Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

W.P. Groenendijk

A conservation-law based approximation algorithm for waiting times in polling systems

Department of Operations Research and System Theory Report OS-R8816 September

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

A Conservation-Law Based Approximation Algorithm for Waiting Times in Polling Systems

W.P. Groenendijk
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

In this paper an algorithm is derived for the approximation of mean waiting times in multi-queue cyclic-service systems with a single server. Mixed service strategies are allowed. The approximation is based on a pseudoconservation law for mean waiting times in these systems. Through numerous numerical examples it is illustrated that, even for cases with high loads and asymmetric traffic, the approximation is accurate. The numerical results are compared with simulation and with previous approximations.

1980 Mathematics Subject Classification: 60K25, 68M20 69C40
Key Words & Phrases: polling system, pseudoconservation law, mixed service strategy, waiting-time approximations
Note: This report will be submitted for publication elsewhere.

1. Introduction

Recently, many studies have been published concerning the analysis of cyclic-service queues, cf. the survey by Takagi [16]. The research has mainly focussed on mean waiting times. In special cases, exact results have been obtained for weighted sums of mean waiting times (Ferguson & Aminetzah [7], Watson [17], Boxma & Groenendijk [1,2]) or even for individual mean waiting times (Ferguson & Aminetzah [7], Takagi [15]). Nevertheless, even in the latter case, simple approximations would be highly preferred, since calculation of these exact results often requires much computing effort.

This paper presents such an approximation for mean waiting times in multi-queue, cyclic-service systems with a single server. The algorithm is very simple in the sense that time and memory requirements on a computer are negligible. The approximation is based on a pseudoconservation law derived in Boxma & Groenendijk [1] and combines ideas of Srinivasan [13] and Groenendijk [10].

The organization of this paper is as follows. First we present a mathematical description of the model. Then we introduce the concept of pseudoconservation laws and give some general cycle-time results. In Section 2 the approximation algorithm is derived. Finally, in Section 3 we discuss the validity of the approximation algorithm, and present some numerical results.

Model description

The model under consideration consists of N queues, $Q_1, \dots, Q_N$. Type-i customers arrive at Q_i according to a Poisson process with intensity λ_i, $i=1,\dots,N$. The service requests of type-i customers are independent, identically distributed stochastic variables with distribution $B_i(\cdot)$, with Laplace Stieltjes Transform (LST) and first and second moment given by $\beta_i(\cdot)$, β_i and $\beta_i^{(2)}$ respectively. The offered traffic at Q_i, ρ_i, is defined as,

$$\rho_i := \lambda_i \beta_i, \quad i=1,\dots,N.$$

The total offered traffic, ρ, is defined as

$$\rho := \sum_{i=1}^{N} \rho_i.$$

The queues are served in cyclic order by a single server S. The switch-over times of the server between the ith and $(i+1)$th queue are independent, identically distributed stochastic variables with

Report OS-R8816
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

first moment s_i and second moment $s_i^{(2)}$. The first moment s of the total switch-over time during a cycle of the server is given by:

$$s := \sum_{i=1}^{N} s_i,$$

its second moment is denoted by $s^{(2)}$. It is assumed that the interarrival processes, the service processes and the switch-over processes are mutually independent.
When upon his arrival S finds Q_i empty, S immediately begins to switch to Q_{i+1}. Otherwise S acts as follows, depending on the service strategy at Q_i:

- Exhaustive service (E): S serves type-i customers until Q_i is empty
- Gated service (G): S serves only those type-i customers present upon his arrival at Q_i (a gate closes upon his arrival)
- 1-Limited service (1L): S serves exactly one customer

We will allow mixed service strategies (e.g., exhaustive at Q_1 and Q_3, gated at Q_4 and Q_5 and 1-limited at Q_2 and $Q_6, \ldots, Q_N$). The order of service at each queue is assumed to be FCFS.

The recent discovery of 'pseudoconservation laws' by Watson [17] and Ferguson & Aminetzah [7] has been of great importance. These laws are exact expressions for a weighted sum of the mean waiting times at the various queues of the cyclic-service system. In particular, they offer an excellent starting point for the construction and testing of approximation algorithms. The pseudoconservation laws are generalized by Boxma & Groenendijk [1] to allow mixed service strategies at the various queues. For the model described above, the pseudoconservation law derived in [1] reduces to the following expression. Denote by e the group of exhaustive queues, by g the group of gated queues, and by $1l$ the group of 1-limited queues. Assume that $\rho<1$, and that, for all $i \in 1l$, $\rho+\lambda_i s<1$. According to Szpankowski & Rego [14] this ensures that the stationary distributions of the waiting times exist. Denote by EW_i the mean waiting time at Q_i. Then

$$\sum_{i \in e,g} \rho_i EW_i + \sum_{i \in 1l} \rho_i [1-\frac{\lambda_i s}{1-\rho}] EW_i = \rho \sum_{\forall i} \frac{\lambda_i \beta_i^{(2)}}{2(1-\rho)} + \rho \frac{s^{(2)}}{2s} + \frac{s}{2(1-\rho)} [\rho^2 - \sum_{i \in e} \rho_i^2 + \sum_{i \in g,1l} \rho_i^2]. \qquad (1.1)$$

A number of recent approximations for mean waiting times in polling systems have been based on the pseudoconservation laws. For a system with exhaustive service at all queues, the mean waiting-time approximation by Bux & Truong [5] appears to fulfill the pseudoconservation law - but this approximation was suggested even before this law was discovered! For systems with either exhaustive or gated service at all queues, a mean waiting-time approximation based on the pseudoconservation law has been devised by Everitt [6], and for systems with 1-limited service at all queues by Boxma & Meister [4]. Groenendijk [10] combined these to obtain an approximation for systems with mixtures of exhaustive, gated and 1-limited queues. For the case of 1-limited service at all queues, Srinivasan [13] improved upon [4] by taking a more detailed look at (conditional) cycle times before eventually applying the pseudoconservation law. The ideas in the current paper are partly based on those of Srinivasan.

Fuhrmann & Wang [9] use an 'approximate conservation law' to obtain a mean waiting-time approximation for the notoriously difficult case of k-limited service at all queues. Pang & Donaldson [12] suggest a very accurate mean waiting-time approximation for discrete-time cyclic-service systems with gated service at all queues. They express the mean waiting time at Q_i in the second moment $v_{i,i}$ of the sum of the visit time at Q_i and the subsequent switch-over time; next they obtain a linear relation between $v_{i+1,i+1}$ and $v_{i,i}$ for all i; and finally they solve for the $v_{i,i}$ by deriving an extra linear relation between $v_{1,1}, \ldots, v_{N,N}$. At this last stage the conservation law is elegantly brought into the picture.

The approximation algorithm in [10] is very simple, and capable of providing qualitative as well as quantitative insight over a wide range of parameters. It provides an explicit formula for $E\mathbf{W}_i$, it is exact in the completely symmetric case (same traffic characteristics, switch-over time distributions and service strategies at all queues), and is an excellent approximation for low and medium traffic. There are however some cases (especially at high and asymmetric loads) in which this algorithm is not quite satisfactory. In this study we present an approximation algorithm for systems with mixtures of exhaustive, gated and 1-limited queues which handles cases with high loads much better. The approximation algorithm is however more complex than the one described in [10]; in particular it involves some iteration.

CYCLE-TIME RESULTS

Below we state some general results on cycle times and related quantities in the cyclic-service system with mixed service strategies. We define the cycle time for Q_i, $\mathbf{C}_i$, as the time between two successive arrivals of the server at Q_i. The visit time at Q_i, $\mathbf{V}_i$, is the time between the arrival of the server at Q_i and its subsequent departure from that queue. Finally, the intervisit time for Q_i, $\mathbf{I}_i$, is defined as

$$\mathbf{I}_i := \mathbf{C}_i - \mathbf{V}_i. \tag{1.2}$$

It may be easily seen that the *mean* cycle time $E\mathbf{C}_i$ is independent of i. Assuming ergodicity, we may balance the flow in and out of the system during a cycle. It follows that the mean cycle time, $E\mathbf{C}$, is equal to the sum of the total switch-over time, s, and the mean time the server is serving customers during an average cycle, $\rho E\mathbf{C}$. So,

$$E\mathbf{C} = \frac{s}{1-\rho}. \tag{1.3}$$

Balancing the flow of customers in and out of the system during a cycle shows that,

$$\lambda_i E\mathbf{C} = \frac{E\mathbf{V}_i}{\beta_i}, \tag{1.4}$$

and hence, from (1.3),

$$E\mathbf{V}_i = \frac{\rho_i s}{1-\rho}. \tag{1.5}$$

Next we introduce $E\mathbf{rc}_i$, the mean residual cycle time for Q_i. By using stochastic mean value theorems it is easily proven that, although successive cycle times for Q_i do not form a renewal process,

$$E\mathbf{rc}_i = \frac{E\mathbf{C}_i^2}{2E\mathbf{C}_i}; \tag{1.6}$$

for an alternative proof cf. [8], Theorem 4.5.1. We shall use $E\mathbf{rc}_i$ most often as the mean forward recurrence time for $\mathbf{C}_i$, i.e. as the expected time until the next arrival of S at Q_i.

2. DERIVATION OF THE APPROXIMATION ALGORITHM

In this section we specify the algorithm for approximating the mean waiting times at the various queues. The algorithm is based on an iteration scheme. In each step of the iteration:

i. We express all mean waiting times in the mean residual cycle time $E\mathbf{rc}_i$. As in [10] it may be shown, that:
   For Q_i gated:

$$E\mathbf{W}_i = (1+\rho_i)E\mathbf{rc}_i. \tag{2.1}$$

   Indeed, the mean waiting time of a tagged type-i customer consists of two components. Firstly, a mean residual cycle time $E\mathbf{rc}_i$, because due to the gating mechanism a customer is never served in

the cycle in which he arrives. Secondly, the mean time from the instant the server arrives at Q_i until the service completion of all type-i customers who arrived before the tagged customer in the same cycle: $(\lambda_i E\mathbf{rc}_i)\beta_i$.
For Q_i exhaustive:

$$E\mathbf{W}_i \approx (1-\rho_i)E\mathbf{rc}_i. \tag{2.2}$$

Actually, it may be proven that $E\mathbf{W}_i = (1-\rho_i)E\tilde{\mathbf{rc}}_i$, where $E\tilde{\mathbf{rc}}_i$ is the mean residual cycle time at Q_i with a cycle starting at a *departure* epoch of the server from Q_i.
For Q_i 1-limited we will derive an expression of the form

$$E\mathbf{W}_i \approx \frac{E\mathbf{rc}_i}{1-\lambda_i E\mathbf{C}_{b,i}} + H_i. \tag{2.3}$$

$E\mathbf{C}_{b,i}$ will be defined in (2.5). H_i will be a function of $E\mathbf{rc}_i$.

ii. We assume that $E\mathbf{rc}_i \equiv E\mathbf{rc}$ for all i, and hence (cf. (1.6) and (1.3)), that the second moment of the cycle time for Q_i is independent of i. Although in most cases this is clearly an approximation, the differences between the second moments of the cycle time are generally quite small (cf. the exact analysis in [3]).

To start the iteration we first take $H_i=0$. From (2.1), (2.2) and (2.3) we have N linear relations between $E\mathbf{W}_i$ and $E\mathbf{rc}_i$. Substituting these relations into the pseudoconservation law (1.1) and solving for $E\mathbf{rc}$, we obtain our first approximation for the mean residual cycle time, which we shall denote by $E\mathbf{rc}^{(1)}$. Note that this iteration step is just the approximation described in [10]. In the second step of the iteration, we use $E\mathbf{rc}^{(1)}$ to compute the extra terms H_i for all the 1-limited queues. Next the approximation is repeated, using the values for the H_i, and substituting the N linear relations into the pseudoconservation law. This yields $E\mathbf{rc}^{(2)}$, the second approximation for the mean residual cycle time. So we switch back and forth between the computation of the mean residual cycle time and the extra terms H_i. The iteration is stopped as soon as the mean waiting times in subsequent steps do not significantly change any more.

Deriving a suitable expression for H_i will be the main concern of this section. Assume that Q_i is served 1-limited. We first introduce some notation. Denote by $\mathbf{X}_i$ the number of type-i customers in the system found by an arbitrary type-i customer - the tagged customer - upon his arrival. Let,

$$p_i(n) := Pr\{\mathbf{X}_i = n\}. \tag{2.4}$$

Let us call the cycle in which the tagged (type-i) customer arrived the A-cycle, and the cycle following the A-cycle the B-cycle. Hence the B-cycle always contains a type-i service. Denote by A_i the indicator function of the event that, during the A-cycle, the arrival of the tagged customer takes place in a visit period of the server at Q_i (so $\bar{A}_i := (1-A_i)$ represents the indicator function of the event that, during the A-cycle, the arrival of the tagged customer takes place in an intervisit period of the server w.r.t. Q_i).

We will be needing some notation to distinguish between several types of conditional cycle times and intervisit times. First of all we define:

$$E\mathbf{C}_{b,i} := E[\mathbf{C}_i \,|\, \text{the cycle contains a service at } Q_i]. \tag{2.5}$$

This quantity plays an important role in several mean waiting-time approximations [4,10,11,13]. Next, denote by $E[\mathbf{C}_i^A \,|\, A_i]$ ($E[\mathbf{C}_i^B \,|\, A_i]$) the average length of an A-cycle (a B-cycle) given the tagged customer arrived during a visit period of the server at Q_i in the A-cycle. Let $E[\mathbf{I}_i^A \,|\, A_i]$ and $E[\mathbf{I}_i^B \,|\, A_i]$ denote the mean intervisit times during the A-cycle and the B-cycle respectively. Note that $E[\mathbf{I}_i^A \,|\, A_i] = E[\mathbf{C}_i^A \,|\, A_i] - \beta_i^{(2)}/\beta_i$, and $E[\mathbf{I}_i^B \,|\, A_i] = E[\mathbf{C}_i^B \,|\, A_i] - \beta_i$. See Fig. 2.1.

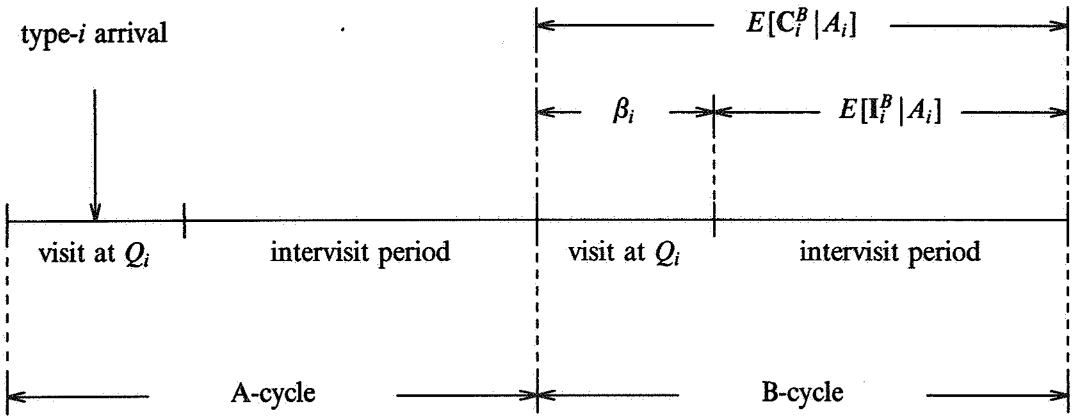

Figure 2.1 *Type-i arrival during visit period at Q_i.*

Furthermore, denote by $E[\mathbf{C}_i^A | \bar{A}_i]$ ($E[\mathbf{C}_i^B | \bar{A}_i]$) the average length of an A-cycle (a B-cycle) given the tagged customer arrived during an intervisit period of the server w.r.t. Q_i in the A-cycle. Let $E[\mathbf{I}_i^A | \bar{A}_i]$ and $E[\mathbf{I}_i^B | \bar{A}_i]$ denote the mean intervisit times during the A-cycle and the B-cycle respectively. Note that $E[\mathbf{I}_i^A | A_i] = E\mathbf{I}_i^2 / E\mathbf{I}_i$, and $E[\mathbf{I}_i^B | A_i] = E[\mathbf{C}_i^B | A_i] - \beta_i$. See Fig. 2.2.

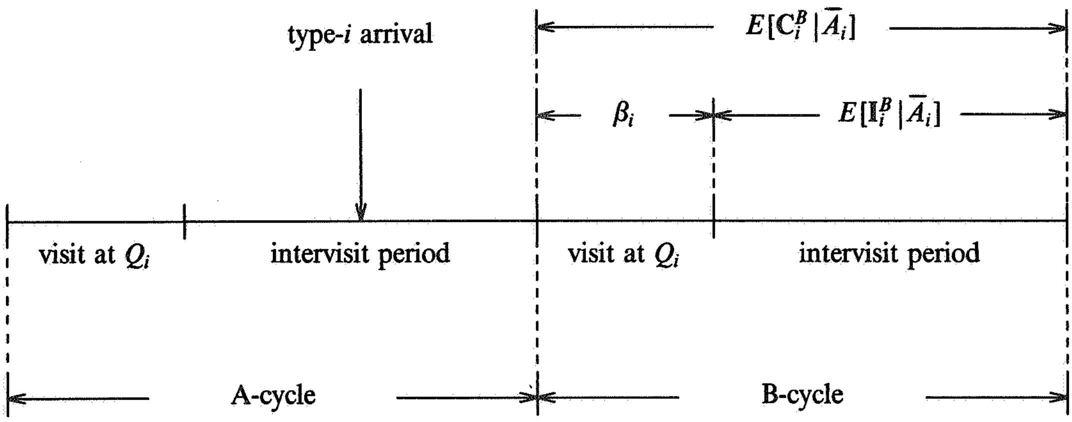

Figure 2.2 *Type-i arrival during intervisit period w.r.t. Q_i.*

As we assume Q_i to be 1-limited, the delay experienced by an arbitrary type-i customer arriving to the system consists of two components: first he has to wait until the server returns to Q_i and subsequently he has to wait as many cycles as there are customers in front of him. Denote the second component by $\mathbf{R}_i$. Then

$$EW_i = E\mathbf{rc}_i + E\mathbf{R}_i. \tag{2.6}$$

In [4] and [10], $E\mathbf{R}_i$ has been approximated by $E\mathbf{X}_i E\mathbf{C}_{b,i}$. Our approximation of $E\mathbf{R}_i$ will take into account whether the tagged customer arrived during a visit period or during an intervisit period for Q_i. This more detailed study of $E\mathbf{R}_i$ leads us to several conditional probabilities and conditional expectations which have to be approximated; but the reward is an approximation for the mean waiting times EW_i which improves upon [10]. The various approximations of conditional probabilities and conditional expectations may be rather inaccurate in some cases; however, even in such cases the conservation-law constraint leads to reasonable approximations for the mean waiting times.

First we condition on the number of customers found upon the arrival of the tagged customer and subsequently on the position of the server at that instant:

$$\begin{aligned} E\mathbf{R}_i &= \sum_{n=0}^{\infty} E[\mathbf{R}_i \mid \mathbf{X}_i = n]p_i(n) \\ &= \sum_{n=0}^{\infty} E[\mathbf{R}_i, A_i \mid \mathbf{X}_i = n]p_i(n) + \sum_{n=0}^{\infty} E[\mathbf{R}_i, \overline{A}_i \mid \mathbf{X}_i = n]p_i(n) \\ &= \sum_{n=2}^{\infty} E[\mathbf{R}_i \mid A_i, \mathbf{X}_i = n]Pr\{A_i \mid \mathbf{X}_i = n\}p_i(n) + \sum_{n=1}^{\infty} E[\mathbf{R}_i \mid \overline{A}_i, \mathbf{X}_i = n]Pr\{\overline{A}_i \mid \mathbf{X}_i = n\}p_i(n). \end{aligned} \tag{2.7}$$

Note that, at an arbitrary moment, the probability that the server is visiting Q_i is equal to ρ_i. Using PASTA [18], we can also write this probability as $Pr\{A_i \mid \mathbf{X}_i \geq 1\}Pr\{\mathbf{X}_i \geq 1\}$. Similarly as in Srinivasan [13], we approximate the probability that an arbitrary type-i customer arrives in a visit period of the server at Q_i, given he finds *n type-i customers* present in the system, by the probability that this arbitrary type-i customer arrives in a visit period of the server at Q_i, given he finds *a non-zero number of customers* present in Q_i:

$$Pr\{A_i \mid \mathbf{X}_i = n\} \approx Pr\{A_i \mid \mathbf{X}_i \geq 1\} = \frac{\rho_i}{Pr\{\mathbf{X}_i \geq 1\}} = \frac{\rho_i}{1 - p_i(0)}, \quad n = 1, 2, \ldots \tag{2.8}$$

We want to take into account the effect that the tagged customer's arrival in the A-cycle has upon the length of the B-cycle. We therefore propose the following approximations:

$$\sum_{n=2}^{\infty} E[\mathbf{R}_i \mid A_i, \mathbf{X}_i = n]p_i(n) \approx \sum_{n=2}^{\infty} (E[\mathbf{C}_i^B \mid A_i] + (n-2)E\mathbf{C}_{b,i})p_i(n), \tag{2.9}$$

$$\sum_{n=1}^{\infty} E[\mathbf{R}_i \mid \overline{A}_i, \mathbf{X}_i = n]p_i(n) \approx \sum_{n=1}^{\infty} (E[\mathbf{C}_i^B \mid \overline{A}_i] + (n-1)E\mathbf{C}_{b,i})p_i(n). \tag{2.10}$$

Note that,

$$\sum_{n=2}^{\infty} (E[\mathbf{C}_i^B \mid A_i] + (n-2)E\mathbf{C}_{b,i})p_i(n) = (E[\mathbf{C}_i^B \mid A_i] - 2E\mathbf{C}_{b,i})(1 - p_i(0)) + E\mathbf{C}_{b,i}E\mathbf{X}_i + (E\mathbf{C}_{b,i} - E[\mathbf{C}_i^B \mid A_i])p_i(1), \tag{2.11}$$

and

$$\sum_{n=1}^{\infty} (E[\mathbf{C}_i^B \mid \overline{A}_i] + (n-1)E\mathbf{C}_{b,i})p_i(n) = (E[\mathbf{C}_i^B \mid \overline{A}_i])(1 - p_i(0)) - E\mathbf{C}_{b,i}(1 - p_i(0)) + E\mathbf{C}_{b,i}E\mathbf{X}_i. \tag{2.12}$$

Combining (2.7) with (2.9), (2.10), (2.11) and (2.12) yields:

$$\begin{aligned} E\mathbf{R}_i \approx{}& \rho_i(E[\mathbf{C}_i^B \mid A_i] - 2E\mathbf{C}_{b,i}) + \frac{\rho_i p_i(1)}{1 - p_i(0)}(E\mathbf{C}_{b,i} - E[\mathbf{C}_i^B \mid A_i]) + E\mathbf{C}_{b,i}E\mathbf{X}_i + \\ & (1 - \rho_i - p_i(0))(E[\mathbf{C}_i^B \mid \overline{A}_i]) - (1 - \rho_i - p_i(0))E\mathbf{C}_{b,i}. \end{aligned} \tag{2.13}$$

Using PASTA [18], applying Little's formula and rearranging terms we obtain:

$$E\mathbf{R}_i \approx \lambda_i E\mathbf{C}_{b,i}E\mathbf{W}_i + (\rho_i - \frac{\rho_i p_i(1)}{1 - p_i(0)})(E[\mathbf{C}_i^B \mid A_i] - E\mathbf{C}_{b,i}) + (1 - \rho_i - p_i(0))(E[\mathbf{C}_i^B \mid \overline{A}_i] - E\mathbf{C}_{b,i}). \tag{2.14}$$

Hence, from (2.6) and (2.14):

$$EW_i \approx \frac{E\mathbf{rc}_i}{1-\lambda_i E\mathbf{C}_{b,i}} + H_i, \tag{2.15}$$

with

$$H_i = \frac{1}{1-\lambda_i E\mathbf{C}_{b,i}} \left[\rho_i \left(1-\frac{p_i(1)}{1-p_i(0)}\right)(E[\mathbf{C}_i^B \mid A_i] - E\mathbf{C}_{b,i}) + (1-\rho_i-p_i(0))(E[\mathbf{C}_i^B \mid \bar{A}_i] - E\mathbf{C}_{b,i}) \right]. \tag{2.16}$$

Note that taking $E[\mathbf{C}_i^B \mid A_i] = E[\mathbf{C}_i^B \mid \bar{A}_i] = E\mathbf{C}_{b,i}$ leads to the approximation discussed in [10]; however, by suitably approximating the terms $E[\mathbf{C}_i^B \mid A_i]$ and $E[\mathbf{C}_i^B \mid \bar{A}_i]$, we can now try to encapsulate the effect of the tagged customer arriving during a visit period or an intervisit period of the server.

(2.16) still contains several unknown terms. The remainder of this section is devoted to a discussion on how to approximate these unknowns.

AN APPROXIMATION FOR $p_i(j)$, $j=0,1$

As in Srinivasan [13], we consider an M/G/1 queue with exceptional service for the customer starting a busy period. Suppose the arrival rate to this system is $\tilde{\lambda}$. Let $\tilde{\beta}_1$ denote the mean service time of the first customer of a busy period, and denote by $\tilde{\beta}_2$ the mean service time of a customer arriving to a non-empty system. It may be easily proven, that the probability that this system is empty at a random point in time is given by

$$\tilde{p}(0) = \frac{1-\tilde{\lambda}\tilde{\beta}_2}{1+\tilde{\lambda}(\tilde{\beta}_1-\tilde{\beta}_2)}. \tag{2.17}$$

Neglecting several dependencies, the cyclic-service system, from the point of view of the type-i customers, behaves the same as the system with exceptional first service as described above with the appropriate substitution of parameters.
We take $\tilde{\lambda}:=\lambda_i$; the 'service time' of a type-i customer arriving to the cyclic-service system and finding no type-i customers present, consists of a residual intervisit period, and a B-cycle:

$$\tilde{\beta}_1 := \frac{E\mathbf{I}_i^2}{2E\mathbf{I}_i} + E[\mathbf{C}_i^B \mid \bar{A}_i]; \tag{2.18}$$

the 'service time' of a type-i customer finding a nonzero number of type-i customers already present in the system is just a cycle with a type-i service:

$$\tilde{\beta}_2 = E\mathbf{C}_{b,i}. \tag{2.19}$$

From (2.17), (2.18) and (2.19) we obtain an approximation for $p_i(0)$:

$$p_i(0) \approx \frac{1-\lambda_i E\mathbf{C}_{b,i}}{1+\lambda_i(E\mathbf{I}_i^2/2E\mathbf{I}_i + E[\mathbf{C}_i^B \mid \bar{A}_i] - E\mathbf{C}_{b,i})}. \tag{2.20}$$

Note that we use PASTA [18] here. The term $E\mathbf{I}_i^2/E\mathbf{I}_i$ in the right-hand side of (2.20) is yet to be determined. We have:

$$E\mathbf{rc}_i = \rho_i\left(\frac{\beta_i^{(2)}}{2\beta_i} + E[\mathbf{I}_i^A \mid A_i]\right) + (1-\rho_i)\frac{E\mathbf{I}_i^2}{E\mathbf{I}_i}. \tag{2.21}$$

From (2.21) and the definition of $E[\mathbf{C}_i^A \mid A_i]$:

$$\frac{E\mathbf{I}_i^2}{E\mathbf{I}_i} = \frac{1}{1-\rho_i}\left(E\mathbf{rc}_i - \rho_i\left(\frac{\beta_i^{(2)}}{2\beta_i} + E[\mathbf{I}_i^A \mid A_i]\right)\right)$$

$$= \frac{\lambda_i\beta_i^{(2)}}{2(1-\rho_i)} + \frac{1}{1-\rho_i}(E\mathbf{rc}_i - \rho_i E[\mathbf{C}_i^A \mid A_i]). \tag{2.22}$$

For $p_i(1)$, the probability that an arriving customer finds exactly 1 type-i customer in the system, we make the heuristic assumption that the $p_i(j)$ for $j=0,1$ follow a distribution as the queue length in an M/G/1 queue with the corresponding arrival and service processes; hence we take $p_i(1) = p_i(0)(1-\beta_i(\lambda_i))/\beta_i(\lambda_i)$. When the service times of type-i customers are exponential, this simplifies to $p_i(1) = p_i(0)(1-p_i(0))$. We have tried more involved approximations for $p_i(1)$, but the influence of this term appears to be rather small.

APPROXIMATIONS FOR THE MEAN CONDITIONAL CYCLE TIMES

In this subsection we give an approximation for $EC_{b,i}$ (occurring in (2.15) and (2.16)), $E[C_i^A\,|A_i]$ (occurring in (2.22)), $E[C_i^B\,|A_i]$ and $E[C_i^B\,|\bar{A}_i]$ (both occurring in (2.16)).

i) *An approximation for $EC_{b,i}$*

The quantity $EC_{b,i}$ was introduced in a paper by Kühn [11]; we closely follow his approach.

Let us call a cycle of the server which starts with a service at Q_i and ends when the server returns to Q_i a 'busy i-cycle'. So $EC_{b,i}$ is the mean length of a busy i-cycle. Assuming balance of flow within a busy i-cycle, the expected number of type-j customers *leaving* the system within this cycle, is equal to the expected number of type-j customers *arriving* at the system within this cycle: $\lambda_j EC_{b,i}$; however this assumption clearly is an approximation.

It is obvious that, for Q_j 1-limited, the mean number of customers leaving the system within the busy i-cycle can not exceed 1; Kühn suggests to suitably limit $\lambda_j EC_{b,i}$ by 1 in that case.

Observing that a busy i-cycle consists of a type-i service and, possibly, services of customers of other types, plus the total switch-over time during that cycle, we propose the following implicit equation for $EC_{b,i}$:

$$EC_{b,i} = \beta_i + \sum_{j\in e,g}\lambda_j EC_{b,i}\beta_j + \sum_{\substack{j\in 1l\\ j\neq i}}\min(1,\lambda_j EC_{b,i})\beta_j + s, \tag{2.23}$$

where e denotes the group of exhaustive queues, g the group of gated queues, and $1l$ the group of 1-limited queues. If $\lambda_j EC_{b,i}\leqslant 1$ for all $j\in 1l$, $j\neq i$, (2.23) simplifies to

$$EC_{b,i} = \frac{\beta_i+s}{1-\rho+\rho_i}, \tag{2.24}$$

otherwise (2.24) is an upper bound for $EC_{b,i}$. In order to compute $EC_{b,i}$ from (2.21) we use an iteration scheme. For $n=1,2,...$:

$$x^{(n)} = \beta_i + s + \sum_{j\in e,g}\rho_j x^{(n-1)} + \sum_{\substack{j\in 1l\\ j\neq i}}\min(1,\lambda_j x^{(n-1)})\beta_j. \tag{2.25}$$

For all starting values $x^{(0)}>0$ and for all $n=1,2,...$, it is easily proven by induction that

$$\left|x^{(n+1)}-x^{(n)}\right| < \left|x^{(n)}-x^{(n-1)}\right|; \tag{2.26}$$

hence, according to the fixed point theorem, the recursion (2.25) has a unique fixed point x^*, which we choose as our approximation for $EC_{b,i}$. Numerical experience suggest that

$$x^{(0)} := \frac{\beta_i+s}{1-\rho+\rho_i} \tag{2.27}$$

is a good starting point for the iteration.

ii) *An approximation for $E[C_i^A\,|A_i]$*

To approximate $E[C_i^A\,|A_i]$, we apply the same ideas as in the approximation for $EC_{b,i}$. Observe that $E[C_i^A\,|A_i]$ consists of a mean type-i service time *given an arrival during that service*, and, possibly, services of customers of other types, plus the total switch-over time in that cycle. So we propose the following implicit equation:

$$E[C_i^A\,|A_i] = \frac{\beta_i^{(2)}}{\beta_i} + s + \sum_{j\in e,g} \lambda_j E[C_i^A\,|A_i]\beta_j + \sum_{\substack{j\in 1l\\ j\neq i}} \min(1,\lambda_j E[C_i^A\,|A_i])\beta_j. \tag{2.28}$$

If $\lambda_j E[C_i^A\,|A_i] \leqslant 1$ for all $j \in 1l$, $j \neq i$, (2.28) simplifies to

$$E[C_i^A\,|A_i] = \frac{\beta_i^{(2)}/\beta_i + s}{1-\rho+\rho_i}. \tag{2.29}$$

In order to compute $E[C_i^A\,|A_i]$ from (2.28) we again use an iteration scheme. For $n = 1,2,...,$

$$x^{(n)} = \frac{\beta_i^{(2)}}{\beta_i} + s + \sum_{j\in e,g} \rho_j x^{(n-1)} + \sum_{\substack{j\in 1l\\ j\neq i}} \min(1,\lambda_j x^{(n-1)})\beta_j. \tag{2.30}$$

As before, it is easily proven that (2.30) has a unique fixed point x^*, which will serve as our approximation for $E[C_i^A\,|A_i]$. Analogously to (2.27),

$$x^{(0)} := \frac{\beta_i^{(2)}/\beta_i + s}{1-\rho+\rho_i}, \tag{2.31}$$

appears to be a good starting point for the iteration.

iii) *An approximation for* $E[C_i^B\,|A_i]$
$E[C_i^B\,|A_i]$ consists of a mean type-i service time and the sum of the mean switch-over times during the cycle. Furthermore, it consists of possible visit times at the other queues. Assuming a strong positive correlation between the A-cycle and the B-cycle, we propose the following approximation for $E[C_i^B\,|A_i]$:

$$E[C_i^B\,|A_i] = \beta_i + s + \sum_{j\in e,g} \rho_j (E[C_i^A\,|A_i] - \frac{\beta_i^{(2)}}{\beta_i} + \beta_i) +$$
$$\sum_{\substack{j\in 1l\\ j\neq i}} \min(1,\lambda_j (E[C_i^A\,|A_i] - \frac{\beta_i^{(2)}}{\beta_i} + \beta_i))\beta_j. \tag{2.32}$$

Note that, when Q_j is served exhaustively or gated, we approximate the mean visit time of the server at Q_j by $\rho_j(E[C_i^A\,|A_i] - \beta_i^{(2)}/\beta_i + \beta_i)$ instead of by $\rho_j s/(1-\rho)$; the latter is the (exact) mean visit time of a cycle we have no information about (cf. (1.5)). Similarly, when Q_j is served 1-limited, we approximate the probability that the server finds Q_j non-empty by $\min(1,\lambda_j(E[C_i^A\,|A_i] - \beta_i^{(2)}/\beta_i + \beta_i))$ instead of by $\lambda_j s/(1-\rho)$; the latter is the exact probability that the server finds Q_j non-empty when there is no information at all about the cycle. In this way we can take advantage of the fact that we know the tagged customer arrived during a visit period of the server at Q_i.

iv) *An approximation for* $E[C_i^B\,|\bar{A}_i]$
Similarly as in iii), $E[C_i^B\,|\bar{A}_i]$ consists of a mean type-i service time and the sum of the mean switch-over times, plus the sum of the mean visit times at the other queues. Again assuming a strong positive correlation between the A-cycle and the B-cycle, we propose the following approximation for $E[C_i^B\,|\bar{A}_i]$:

$$E[C_i^B\,|\bar{A}_i] = \beta_i + s + \sum_{j\in e,g} \rho_j (\frac{E\mathrm{I}_i^2}{E\mathrm{I}_i} + \beta_i) + \sum_{\substack{j\in 1l\\ j\neq i}} \min(1,\lambda_j(\frac{E\mathrm{I}_i^2}{E\mathrm{I}_i} + \beta_i))\beta_j. \tag{2.33}$$

The same remarks as above apply here as well: when Q_j is served exhaustively or gated, we approximate the mean visit time of the server at Q_j by $\rho_j(E\mathrm{I}_i^2/E\mathrm{I}_i + \beta_i)$. Similarly, when Q_j is served 1-limited, we approximate the probability that the server finds Q_j non-empty by $\min(1, \lambda_j(EI_i^2 / EI_i + \beta_i))$. So again we take advantage of the fact that we know in which part of the A-cycle the tagged customer has arrived.

3. Numerical results and conclusions

In this section we will highlight the features of the approximation algorithm and present some numerical results. To our best knowledge, the approximation algorithms described in [10] and in this paper are the only ones available for cyclic-service systems with mixed service strategies. Therefore, in most cases we can only compare the results of these approximations with simulation.

Numerical results

In this subsection we present some results of the approximation. The numerical results have been collected in nine tables at the end of the paper. We have chosen only cases in which the system contains queues with 1-limited service, since for other cases the approximation is identical to the one described in [10]. Representative examples have been chosen to investigate the accuracy of the approximation for a wide range of parameter values. The simulation results in the tables are generated using the simulation language Simula'67. The relative error given in the tables is defined as

$$\frac{\textit{approximation result} - \textit{simulation result}}{\textit{simulation result}} 100\%.$$

In the tables, an asterisk indicates that the mean waiting times have been averaged over the corresponding group of queues. In all cases considered, the service-time distributions are taken negative exponential.

In Tables 1 and 2, four 3-queue systems are studied. In the first system Q_1 is served exhaustively and in the second system Q_1 is served gated; Q_2 and Q_3 are served 1-limited. In the third and fourth system Q_1 and Q_3 are served 1-limited; Q_2 is served exhaustively and gated respectively. In both tables, the systems have been studied under loads of $\rho=0.3$, $\rho=0.5$ and $\rho=0.8$; the individual switch-over times are constant and equal to 0.1. The current approximation (indicated by approx.) is compared with simulation results and with the approximation described in [10] (indicated by $\text{app}_{[10]}$).

In Tables 3a and 3b a 3-queue system is studied, with all queues receiving 1-limited service. In both tables, the system has been studied under loads of $\rho=0.3$, $\rho=0.5$ and $\rho=0.8$. The switch-over time distributions are taken negative exponential; in Table 3a, the mean switch-over times are equal to 0.1, in Table 3b they are equal to 0.05. The approximation is compared with simulation results, with the approximation in Boxma & Meister [4] ($\text{app}_{[4]}$), with the still often used approximation by Kühn [11] ($\text{app}_{[11]}$) and with the approximation of Srinivasan [13] ($\text{app}_{[13]}$). The errors are indicated in parentheses.

In Tables 4, 5 and 6 two 12-queue systems have been studied. In Tables 4a, 5a and 6a Q_1, Q_2 and Q_3 are served exhaustively, whereas in Tables 4b, 5b and 6b these queues are served gated; $Q_4, \ldots, Q_{12}$ are served 1-limited. In Tables 4a and 4b the systems are studied under a load of $\rho=0.3$, in Tables 5a and 5b under a load of $\rho=0.5$ and in Tables 6a and 6b under a load of $\rho=0.8$. The individual switch-over times are constant and equal to 0.16. In each table, five cases are distinguished: Case A, B, C, D and E.
In Case A, all traffic characteristics are symmetric: $\lambda_1 = \cdots = \lambda_{12} = 1/12$; $\beta_1 = \cdots = \beta_{12}$.
In Case B, all mean service times are equal: $\beta_1 = \cdots = \beta_{12}$. The arrival intensities are given by $\lambda_1 = \cdots = \lambda_3$, $\lambda_4 = \cdots = \lambda_{12}$ and $3\lambda_1 = 4/5$, $9\lambda_4 = 1/5$. Hence $\lambda_1 = 12\lambda_4$.
In Case C again $\beta_1 = \cdots = \beta_{12}$ and $\lambda_1 = \cdots = \lambda_3$, $\lambda_4 = \cdots = \lambda_{12}$; but now $3\lambda_1 = 1/5$, $9\lambda_4 = 4/5$. Hence $\lambda_1 = 0.75\lambda_4$.
In Case D, all arrival intensities are equal: $\lambda_1 = \cdots = \lambda_{12} = 1/12$. The mean service times are given by $\beta_1 = \cdots = \beta_3$, $\beta_4 = \cdots = \beta_{12}$ and $\beta_1 = 12\beta_4$.

In Case E again $\lambda_1 = \cdots = \lambda_{12} = 1/12$ and $\beta_1 = \cdots = \beta_3$, $\beta_4 = \cdots = \beta_{12}$; but now $\beta_1 = 0.75\beta_4$. The approximation results are compared with simulation and with the approximation in [10] ($app_{[10]}$). Note that, with respect to the mean waiting times at the 1-limited queues, it makes relatively little difference whether the service strategy at the first three queues is exhaustive or gated.

Finally, in Tables 7, 8 and 9, a 12-queue system is studied, with $Q_1, \ldots, Q_3$ served exhaustively, $Q_4, \ldots, Q_7$ served gated, and $Q_8, \ldots, Q_{12}$ served 1-limited. The total load is kept fixed at $\rho = 0.5$; the individual switch-over times are constant and equal to 0.05. In Case A we have asymmetric arrival streams, while $\beta_1 = \cdots = \beta_{12}$. In Case B one queue has a larger mean service time than the other queues; here $\lambda_1 = \cdots = \lambda_{12} = 1/12$.
In Table 7 there is one heavily loaded exhaustively served queue. In Case A, $\lambda_1 = 0.56$, $\lambda_2 = \cdots = \lambda_{12} = 0.04$; in Case B, $\beta_1 = 3.36$, $\beta_2 = \cdots = \beta_{12} = 0.24$.
In Table 8 we study the case of one heavily loaded gated queue. In Case A, $\lambda_4 = 0.56$, $\lambda_1 = \cdots = \lambda_3 = \lambda_5 = \cdots = \lambda_{12} = 0.04$; in Case B, $\beta_4 = 3.36$, $\beta_1 = \cdots = \beta_3 = \beta_5 = \cdots = \beta_{12} = 0.24$.
In Table 9 there is one heavily loaded queue receiving 1-limited service. In Case A, $\lambda_8 = 0.56$, $\lambda_1 = \cdots = \lambda_7 = \lambda_9 = \cdots = \lambda_{12} = 0.04$; in Case B, $\beta_8 = 3.36$, $\beta_1 = \cdots = \beta_7 = \beta_9 = \cdots = \beta_{12} = 0.24$.
The approximation results are compared with simulation and with the approximation in [10] ($app_{[10]}$).

CONCLUSIONS

Generally speaking, the approximation algorithm has the following properties:

- It is exact in the completely symmetric case with the same traffic characteristics, switch-over time distributions and service strategies at all queues (because the approximation is based on the pseudoconservation law);
- Time and memory requirements on a (personal) computer are negligible;
- For systems with a mixture of only gated and exhaustive service at all queues the approximation is identical to the approximation described in [10], and thus very accurate over the whole range of admissible parameter values;
- When the system contains queues receiving 1-limited service, the approximation is more accurate than the one described in [10]. This is most noticeable at high loads. However, the approximation in [10] provides an explicit formula for EW_i, which gives much qualitative insight into the behavior of the systems under consideration, whereas in the current approximation EW_i must be calculated iteratively. So the increase in the accuracy goes at the expense of a decrease in transparency.
- As may be seen from Tables 7,8 and 9, the position of a queue in the network has little influence on the mean waiting time at that queue. This supports the assumption that the second moment of the cycle time for Q_i is independent of i.

Furthermore, we can make the following remarks.

REMARK 3.1

Application of the approximation algorithm is not limited to cyclic-service systems *with* switch-over times. By giving the mean total switch-over time an arbitrary small, but fixed, value, we can just as well obtain approximations for cyclic-service systems *without* switch-over times. The accuracy of the approximation results does not seem to deteriorate as the switch-over times tend to zero.

REMARK 3.2

As in Boxma & Meister [4] we can apply a modification procedure for the case that there is a heavily loaded 1-limited queue in the system. This procedure is based on the following idea. Remove the 1-limited queue(s) with a relatively high arrival rate from the system, and enlarge the switch-over times to compensate for the service times at the removed queues. The resulting, reduced system has a lower and more symmetric traffic load, and can be more accurately approximated. By substituting the mean waiting times for the reduced system in the pseudoconservation law for the original system, we finally achieve a more accurate mean waiting-time approximation for the removed queue, and also make the resulting mean waiting times satisfy the pseudoconservation law. However, this method leads to two potential problems. First, the method only works when the arrival rates are highly asymmetric and the total switch-over time is sufficiently large. Second, and most important: there is no clear criterion as to when this method will improve the estimates for the mean waiting times. In the tables, this modification procedure is only applied once, viz. in Table 3b in the Boxma & Meister approximation for $\rho=0.8$.

ACKNOWLEDGEMENT
The author is indebted to Prof. O.J. Boxma and Prof. J.W. Cohen for several stimulating discussions.

REFERENCES

[1] BOXMA, O.J., GROENENDIJK, W.P. (1987). Pseudo-conservation laws in cyclic-service systems. *J. Appl. Prob.* **24**, 949-964.

[2] BOXMA, O.J., GROENENDIJK, W.P. (1988). Waiting times in discrete-time cyclic-service systems. *IEEE Trans. Commun.*, Vol. COM-**36**, 164-170.

[3] BOXMA, O.J., GROENENDIJK, W.P. (1988). Two queues with alternating service and switching times. In: *Queueing Theory and its Applications - Liber Amicorum for J.W. Cohen,* eds. O.J. Boxma and R. Syski, North-Holland Publ. Cy., Amsterdam, 261-282.

[4] BOXMA, O.J., MEISTER, B. (1986). Waiting-time approximations for cyclic-service systems with switch-over times. *Performance Evaluation Review* **14**, 254-262.

[5] BUX, W., TRUONG, H.L. (1983). Mean-delay approximations for cyclic-service queueing systems. *Performance Evaluation* **3**, 187-196.

[6] EVERITT, D.E. (1986). Simple approximations for token rings. *IEEE Trans. Commun.,* Vol. COM-**34**, 719-721.

[7] FERGUSON, M.J., AMINETZAH, Y.J. (1985). Exact results for nonsymmetric token ring systems. *IEEE Trans. Commun.,* Vol. COM-**33**, 223-231.

[8] FRANKEN, P., KÖNIG, D., ARNDT, U., SCHMIDT, V. (1982). *Queues and Point Processes.* Wiley, New York.

[9] FUHRMANN, S.W., WANG, Y.T. (1988). Mean waiting time approximations of cyclic service systems with limited service. In: *Performance '87,* eds. P.-J. Courtois and G. Latouche, North-Holland Publ. Cy., Amsterdam, 253-265.

[10] GROENENDIJK, W.P. (1988). Waiting-time approximations for cyclic-service systems with mixed service strategies. In: *Proc. 12th ITC,* North-Holland Publ. Co., Amsterdam.

[11] KÜHN, P.J. (1979). Multiqueue systems with nonexhaustive cyclic service, *The Bell System Techn. Journ.* **58**, 671-698.

[12] PANG, J.W.M., DONALDSON, R.W. (1986). Approximate delay analysis and results for asymmetric token-passing and polling networks. *IEEE J. Sel. Areas in Commun.,* Vol. SAC-**4**, 783-793.

[13] SRINIVASAN, M.M. (1986). An approximation for mean waiting times in cyclic server systems with non-exhaustive service. *Report Department of Industrial and Operations Engineering, University of Michigan.*

[14] SZPANKOWSKI, W., REGO, V. (1987). Ultimate stability conditions for some multidimensional distributed systems. *Report Department of Computer Science, Purdue University.*

[15] TAKAGI, H. (1986). *Analysis of Polling Systems.* The MIT Press, Cambridge, Massachusetts.

[16] TAKAGI, H. (1987). A survey of queueing analysis of polling models. In: *Proceedings of the Third International Conference on Data Communication Systems and their Performance,* Rio de Janeiro.

[17] WATSON, K.S. (1984). Performance evaluation of cyclic service strategies - a survey. In: *Performance '84,* ed. E. Gelenbe, North-Holland Publ. Cy., Amsterdam, 521-533.

[18] WOLFF, R.W. (1982). Poisson arrivals see time averages. *Oper. Res.* **30**, 223-231.

Table 1. $\lambda_1=\lambda_2=\lambda_3=\frac{1}{3}$; $\frac{1}{3}\beta_1=\beta_2=\beta_3$; $s_1=s_2=s_3=0.1$.

3 queues		$(e, 1l, 1l)$		$(g, 1l, 1l)$		$(1l, e, 1l)$			$(1l, g, 1l)$		
		Q_1	$Q_2-Q_3^*$	Q_1	$Q_2-Q_3^*$	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
$\rho=0.3$	simul.	0.321	0.507	0.420	0.468	0.515	0.323	0.452	0.516	0.357	0.453
	approx.	0.317	0.512	0.420	0.467	0.512	0.328	0.457	0.508	0.367	0.453
	$app_{[10]}$	0.324	0.500	0.426	0.457	0.515	0.330	0.445	0.511	0.369	0.441
	error%	-1.1	1.0	-0.0	-0.2	-0.6	1.5	1.1	-1.6	2.8	-0.0
	$err_{[10]}$%	1.0	-1.1	1.4	-2.1	0.1	2.2	-1.6	-0.9	3.4	-2.6
$\rho=0.5$	simul.	0.68	1.59	0.97	1.39	1.44	0.63	1.12	1.43	0.73	1.20
	approx.	0.64	1.66	1.00	1.34	1.44	0.64	1.22	1.42	0.77	1.20
	$app_{[10]}$	0.71	1.53	1.06	1.22	1.47	0.66	1.10	1.45	0.79	1.08
	error%	-5.4	4.6	2.9	-3.9	-0.3	2.3	8.9	-0.7	5.5	-0.1
	$err_{[10]}$%	4.6	-3.7	8.9	-11.6	1.9	5.5	-1.8	1.2	9.6	-9.4
$\rho=0.8$	simul.	1.99	18.45	3.06	17.02	12.99	1.49	10.33	12.53	1.85	10.04
	approx.	1.72	18.88	3.90	14.48	13.41	1.57	8.54	13.24	2.13	8.41
	$app_{[10]}$	2.42	16.77	4.82	11.73	13.74	1.70	7.28	13.53	2.31	7.16
	error%	-13.6	2.0	27.5	-14.6	3.2	5.4	-17.3	5.7	15.1	-16.2
	$err_{[10]}$%	21.5	-9.1	57.6	-31.1	5.8	14.0	-29.6	8.0	24.7	-28.6

Table 2. $\lambda_1=0.6$, $\lambda_2=\lambda_3=0.2$; $\beta_1=\beta_2=\beta_3$; $s_1=s_2=s_3=0.1$.

3 queues		$(e, 1l, 1l)$		$(g, 1l, 1l)$		$(1l, e, 1l)$			$(1l, g, 1l)$		
		Q_1	$Q_2-Q_3^*$	Q_1	$Q_2-Q_3^*$	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
$\rho=0.3$	simul.	0.286	0.419	0.381	0.393	0.535	0.294	0.373	0.532	0.328	0.370
	approx.	0.288	0.420	0.382	0.388	0.535	0.295	0.375	0.532	0.330	0.373
	$app_{[10]}$	0.288	0.416	0.383	0.386	0.534	0.297	0.375	0.531	0.332	0.372
	error%	0.6	0.2	0.2	-1.2	-0.0	0.2	0.5	-0.1	0.7	0.7
	$err_{[10]}$%	0.6	-0.7	0.5	-1.8	-0.1	0.8	0.4	-0.2	1.3	0.6
$\rho=0.5$	simul.	0.58	1.17	0.86	1.00	1.57	0.54	0.81	1.56	0.63	0.79
	approx.	0.56	1.19	0.88	0.97	1.56	0.54	0.84	1.54	0.65	0.83
	$app_{[10]}$	0.59	1.15	0.89	0.94	1.55	0.56	0.84	1.53	0.67	0.83
	error%	-2.7	1.9	1.9	-3.3	-0.9	0.1	4.9	-1.6	3.4	4.6
	$err_{[10]}$%	2.3	-1.8	4.6	-6.0	-1.5	4.2	4.9	-2.4	6.9	4.8
$\rho=0.8$	simul.	1.64	9.96	2.99	8.78	57.83	1.18	2.56	60.38	1.47	2.48
	approx.	1.50	10.29	3.42	7.73	58.56	1.25	2.87	58.00	1.70	2.82
	$app_{[10]}$	1.91	9.42	3.88	6.74	57.49	1.42	3.09	56.68	1.93	3.04
	error%	-8.5	3.3	14.4	-12.0	1.3	5.9	12.1	-3.9	15.6	13.7
	$err_{[10]}$%	15.9	-5.4	29.9	-23.3	-0.6	20.3	20.7	-6.1	31.3	22.6

Table 3a. Three queues, all served 1-limited; $\lambda_1=\lambda_2=\lambda_3=\frac{1}{3}$; $\frac{1}{3}\beta_1=\beta_2=\beta_3$; $s_1=s_2=s_3=0.1$.

	$\rho=0.3$				$\rho=0.5$				$\rho=0.8$			
	Q_1		$Q_2-Q_3^*$		Q_1		$Q_2-Q_3^*$		Q_1		$Q_2-Q_3^*$	
simul.	0.564		0.503		1.44		1.21		11.59		9.19	
approx.	0.563	(-0.2)	0.504	(0.2)	1.43	(-0.3)	1.22	(0.8)	12.44	(7.3)	7.77	(-15.4)
$app_{[4]}$	0.570	(1.1)	0.493	(-2.0)	1.49	(4.0)	1.12	(-7.0)	13.02	(12.3)	6.89	(-25.1)
$app_{[11]}$	0.545	(-3.4)	0.459	(-8.7)	1.36	(-5.6)	0.96	(-20.7)	10.71	(-7.6)	5.34	(-41.9)
$app_{[13]}$	0.570	(1.1)	0.493	(-2.0)	1.47	(2.4)	1.16	(-4.0)	12.59	(8.6)	7.55	(-17.8)

Table 3b. Three queues, all served 1-limited; $\lambda_1=0.6$, $\lambda_2=\lambda_3=0.2$; $\beta_1=\beta_2=\beta_3$; $s_1=s_2=s_3=0.05$.

	$\rho=0.3$				$\rho=0.5$				$\rho=0.8$			
	Q_1		$Q_2-Q_3^*$		Q_1		$Q_2-Q_3^*$		Q_1		$Q_2-Q_3^*$	
simul.	0.334		0.262		0.98		0.60		9.34		1.90	
approx.	0.334	(0.0)	0.262	(0.0)	0.96	(-2.0)	0.63	(4.2)	8.98	(-3.9)	2.26	(19.1)
$app_{[4]}$	0.334	(0.0)	0.262	(0.0)	0.96	(-2.2)	0.63	(4.6)	9.79	(4.8)	1.48	(-22.2)
$app_{[11]}$	0.313	(-6.3)	0.245	(-6.5)	0.84	(-14.3)	0.55	(-8.3)	6.34	(-32.1)	2.48	(30.5)
$app_{[13]}$	0.334	(0.0)	0.261	(-0.4)	0.97	(-1.1)	0.61	(2.2)	9.16	(-1.9)	2.08	(9.7)

Table 4a. Twelve queues; $Q_1,\ldots,Q_3$ served exhaustively, $Q_4,\ldots,Q_{12}$ served 1-limited.

$\rho=0.3$	Case A*		Case B*		Case C*		Case D*		Case E*	
	e	1*l*	e	1*l*	e	1*l*	e	1*l*	e	1*l*
simul.	1.44	2.00	1.39	1.63	1.45	2.04	1.58	2.30	1.45	2.00
approx.	1.44	2.00	1.39	1.62	1.45	2.04	1.58	2.31	1.45	2.00
$app_{[10]}$	1.45	2.00	1.39	1.62	1.45	2.04	1.59	2.26	1.46	2.00
error%	0.1	-0.0	-0.1	-0.2	-0.0	-0.2	-0.0	0.6	0.0	0.1
$err_{[10]}$%	0.5	-0.1	-0.1	-0.3	0.4	-0.3	0.6	-1.5	0.4	-0.0

Table 4b. Twelve queues; $Q_1,\ldots,Q_3$ served gated, $Q_4,\ldots,Q_{12}$ served 1-limited.

$\rho=0.3$	Case A*		Case B*		Case C*		Case D*		Case E*	
	g	1*l*	g	1*l*	g	1*l*	g	1*l*	g	1*l*
simul.	1.51	2.00	1.61	1.60	1.51	2.05	1.81	2.25	1.45	2.00
approx.	1.52	2.00	1.61	1.61	1.51	2.04	1.81	2.25	1.45	2.00
$app_{[10]}$	1.52	1.99	1.61	1.60	1.51	2.04	1.82	2.21	1.46	2.00
error%	0.3	0.2	-0.0	0.1	-0.2	-0.4	0.3	0.0	0.0	-0.0
$err_{[10]}$%	0.7	0.0	-0.0	-0.0	-0.4	-0.3	0.7	-2.0	0.4	-0.1

Table 5a. Twelve queues; $Q_1, \ldots, Q_3$ served exhaustively, $Q_4, \ldots, Q_{12}$ served 1-limited.

$\rho=0.5$	Case A*		Case B*		Case C*		Case D*		Case E*	
	e	1l	e	1l	e	1l	e	1l	e	1l
simul.	2.18	3.75	2.14	2.80	2.19	3.86	2.82	5.66	2.19	3.74
approx.	2.19	3.75	2.14	2.79	2.20	3.86	2.82	5.64	2.21	3.75
$app_{[10]}$	2.24	3.73	2.14	2.76	2.25	3.84	2.92	5.07	2.25	3.73
error%	0.4	0.0	-0.0	-0.2	0.5	0.1	0.1	-0.2	0.9	0.7
$err_{[10]}$%	2.7	-0.6	0.3	-1.2	2.8	-0.4	3.5	-10.4	3.1	-0.2

Table 5b. Twelve queues; $Q_1, \ldots, Q_3$ served gated, $Q_4, \ldots, Q_{12}$ served 1-limited.

$\rho=0.5$	Case A*		Case B*		Case C*		Case D*		Case E*	
	g	1l	g	1l	g	1l	g	1l	g	1l
simul.	2.36	3.74	2.68	2.68	2.33	3.87	3.39	5.31	2.33	3.76
approx.	2.38	3.74	2.68	2.66	2.34	3.86	3.42	5.13	2.35	3.74
$app_{[10]}$	2.43	3.71	2.68	2.64	2.40	3.84	3.50	4.65	2.40	3.73
error%	0.6	-0.0	-0.0	-0.6	0.6	-0.3	1.0	-3.3	0.9	-0.4
$err_{[10]}$%	2.8	-0.7	0.1	-1.3	2.9	-0.8	3.5	-12.4	3.0	-0.9

Table 6a. Twelve queues; $Q_1, \ldots, Q_3$ served exhaustively, $Q_4, \ldots, Q_{12}$ served 1-limited.

$\rho=0.8$	Case A*		Case B*		Case C*		Case D*		Case E*	
	e	1l	e	1l	e	1l	e	1l	e	1l
simul.	5.2	46.1	6.3	13.4	5.1	59.9	8.6	136.7	5.2	44.4
approx.	5.5	45.3	6.5	12.9	5.4	60.8	9.7	115.1	5.5	44.8
$app_{[10]}$	6.2	44.1	6.7	11.8	6.1	59.7	11.5	79.4	6.1	44.0
error%	5.2	-1.6	2.2	-3.4	5.9	-0.3	13.4	-15.8	5.4	0.8
$err_{[10]}$%	19.0	-4.2	5.7	-11.9	18.9	-0.4	34.3	-41.9	18.1	-1.1

Table 6b. Twelve queues; $Q_1, \ldots, Q_3$ served gated, $Q_4, \ldots, Q_{12}$ served 1-limited.

$\rho=0.8$	Case A*		Case B*		Case C*		Case D*		Case E*	
	g	1l	g	1l	g	1l	g	1l	g	1l
simul.	5.9	46.0	8.73	11.8	5.7	61.6	10.8	135.8	5.8	45.4
approx.	6.2	45.1	8.9	11.0	6.0	60.7	12.8	93.9	6.1	44.7
$app_{[10]}$	7.0	43.8	9.0	10.3	6.8	59.4	14.3	64.1	6.8	43.8
error%	6.3	-2.0	1.8	-6.6	6.2	-1.5	18.7	-30.8	5.9	-1.5
$err_{[10]}$%	19.7	-4.9	3.4	-12.5	19.0	-3.5	32.5	-52.8	18.4	-3.5

Table 7. Twelve queues; Q_1 heavily loaded.

	Case A					Case B				
Q_i	simul.	approx.	$app_{[10]}$	error	$err_{[10]}$	simul.	approx.	$app_{[10]}$	error	$err_{[10]}$
1 (e)	0.84	0.85	0.85	0.6	1.4	1.91	1.98	2.11	3.8	10.5
2 (e)	1.15	1.15	1.16	0.1	0.9	2.65	2.70	2.87	1.7	8.2
3 (e)	1.14	1.15	1.16	0.6	1.4	2.71	2.70	2.87	-0.5	5.9
4 (g)	1.19	1.20	1.21	0.8	1.5	2.83	2.81	2.99	-0.8	5.6
5 (g)	1.20	1.20	1.21	0.2	1.0	2.86	2.81	2.99	-2.0	4.3
6 (g)	1.21	1.20	1.21	-0.8	-0.1	2.95	2.81	2.99	-5.0	1.1
7 (g)	1.20	1.20	1.21	0.2	0.9	3.03	2.81	2.99	-7.4	-1.5
8 (1*l*)	1.32	1.32	1.29	0.6	-1.7	4.02	4.01	3.38	-0.3	-15.9
9 (1*l*)	1.33	1.32	1.29	-0.2	-2.6	4.04	4.01	3.38	-0.6	-16.2
10 (1*l*)	1.32	1.32	1.29	-0.0	-2.4	4.10	4.01	3.38	-2.0	-17.4
11 (1*l*)	1.33	1.32	1.29	-0.5	-2.9	4.08	4.01	3.38	-1.6	-17.0
12 (1*l*)	1.33	1.32	1.29	-0.3	-2.6	4.15	4.01	3.38	-3.2	-18.4

Table 8. Twelve queues; Q_4 heavily loaded.

	Case A					Case B				
Q_i	simul.	approx.	$app_{[10]}$	error	$err_{[10]}$	simul.	approx.	$app_{[10]}$	error	$err_{[10]}$
1 (e)	1.00	1.00	1.01	0.5	0.8	2.09	2.18	2.26	4.1	7.9
2 (e)	1.00	1.00	1.01	0.2	0.6	2.13	2.18	2.26	2.1	5.9
3 (e)	1.00	1.00	1.01	0.0	0.4	2.17	2.18	2.26	0.2	4.0
4 (g)	1.31	1.31	1.32	0.4	0.8	2.81	2.84	2.95	1.1	4.9
5 (g)	1.05	1.04	1.05	-1.0	-0.6	1.99	2.26	2.35	13.8	17.9
6 (g)	1.06	1.04	1.05	-1.3	-0.9	2.02	2.26	2.35	12.1	16.5
7 (g)	1.05	1.04	1.05	-0.3	0.0	2.05	2.26	2.35	10.4	14.7
8 (1*l*)	1.15	1.14	1.12	-1.0	-2.5	3.15	3.11	2.66	-1.3	-15.6
9 (1*l*)	1.15	1.14	1.12	-1.0	-2.5	3.20	3.11	2.66	-2.8	-16.9
10 (1*l*)	1.15	1.14	1.12	-1.0	-2.5	3.23	3.11	2.66	-3.7	-17.7
11 (1*l*)	1.15	1.14	1.12	-1.1	-2.6	3.21	3.11	2.66	-3.1	-17.0
12 (1*l*)	1.15	1.14	1.12	-0.8	-2.3	3.25	3.11	2.66	-4.3	-18.2

Table 9. Twelve queues; Q_8 heavily loaded.

	Case A					Case B				
Q_i	simul.	approx.	$app_{[10]}$	error	$err_{[10]}$	simul.	approx.	$app_{[10]}$	error	$err_{[10]}$
1 (e)	0.85	0.86	0.89	0.9	4.4	1.78	1.95	1.99	9.7	11.9
2 (e)	0.85	0.86	0.89	0.9	4.3	1.81	1.95	1.99	7.8	9.9
3 (e)	0.85	0.86	0.89	0.9	4.4	1.83	1.95	1.99	6.5	8.6
4 (g)	0.89	0.89	0.93	0.7	4.2	1.91	2.03	2.07	6.3	8.4
5 (g)	0.89	0.89	0.93	0.4	3.9	1.94	2.03	2.07	4.8	6.9
6 (g)	0.89	0.89	0.93	0.0	3.5	1.97	2.03	2.07	3.4	5.4
7 (g)	0.89	0.89	0.93	0.1	3.6	2.01	2.03	2.07	1.2	3.2
8 (1*l*)	4.40	4.38	4.31	-0.4	-2.0	3.50	3.43	3.52	-2.0	0.6
9 (1*l*)	0.93	0.96	0.99	2.8	5.9	2.59	2.74	2.35	5.8	-9.2
10 (1*l*)	0.93	0.96	0.99	2.6	5.6	2.60	2.74	2.35	5.3	-9.6
11 (1*l*)	0.94	0.96	0.99	2.3	5.3	2.60	2.74	2.35	5.1	-9.8
12 (1*l*)	0.94	0.96	0.99	2.0	5.0	2.60	2.74	2.35	5.2	-9.7