since the dimension of the null space of the coefficient matrix of (31) equals one, adding a single *non-homogeneous* equation would render a unique solution for the unknown scaling factor c . This additional equation can be readily obtained from a scaled version of the PCL (23) as done in the lemma below.

Lemma 3.5. *The quantity c is given by*

$$c = \frac{r\sigma^2}{\sum_{i=1}^N \hat{\rho}_i(1 + \hat{\rho}_i)}. \quad (35)$$

Proof: Via Lemma 3.4 in combination with (4) and (5), one obtains the mean asymptotic scaled delays, for $i = 1, 2, \dots, N$,

$$\omega_i = \frac{(1 + \hat{\rho}_i)}{2} \left(\frac{c}{r} + r \right), \quad (36)$$

which satisfies a scaled version of the PCL (23). That is, multiplying both sides of (23) by $(1 - \rho)$ and letting $\rho \uparrow 1$ yields

$$\sum_{i=1}^N \hat{\rho}_i \omega_i = \frac{\sigma^2}{2} + \frac{r}{2} \sum_{i=1}^N \hat{\rho}_i(1 + \hat{\rho}_i), \quad (37)$$

where we have used the definition of σ^2 as given in (3). Combining (36) and (37) completes the proof. \square

Lemma 3.5 has the following immediate consequence for the mean asymptotic scaled delay ω_i at each of the queues, which is the main result of the present note.

Corollary 3.6. *For $i = 1, 2, \dots, N$,*

$$\omega_i = \frac{(1 + \hat{\rho}_i)}{2} \left(\frac{\sigma^2}{\sum_{j=1}^N \hat{\rho}_j(1 + \hat{\rho}_j)} + r \right). \quad (38)$$

For Poisson arrival processes, the result in Corollary 3.6 has been obtained before in the literature, see, e.g., [13]. For general renewal arrivals, only conjectures [5, 6, 8] have been known so far and as such our approach is the first to give a rigorous proof of these conjectures. We close this subsection with a remark.

Remark 3.7. In Section 2, the assumption is made that the mean delay incurred at each of the queues, considered as a function of ρ , has a first order pole at $\rho = 1$. However, the approach presented here is actually the first to rigorously prove the validity of this assumption in case of renewal arrivals. In fact, the scaled version of the PCL in (37), which dictates that $\sum_{i=1}^N \rho_i E[W_i]$ has a first-order pole at $\rho = 1$, together with (36) imply that, for $i, j = 1, 2, \dots, N$,

$$\lim_{\rho \uparrow 1} \frac{E[W_i]}{E[W_j]} = \frac{1 + \hat{\rho}_i}{1 + \hat{\rho}_j}, \quad (39)$$

which in turn implies that $E[W_i]$, $i = 1, 2, \dots, N$, indeed has a first-order pole at $\rho = 1$.

3.4 Implications

The following properties about the impact of the system parameters on the mean asymptotic scaled delay are revealed by Theorem 2.1.

Corollary 3.8 (Insensitivity).

For $i = 1, 2, \dots, N$, the mean asymptotic scaled delay ω_i ,

- (1) depends on the interarrival-time distributions only through σ^2 , defined in Equation (3);
- (2) is independent of the visit order;
- (3) depends on the second moments of the service-time distributions only through $b^{(2)}$, i.e., the second moment of the service time of an arbitrary customer;
- (4) depends on the switch-over time distributions only through the first moment of the total switch-over time in a cycle.

The insensitivity properties summarized in Corollary 1 are, in general, not true for stable systems, i.e., for $\rho < 1$.

Corollary 3.9 (Approximations).

For $\rho < 1$, the expected delay at Q_i can be approximated by, for $i = 1, \dots, N$,

$$E[W_i] \approx \frac{\omega_i}{1 - \rho}, \tag{40}$$

where ω_i is given by (38).

The accuracy of this approximation is validated in [8]; the results show that the approximation is highly accurate when the load is roughly 80% or more.

4 Topics for Further Research

In this note we have proposed a new method to rigorously prove HT limits for the mean asymptotic scaled delay in a gated polling model with a general number of queues under the assumption of general renewal arrivals. The method might be extended to derive HT results for the complete waiting-time distributions as well. That is, decomposition results for the waiting time distributions obtained in [3] may form a starting point to obtain such results, opening up a very challenging area for further research.

References

- [1] E. Altman, P. Konstantopoulos, Z. Liu (1992). Stability, monotonicity and invariant quantities in general polling systems. *Queueing Systems* 11, 35-57.
- [2] D. Bertsimas and G. Mourtzinou (1997). Multiclass queueing systems in heavy traffic: an asymptotic approach based on distributional and conservation laws. *Oper. Res.* 45, 470-487.
- [3] D. Bertsimas and G. Mourtzinou (1999). Decomposition results for general polling systems and their applications. *Queueing Systems* 31, 295-316.
- [4] O.J. Boxma (1989). Workloads and waiting times in single-server systems with multiple customer classes. *Queueing Systems* 5, 185-214.
- [5] E.G. Coffman, A.A. Puhalskii and M.I. Reiman (1995). Polling systems with zero switch-over times: a heavy-traffic principle. *Ann. Appl. Prob.* 5, 681-719.
- [6] E.G. Coffman, A.A. Puhalskii and M.I. Reiman (1998). Polling systems in heavy-traffic: a Bessel process limit. *Math. Oper. Res.* 23, 257-304.

- [7] D. Down (1998). On the stability of polling models with multiple servers. *Journal of Applied Probability* 35, 925-935.
- [8] T.L. Olsen and R.D. van der Mei (2005). Periodic polling systems in heavy-traffic: renewal arrivals. *OR Letters* 33, 17-25.
- [9] J.A.C. Resing (1993). Polling systems and multitype branching processes. *Queueing Systems* 13, 409-426.
- [10] D. Sarkar and W.I. Zangwill (1989). Expected waiting time for nonsymmetric cyclic queueing systems - Exact results and applications. *Mgmt. Sc.* 35, 1463-1474.
- [11] H. Takagi (1986). *Analysis and of Polling Systems* (MIT Press, Cambridge, MA).
- [12] H. Takagi (2000). Analysis and application of polling models. In: *Performance Evaluation: Origins and Directions*, eds. G. Haring, C. Lindemann and M. Reiser, (Lecture Notes in Computer Science 1769, Springer, Berlin), 423-442.
- [13] R.D. van der Mei (1999). Distributions of the delay in polling systems in heavy traffic. *Perf. Eval.* 38, 133-148.
- [14] R.D. van der Mei and E.M.M. Winands (2007). Polling models with renewal arrivals: a new method to derive heavy-traffic asymptotics. To appear in *Perf. Eval.*