A note on polling models with renewal arrivals and nonzero switch-over times
Mei, van der, R.D.; Winands, E.M.M.

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 author's 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

Citation for published version (APA):
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 \( Q_1, \ldots, Q_N \), 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 \( \rho \), we denote its values evaluated at \( \rho = 1 \) by \( \hat{x} \). Furthermore, we use the notation that \( h(x) \sim g(x) \) as \( x \to 1 \) means that \( \lim_{x \to 1} h(x)/g(x) = 1 \). 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 \( \rho \) 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 \( i = 1, 2, \ldots, N \),

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

where \( g(x) = o(f(x)) \) means that \( g(x)/f(x) \to 0 \) as \( x \to 1 \). More colloquially, we can say that \( \omega_i \), 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 \( i = 1, 2, \ldots, N \),

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

with

\[
\sigma^2 := \sum_{i=1}^{N} \lambda_i \left( Var[B_i] + \hat{\rho}_i Var[\hat{A}_i] \right).
\]
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 $A_i$ ($i = 1, \ldots, N$) fixed up to a common scaling constant $\rho$ (i.e., $A_i \sim \rho I A_i$, where $I A_i$ ($i = 1, \ldots, N$) 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 $i = 1, 2, \ldots, N$ (cf. [2]),

$$E[W_i] \sim \frac{(1 + \rho)}{2} \left( \frac{\text{Var}[C_i]}{E[C_i]} + E[C_i] \right) + \frac{(c_i^2 - 1) b_i(1)}{2},$$

(4)

where $c_i^2$ 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 $i = 1, 2, \ldots, N$ and $\rho < 1$ (see, e.g., [1]),

$$E[C_i] = \frac{r}{1 - \rho},$$

(5)

whereas the variances of the cycle lengths $\text{Var}[C_i]$ ($i = 1, \ldots, N$), 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 $\text{Var}[C_i]$ ($i = 1, \ldots, N$), satisfy the following set of $N$ linear equations, for $i = 1, 2, \ldots, N$ as $\rho \uparrow 1$,

$$\left( \frac{1 + 2\rho_i - \rho_i^3}{2(1 + \rho_i)} - \sum_{l=1}^{i-1} F_{i,l}^{(i)} - \sum_{l=i+1}^{N} E_{i,l}^{(i)} \right) \text{Var}[C_i]$$

$$- \left( \frac{1}{2(1 + \rho_i)} + \sum_{l=1}^{i-1} F_{i,l}^{(i+1)} + \sum_{l=i+1}^{N} E_{i,l}^{(i+1)} \right) \text{Var}[C_{i+1}]$$

$$- \sum_{k \neq i, i+1} \left( \sum_{l=1}^{i-1} F_{i,l}^{(k)} + \sum_{l=i+1}^{N} E_{i,l}^{(k)} \right) \text{Var}[C_k] \sim \frac{H_i \rho_i}{1 + \rho_i} + \sum_{l=1}^{i-1} F_{i,l}^{(0)} + \sum_{l=i+1}^{N} E_{i,l}^{(0)},$$

(6)

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

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

(7)

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

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

(8)

for $i - j = 2$,

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

$$E_{i,j}^{(k)} \sim (a_i - \rho_i e_j) E_{i-1,j}^{(k)} - a_i f_j E_{i-1,j+1}^{(k)} + f_j E_{i,j+1}^{(k)},$$

(9)

for $i - j = 2$,
and for $k = 0, 1, \ldots, N$ as $\rho \uparrow 1$,
\[
E_{i,j}^{(k)} = (a_i - \rho_i e_i \rho_{i-1})E_{i,j}^{(k)} - a_i f_j E_{i-1,j}^{(k)} - a_i f_j E_{i,j+1}^{(k)} + f_j E_{i,j+1}^{(k)} + f_j E_{i,j+1}^{(k)},
\]
for $i - j \geq 3$, \hspace{1cm} (10)
\[
F_{i,j}^{(k)} = (a_i - \rho_i e_i \rho_{i-1})F_{i,j}^{(k)} - a_i f_j F_{i-1,j}^{(k)} - a_i f_j F_{i,j+1}^{(k)} + f_j F_{i,j+1}^{(k)},
\]
for $i - j \geq 2$, \hspace{1cm} (11)

with initial conditions, for $j = 1, 2, \ldots, N$ as $\rho \uparrow 1$,
\[
E_{j,j}^{(0)} \sim H_j,
\]
\[
E_{j,j}^{(k)} \sim \begin{cases} 
\rho_j^2, & k = j, \\
0, & \text{else},
\end{cases}
\]
and, for $j = 1, 2, \ldots, N - 1$ as $\rho \uparrow 1$,
\[
E_{j,j+1}^{(0)} \sim \frac{H_j \rho_j}{1 + \rho_j},
\]
\[
E_{j,j+1}^{(k)} \sim \begin{cases} 
\frac{\rho_j^2}{1 + \rho_j}, & k = j, \\
\frac{\rho_j^3}{1 + \rho_j}, & k = j + 1, \\
0, & \text{else},
\end{cases}
\]

Moreover, for $j = 1, 2, \ldots, N$ as $\rho \uparrow 1$,
\[
E_{j,j}^{(0)} \sim \frac{H_j \rho_j}{1 + \rho_j},
\]
\[
E_{j,j}^{(k)} \sim \begin{cases} 
\frac{\rho_j^2}{1 + \rho_j}, & k = j, \\
\frac{\rho_j^3}{1 + \rho_j}, & k = j + 1, \\
0, & \text{else},
\end{cases}
\]
and, for $j = 1, 2, \ldots, N - 1$ as $\rho \uparrow 1$,
\[
F_{j+1,j}^{(0)} \sim \frac{e_j \rho_j}{1 + \rho_j} H_j + \frac{f_j \rho_j}{1 + \rho_j} H_{j+1},
\]
\[
F_{j+1,j}^{(k)} \sim \begin{cases} 
\frac{e_j \rho_j (1 + 2 \rho_j) \rho_{j+1}}{2(1 + \rho_j)} & k = j, \\
\frac{f_j \rho_j (1 + 2 \rho_j) \rho_{j+1}}{2(1 + \rho_j)} & k = j + 1, \\
0 & \text{else},
\end{cases}
\]

Finally, the constants $a_i$, $e_i$ and $f_i$ are defined as, respectively, for $i = 1, 2, \ldots, N$ as $\rho \uparrow 1$,
\[
a_i \sim \frac{\rho_i (1 + \rho_{i-1})}{\rho_{i-1}}, \quad e_i \sim \frac{\rho_i}{1 + \rho_i} \quad \text{and} \quad f_i \sim \frac{1}{a_{i+1}}.
\]

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 $\text{Var}[C_i], i = 1, \ldots, N$, 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 $\sim$-signs could be replaced by $\approx$-signs), where we note that in this case, for $i = 1, 2, \ldots, N$,
\[
c_i^2 A_i = 1,
\]

(21)
Lemma 3.3. shown in the following lemma.

The present subsection derives a PCL for the mean delays for the model described in Section 2 as \( \rho < 1 \) and case of Poisson arrivals. Recall that in this Poisson case (6) is exact under any traffic intensity \( \rho < 1 \).

The starting point of our proof is the set of equations for \( V a r[C_i] \) given by (6) in the special case of renewal arrivals. 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 \( V a r[C_i] \) of (6) are linear functions of \( H_i \) as well. That is, for \( i = 1, 2, \ldots, N \) and \( \rho < 1 \),

\[
V a r[C_i] = g_i(H_1, H_2, \ldots, H_N),
\]

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

\[
g_i(H_1, \ldots, H_N) = c_{i,0} + \sum_{j=1}^{N} c_{i,j} H_j.
\]

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 \( \rho < 1 \),

\[
\sum_{i=1}^{N} \rho_i E[W_i] = \frac{\rho}{2\nu} \sum_{i=1}^{N} H_i + \frac{r}{2(1-\rho)} \sum_{i=1}^{N} \rho_i (1 + \rho_i) + \sum_{i=1}^{N} \rho_i \left( \frac{\mu_i^2 - 1}{2} \right).
\]

Proof: Starting point of our proof is the set of equations for \( V a r[C_i] \) 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 \( V a r[C_i] \) of (6) are linear functions of \( H_i \) as well. That is, for \( i = 1, 2, \ldots, N \) and \( \rho < 1 \),

\[
V a r[C_i] = g_i(H_1, H_2, \ldots, H_N),
\]

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

\[
g_i(H_1, \ldots, H_N) = c_{i,0} + \sum_{j=1}^{N} c_{i,j} H_j.
\]

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 \( \rho < 1 \),

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

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 \( i = 1, 2, \ldots, N \) and \( \rho < 1 \) (cf. [11]),

\[
E[W_i] = \frac{1 + \rho_i}{2} \left( V a r[C_i] \right) + E[C_i] = \frac{1 + \rho_i}{2(1-\rho)} \left( g_i(H_1, H_2, \ldots, H_N) \frac{(1-\rho)^2}{r} + r \right),
\]

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

\[
\sum_{i=1}^{N} \rho_i (1 + \rho_i) g_i(H_1, H_2, \ldots, H_N) = \frac{\rho}{1-\rho} \sum_{i=1}^{N} H_i.
\]

Returning to the general case of renewal arrivals, (6) states that asymptotically \( V a r[C_i] \) (\( i = 1, 2, \ldots, N \)) satisfy the same set of linear equations as in the Poisson case, where the variables
$H_i \ (i = 1, 2, \ldots, N)$ 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 $H_i \ (i = 1, 2, \ldots, N)$, defined in (7), only show up at the right-hand side of (6), we have that as $\rho \uparrow 1, \ i = 1, \ldots, N$,

$$Var[C_i] \sim g_i(H_1, \ldots, H_N) = c_i,0 + \sum_{j=1}^{N} c_{i,j} H_j,$$

where the last equality follows from (25). Note that the variables $H_i \ (i = 1, 2, \ldots, N)$ 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 $\rho \uparrow 1$,

$$\sum_{i=1}^{N} \rho_i (1+\rho_i) g_i(H_1, H_2, \ldots, H_N) \sim \frac{\rho}{1-\rho} \sum_{i=1}^{N} H_i.$$  \hspace{1cm} (30)

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

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 $Var[C_i]$ 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 $i = 1, 2, \ldots, N$,

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

$$- \sum_{k \neq i,i+1} \left(\sum_{l=1}^{i-1} F_{i,l} + \sum_{l=i+1}^{N} E_{l,i}\right) \xi_k = 0,$$

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

$$\xi_i = \lim_{\rho \downarrow 1} (1-\rho)^2 Var[C_i],$$

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 $i = 1, 2, \ldots, N$,

$$\xi_i = c,$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 $\rho$, 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 $i, j = 1, 2, \ldots, N$,

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

which in turn implies that $E[W_i], i = 1, 2, \ldots, N$, indeed has a first-order pole at $\rho = 1$. 7
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 $i = 1, 2, \ldots, N$, the mean asymptotic scaled delay $\omega_i$,

1. depends on the interarrival-time distributions only through $\sigma^2$, 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 $i = 1, \ldots, N$,

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

where $\omega_i$ 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


