Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

O.J. Boxma, W.P. Groenendijk, J.A. Weststrate

A pseudoconservation law for service systems with a polling table

Department of Operations Research and System Theory Report OS-R8813 August

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 Pseudoconservation Law for Service Systems with a Polling Table

O.J. Boxma
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands;
Faculty of Economics, Tilburg University
P.O. Box 90153, 5000 LE Tilburg, The Netherlands

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

J.A. Weststrate
Faculty of Economics, Tilburg University
P.O. Box 90153, 5000 LE Tilburg, The Netherlands

This paper is devoted to the analysis of waiting times in polling systems in which the stations are polled according to a general service order table. Such systems may, e.g., be used to represent token-bus local area networks in which the 'routing' of the token is fixed. Stations are given higher priority by being listed more frequently in the table, or by receiving service according to the exhaustive service discipline. The polling system is modelled by a single-server multi-queue system in discrete time. Non-zero switch-over times between the queues are assumed.

An extension of the principle of work conservation to systems with non-zero switch-over times leads to the main result of the paper, an exact expression for a weighted sum of the mean waiting times at the various queues. Via a limiting procedure, the discrete-time results are translated to continuous-time results. Finally, the special case of polling in a star network is discussed and compared with polling in a corresponding network with strictly cyclic service order.

1980 Mathematics Subject Classification: 60K25, 68M20. 69C40
Key Words & Phrases: discrete-time queueing system, polling, (pseudo-)conservation law, mean waiting times.
Note: This report has been submitted for publication elsewhere.

1. Introduction

A system of multiple queues attended to by a single server is commonly referred to as a polling system. Such systems arise naturally in computer-communication networks and in switching systems with distributed control. A very large number of studies has been devoted to the queueing analysis of polling systems; see Takagi [13,14]. The vast majority of these studies considers polling systems in which the server attends to the queues in a fixed strictly cyclic order. Several service strategies at the queues have been considered, ranging from exhaustive service (a queue is served until it is empty, before the server moves on to the next queue) to 1-limited service (the server serves exactly one customer from a non-empty queue). The main performance measure under consideration is the mean waiting time of a customer. In the case of exhaustive service at all queues, the exact mean waiting times at all queues can be determined by solving an intricate system of linear equations. For most other service disciplines, like 1-limited, exact mean waiting times are only known in special cases; see Takagi [13,14] for detailed results and further references.

Recent studies of Ferguson & Aminetzah [6] and Watson [16] have revealed that in some special cyclic-service systems there exists a simple expression for a weighted sum of the mean waiting times. Such *pseudoconservation laws* can be used to obtain or test approximations for the individual mean

Report OS-R8813
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

waiting times. In [3,4] they have been generalized by allowing a mixture of different service strategies at different queues. The proof of the resulting unified pseudoconservation law is based on a stochastic decomposition of the workload in the cyclic-service system.

Although the extensive research on cyclic-service systems has been useful for performance evaluation, it has not yet led to a clear ability to control the systems under consideration and to affect their design. Recent advances of computer and communication technology enable the use of more sophisticated scheduling and service strategies, while the need to control complex networks makes the use of such strategies imperative.

At last, a few studies are emerging which open up possibilities for optimization. Levy [11] studies a cyclic-service system with a binomial-gated service strategy at the queues. This leads to a mathematically tractable model in which the choice of binomial probabilities of numbers of customers served at the queues allows prioritization. Browne & Yechiali [5] present a semi-dynamic polling policy in which, after each visit of the server to a queue, the next queue to be visited is chosen so as to minimize some objective function. Finally, Baker & Rubin [1] derive the mean waiting times in a multi-queue single-server system, in which the server visits the queues according to a general service order table and serves each queue exhaustively. Stations are given higher priority by listing them more often in the polling table.

Apart from offering possibilities for optimization over a large class of service strategies, polling with a general service order table is also frequently encountered in practice. The token bus protocol in Local Area Networks gives rise to non-cyclic polling. The introduction of the Manufacturing Automation Protocol (MAP), which is becoming the standard for communication in automated manufacturing and which implements the token bus protocol, will add to the importance of token bus. As a second example, consider a computer with multi-drop terminals in which the computer, after polling a terminal, transmits its outbound traffic and then polls the next terminal. Such a scheme leads to star polling, in which a central queue is visited after each visit to one of the other queues. Another important example of non-cyclic polling is scan polling, which is named after a seek policy in the moving-arm disk device in computer storage. Here the order of polling is $1,2,...,N-1,N,N,N-1,...,2,1$.

The present study is devoted to polling systems with a general service order table. The model is the same as in Baker & Rubin [1], apart from the fact that we allow various service strategies at the queues. By extending the concept of stochastic decomposition of the workload to polling systems with a service order table, we derive a pseudoconservation law for such systems. These results yield new insight into the behaviour of polling systems, and may be used to obtain approximations for individual mean waiting times.

The rest of the paper is organized in the following way. Section 2 contains a model description, and some preliminary results concerning cycle times and visit times; it also presents a work decomposition result for single-server, multi-queue systems with a polling table. Section 3 is devoted to the derivation of the main result, the pseudoconservation law. Finally, the special case of polling in a star network is discussed in detail in Section 4; the mean workload in the star system is compared with the mean workload in a corresponding network with strictly cyclic service order, and shown to be smaller. The results are summarized in Section 5.

Our model formulation is discrete-time. Such a formulation naturally fits the generally time-synchronized configuration of practical communication networks. Furthermore, we feel that a discrete-time approach to polling systems is often slightly easier than a continuous-time approach, in particular in the increasingly important variants in which there are time restrictions on visits and cycles. Continuous-time results are in our study easily obtained via a limiting procedure.

2. Model description and preliminary results

We consider a discrete-time queueing system with N stations (queues) $Q_1,...,Q_N$, where each station has an infinite buffer capacity to store waiting messages (customers). Each message consists of a number of packets; packets are assumed to be of fixed length. Time is slotted with slot size equal to the transmission time of the data contained in a packet (the service time of a packet). We shall call the time interval $[j,j+1[$ the jth slot.

Message arrival process

Let $\mathbf{x}_n(j)$ denote the number of messages arriving at station n in the jth slot. The $\mathbf{x}_n(j)$, $j=1,2,...$ are assumed to be independent, identically distributed random variables with z transform, first and second moment:

$$A_n(z) := E[z^{\mathbf{x}_n(j)}], \quad \lambda_n := E[\mathbf{x}_n(j)], \quad \lambda_n^{(2)} := E[\mathbf{x}_n^2(j)]. \tag{2.1}$$

Let

$$\lambda := \sum_{n=1}^{N}\lambda_n, \quad \lambda^{(2)} := E[(\sum_{n=1}^{N}\mathbf{x}_n(j))^2]. \tag{2.2}$$

The message arrival process at each station is assumed to be independent of those at other stations.

Service process

Denote by $\mathbf{b}_n$ the number of packets included in a message at station n. The z transform, first and second moment of $\mathbf{b}_n$ are given by:

$$B_n(z) := E[z^{\mathbf{b}_n}], \quad \beta_n := E[\mathbf{b}_n], \quad \beta_n^{(2)} := E[\mathbf{b}_n^2]. \tag{2.3}$$

Further introduce:

$$\beta := \sum_{n=1}^{N}\frac{\lambda_n}{\lambda}\beta_n, \quad \beta^{(2)} := \sum_{n=1}^{N}\frac{\lambda_n}{\lambda}\beta_n^{(2)}. \tag{2.4}$$

Note that $B_n(0) = Pr\{\mathbf{b}_n=0\} = 0$ by definition. The offered traffic at the nth station, ρ_n, is defined as

$$\rho_n := \lambda_n\beta_n, \quad n=1,2,...,N. \tag{2.5}$$

The total offered traffic, ρ, is defined as

$$\rho := \sum_{n=1}^{N}\rho_n. \tag{2.6}$$

Polling strategy

The N stations are served by a single server S. The order in which S visits the stations is specified by a *polling table* $T = \{T(m),\ m=1,...,M\}$. The first entry in the polling table, $T(1)$, is the index of the first station polled in a cycle, $T(2)$ the index of the second, etc. After station $T(M)$ is polled, the next cycle starts with $T(1)$.

Following the approach of Baker & Rubin [1], a unique *pseudostation* will be associated with each entry in the polling table; as a result of this, the M pseudostations are visited in a strictly cyclic order. Denote by PS_m the pseudostation associated with the mth entry in the polling table; its corresponding station has index $T(m)$ (as much as possible, we reserve n as an index for stations and m as an index for pseudostations). For simplicity of notation, all references to station indices and pseudostation indices are implicitly assumed to be modulo N and M respectively. We shall say that 'PS_i is connected with PS_j', if $T(i)=T(j)$, that is, if PS_i and PS_j correspond with the same station.

On a few occasions, we shall have need to specify the exact position of work. Suppose PS_j is the first pseudostation after PS_i that is connected with PS_i. By convention, the work in PS_i is shifted to PS_j immediately before S arrives at PS_j.

Service strategy
For the service strategies at the pseudostations there are various possibilities, which differ in the number of messages which may be transmitted during a visit of server S to that pseudostation. Assume that S visits PS_m. When PS_m (and hence $Q_{T(m)}$) is empty, S immediately begins to switch to PS_{m+1}. Otherwise, depending on the service strategy at PS_m:

1) Exhaustive service (E): messages are transmitted from PS_m until PS_m is empty;
2) Gated service (G): only messages present in PS_m upon arrival of S at PS_m are transmitted;
3) 1-Limited service ($1-L$): exactly one message is transmitted from PS_m.

To carry out an exact analysis, we impose the restriction that stations with a 1-limited service strategy are served only once during a cycle.

In the sequel we will allow mixed service strategies (e.g., exhaustive at PS_1, 1-limited at PS_2 and PS_4, gated at PS_3, etc.). It is assumed that pseudostations corresponding to the same station have the same service strategy, but this assumption is not essential for the analysis.

REMARK 2.1
We have restricted ourselves here to the three main disciplines in polling systems. We could have included other strategies, like the binomial-gated strategy of Levy [11]. The derivation of the pseudoconservation law in Section 3 will clearly expose where the choice of service strategy matters, and how adaptation of the result to another service strategy can be made.

Switching process
A switch-over time is needed to switch from one pseudostation to the next. The switch-over times of the server between the mth and the $(m+1)$th pseudostation (measured in slots) are independent, identically distributed random variables with first moment s_m and second moment $s_m^{(2)}$. The first moment s of the total switch-over time during a cycle of the server is given by

$$s := \sum_{m=1}^{M} s_m, \tag{2.7}$$

its second moment is given by $s^{(2)}$. The message arrival process, the service process and the switching process are assumed to be mutually independent.

For ease of notation we define a 'cyclic sum' as

$$\sum_{m=i}^{j}{}^{\bullet} x_m \stackrel{def}{=} \begin{cases} \sum_{m=i}^{j} x_m, & \text{if } i \leqslant j \\ \sum_{m=i}^{M} x_m + \sum_{m=1}^{j} x_m, & \text{if } i > j \end{cases} \tag{2.8a}$$

and, analogously, a 'cyclic product' as

$$\prod_{m=i}^{j}{}^{\bullet} x_m \stackrel{def}{=} \begin{cases} \prod_{m=i}^{j} x_m, & \text{if } i \leqslant j \\ \prod_{m=i}^{M} x_m \times \prod_{m=1}^{j} x_m, & \text{if } i > j \end{cases} \tag{2.8b}$$

Preliminary results
Below, we state a few results for future reference.

Ergodicity conditions
A necessary condition for ergodicity of the system is $\rho<1$. When the service strategy at each queue is either exhaustive or gated, this condition is also sufficient. However, for each queue Q_n with a 1-limited service strategy an additional ergodicity condition is needed, viz.

$$\rho + \lambda_n s < 1, \tag{2.9}$$

see also [4]. In the sequel it will be assumed that the ergodicity conditions are fulfilled, and that the system is in equilibrium.

Cycle- and visit-time results
For the (strictly) cyclic service system with the M pseudostations we can define the cycle time $\mathbf{C}_m$ for PS_m as the time between two successive arrivals of the server at PS_m. It is easily seen that $E\mathbf{C}_m$ is independent of m, and from a balancing argument it follows that the mean cycle time equals $E\mathbf{C}$ with

$$E\mathbf{C} = \frac{s}{1-\rho}. \tag{2.10}$$

Furthermore we define the visit time $\mathbf{V}_m$ of the server for PS_m as the time between the arrival of the server at PS_m and his subsequent departure from PS_m. The mean visit times play an important role in the next section in the waiting-time analysis. To calculate them, we must first introduce the $M \times M$ (0-1) matrix $H = (h_{ij})$ (this is the transposed of H in Baker & Rubin [1]), where

$$h_{ij} := \min\{1, \prod_{m=i}^{j-1}{}^{\bullet} |T(j)-T(m)|\}. \tag{2.11}$$

Note that, for $i \neq j$, h_{ij} equals 1 iff $PS_i, \ldots, PS_{j-1}$ are not connected with the same pseudostation as PS_j.

Example 2.1
Suppose $N=3$ and $T=[1,2,1,3]$. Then

$$H = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

Example 2.2
In the strictly cyclic case $h_{ij}=1$ for all i,j, $i \neq j$.

To calculate the mean visit times for a pseudostation we distinguish between the cases that PS_m has a 1-limited, exhaustive or gated service strategy. Our analysis follows Baker & Rubin [1].

1) *PS_m has a 1-limited service strategy.* By assumption, PS_m is only visited once per cycle. Hence, balancing the flow of customers at PS_m in and out of the system during a cycle of the server shows that

$$\lambda_{T(m)} E\mathbf{C} = \frac{E\mathbf{V}_m}{\beta_{T(m)}}, \tag{2.12}$$

and hence from (2.10):

$$E\mathbf{V}_m = \frac{\rho_{T(m)} s}{1-\rho}. \tag{2.13}$$

2) *PS_m has an exhaustive strategy.* The fact that stations with an exhaustive strategy may be served more than once during a cycle, complicates the determination of the mean visit times for their corresponding pseudostations. Note however, that we can write EV_m as the mean number of messages found by the server upon his arrival at PS_m, multiplied by the mean length of a busy period started by one message. When the server arrives at PS_m, messages have accumulated during the interval

$$\sum_{i=m+1}^{m-1} {}^{\bullet}\, h_{im}(s_{i-1}+EV_i)+s_{m-1}. \tag{2.14}$$

(2.14) represents the mean time (measured in slots) between the departure of S from the last pseudostation before PS_m which is connected with PS_m, and the arrival of S at PS_m. From (2.14) and the fact that the mean length of a busy period started by one message is $\beta_{T(m)}/(1-\rho_{T(m)})$, we obtain for the mean visit time at PS_m (note that $h_{mm}=0$ by definition):

$$EV_m = \lambda_{T(m)}[\sum_{i=1}^{M} h_{im}(s_{i-1}+EV_i)+s_{m-1}]\frac{\beta_{T(m)}}{1-\rho_{T(m)}}. \tag{2.15}$$

3) *PS_m has a gated service strategy.* In this case, we can write EV_m as the mean number of messages found by the server upon his arrival at PS_m, multiplied by the mean transmission time of a message at PS_m. When the server arrives at PS_m, messages have accumulated during the interval

$$\sum_{i=m+1}^{m-1} {}^{\bullet}\, h_{im}(EV_{i-1}+s_{i-1})+EV_{m-1}+s_{m-1}. \tag{2.16}$$

So for the mean visit times at the gated pseudostations we obtain:

$$EV_m = \lambda_{T(m)}[\sum_{i=1}^{M} h_{im}(EV_{i-1}+s_{i-1})+EV_{m-1}+s_{m-1}]\beta_{T(m)}. \tag{2.17}$$

Balancing the flow of messages at Q_n in and out of the system during a cycle shows that

$$\sum_{\{m\,|\,T(m)=n\}} \frac{EV_m}{\rho_{T(m)}} = EC = \frac{s}{1-\rho}. \tag{2.18}$$

Note that (2.15), (2.17) and (2.18) represent a set of $M-N$ simultaneous linear equations for the mean visit times of the server at the pseudostations with a gated or exhaustive service strategy.

Work decomposition

Suppose for the moment that there are no switch-over times. Then the principle of work conservation implies that the amount of work in the polling system equals the amount of work in a 'corresponding' single-server, single-queue system with batch arrivals, the so-called Geom/G/1 queue [9]. The arrival process in this discrete-time queue is constructed as follows: the arrival streams at all N stations of our original system are aggregated into a single stream. The batch of all the messages arriving in a slot is called a train. In any slot no train arrives with probability $\prod_{n=1}^{N} A_n(0)$ and a train does arrive with probability $1-\prod_{n=1}^{N} A_n(0)$. An arbitrarily chosen message in this train poses a service request whose z transform is the mixture $\sum_{n=1}^{N}(\lambda_n/\lambda)B_n(z)$.

As customary in discrete-time queueing literature an *arbitrary epoch* is supposed to be the instant just after the beginning of an arbitrary slot. Define $V_{Geom/G/1}$ as the amount of work at an arbitrary epoch in this corresponding Geom/G/1 system with batch arrivals. According to Kobayashi & Konheim [10], the mean number of messages in this system at an arbitrary epoch is given by

$$EX_{Geom} = \frac{\lambda^2\beta^{(2)}}{2(1-\rho)} + \frac{(\lambda^{(2)}-\lambda^2-\lambda)\beta}{2(1-\rho)} + \rho. \tag{2.19}$$

Note that the second term in the right-hand side disappears when the number of messages arriving per slot has a Poisson distribution. The mean number of messages in service is ρ; the mean residual service time of the message in service is $\beta^{(2)}/2\beta+1/2$ (cf. Hunter [9]). Hence,

$$EV_{Geom} = \left[\frac{\lambda^2\beta^{(2)}}{2(1-\rho)}+\frac{(\lambda^{(2)}-\lambda^2-\lambda)\beta}{2(1-\rho)}\right]\beta + \rho\left[\frac{\beta^{(2)}}{2\beta}+\frac{1}{2}\right]. \tag{2.20}$$

Since there are switch-over times incorporated in the original system, the server may be idle (switching) although there is work in the system. Hence the principle of work conservation can not be applied. However, for the service system with a polling table, just as for systems with a strictly cyclic polling strategy [3,4], there appears to exist a natural extension of the work conservation principle. The amount of work in the system can be decomposed into the amount of work in the corresponding Geom/G/1 queue and an extra quantity. The decomposition is presented in Theorem 2.1 below. In the theorem, an arbitrary epoch is considered to be 'in' a switching interval if it marks the beginning of a switching slot.

THEOREM 2.1
Consider a single-server, multi-queue service system with a polling table, as described in the beginning of this section. Suppose the system is ergodic and stationary. Then the amount of work V_P in this system at an arbitrary epoch is distributed as the sum of the amount of work $V_{Geom/G/1}$ in the 'corresponding' Geom/G/1 system with batch arrivals at an arbitrary epoch and the amount of work Y in the system with polling table at an arbitrary epoch in a switching interval. In other words,

$$V_P \overset{D}{=} V_{Geom/G/1} + Y, \tag{2.21}$$

where $\overset{D}{=}$ *stands for equality in distribution. Furthermore,* $V_{Geom/G/1}$ *and* Y *are independent.*

PROOF: See Boxma [2].

3. THE PSEUDOCONSERVATION LAW

As a consequence of Theorem 2.1:

$$EV_P = EV_{Geom/G/1} + EY, \tag{3.1}$$

and hence, cf. (2.20):

$$EV_P = \frac{\lambda\beta^{(2)}}{2(1-\rho)}\rho + \frac{(\lambda^{(2)}-\lambda^2-\lambda)\beta}{2\lambda(1-\rho)}\rho + \rho\{\frac{\beta^{(2)}}{2\beta}+\tfrac{1}{2}\} + EY. \tag{3.2}$$

On the other hand, we can write EV_P as:

$$EV_P = \sum_{n=1}^{N}\beta_n EX_n^w + \sum_{n=1}^{N}\rho_n\{\frac{\beta_n^{(2)}}{2\beta_n}+\tfrac{1}{2}\} = \sum_{n=1}^{N}\rho_n EW_n + \sum_{n=1}^{N}\rho_n\{\frac{\beta_n^{(2)}}{2\beta_n}+\tfrac{1}{2}\}, \tag{3.3}$$

where X_n^w denotes the number of waiting type-n messages at an arbitrary epoch, and W_n the waiting time of a type-n message; the waiting time is counted from the beginning of the slot following the one in which the message arrived. The second equality is based on Little's formula. From (3.2) and (3.3) we obtain the following expression for a weighted sum of the mean message waiting times:

$$\sum_{n=1}^{N}\rho_n EW_n = \frac{\lambda\beta^{(2)}}{2(1-\rho)}\rho + \frac{(\lambda^{(2)}-\lambda^2-\lambda)\beta}{2\lambda(1-\rho)}\rho + EY. \tag{3.4}$$

Note that in the derivation of (3.4) the notion of pseudostations has played no role. Only the last term in (3.4) depends on this notion. To obtain an expression for the weighted sum of the mean message waiting times at the various queues, we now derive an expression for EY, the mean amount of work in the modified system of pseudostations at an arbitrary epoch in a switching interval. Let $\mathbf{Y}_m$ denote the amount of work in the cyclic-service system at an arbitrary switching epoch during a switch-over from PS_m to PS_{m+1}. Obviously,

$$E\mathbf{Y} = \sum_{m=1}^{M} \frac{s_m}{s} E\mathbf{Y}_m. \tag{3.5}$$

As in [3,4], $E\mathbf{Y}_m$ is composed of three terms:

1. $E\mathbf{M}_m^{(1)}$: the mean amount of work in PS_m at a departure epoch of the server from PS_m.
2. $E\mathbf{M}_m^{(2)}$: the mean amount of work in the rest of the system at a departure epoch of S from PS_m.
3. $\rho\{\frac{s_m^{(2)}}{2s_m} - \frac{1}{2}\}$: the mean amount of work that arrived in the system during the past part of the switching interval under consideration.

It will turn out that only $E\mathbf{M}_m^{(1)}$ depends on the choice of the service strategies at the various (pseudo)stations. To calculate $E\mathbf{M}_m^{(2)}$, we must introduce the $M \times M$ (0-1) matrix $Z = (z_{ij})$, where

$$z_{ij} := \min\{1, \prod_{k=i+1}^{j}{}^{\bullet} |T(i) - T(k)|\}. \tag{3.6}$$

Note that for $i \neq j$, $z_{ij} = 1$ iff $PS_{i+1}, \ldots, PS_j$ are not connected with PS_i.

EXAMPLE 3.1
Suppose $N=3$ and $T=[1,2,1,3]$. Then

$$Z = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

EXAMPLE 3.2
In the strictly cyclic case, $z_{ij} = 1$ for all i,j, $i \neq j$.

We shall now consider $E\mathbf{M}_m^{(2)}$, the mean amount of work in $PS_{m-1}, \ldots, PS_1, PS_M, \ldots, PS_{m+1}$ at a departure epoch of the server from PS_m. PS_k can make two contributions to $E\mathbf{M}_m^{(2)}$, viz. (i) the mean amount of work $E\mathbf{M}_k^{(1)}$ left behind in PS_k by S, and (ii) the mean amount of work that has arrived in $Q_{T(k)}$ during the switch-over times from PS_k to PS_m and the visit times of $PS_{k+1}, \ldots, PS_m$. Both contributions disappear when any of the pseudostations $PS_{k+1}, \ldots, PS_m$ is connected with PS_k, i.e. when $z_{km} = 0$ (but not when PS_k is connected with, say, PS_{m+1}; cf. the convention introduced in Section 2).

Hence we have:

$$E\mathbf{M}_m^{(2)} = \sum_{k \neq m} z_{km} E\mathbf{M}_k^{(1)} + \sum_{k \neq m} z_{km} \rho_{T(k)} \sum_{j=k}^{m-1}{}^{\bullet} (s_j + E\mathbf{V}_{j+1}). \tag{3.7}$$

Still leaving $E\mathbf{M}_m^{(1)}$ unspecified, we obtain the following expression for $E\mathbf{Y}$ from (3.5) and (3.7) (note that $z_{mm} = 0$ by definition):

$$E\mathbf{Y} = \sum_{m=1}^{M} \frac{s_m}{s} [E\mathbf{M}_m^{(1)} + \sum_{k=1}^{M} z_{km} E\mathbf{M}_k^{(1)}] + \sum_{k=1}^{M} \rho_{T(k)} \sum_{m=1}^{M} \frac{s_m}{s} z_{km} \sum_{j=k}^{m-1}{}^{\bullet} (s_j + E\mathbf{V}_{j+1})$$
$$+ \rho \sum_{m=1}^{M} \frac{s_m^{(2)}}{2s} - \frac{1}{2}\rho. \tag{3.8}$$

Finally, from (3.4) and (3.8):

$$\sum_{n=1}^{N}\rho_n E\mathbf{W}_n = \frac{\lambda\beta^{(2)}}{2(1-\rho)}\rho + \frac{(\lambda^{(2)}-\lambda^2-\lambda)\beta}{2\lambda(1-\rho)}\rho + \sum_{m=1}^{M}\frac{s_m}{s}[E\mathbf{M}_m^{(1)} + \sum_{k=1}^{M}z_{km}E\mathbf{M}_k^{(1)}]$$
$$+ \sum_{k=1}^{M}\rho_{T(k)}\sum_{m=1}^{M}\frac{s_m}{s}z_{km}\sum_{j=k}^{m-1}{}^{\bullet}(s_j+E\mathbf{V}_{j+1}) + \rho\sum_{m=1}^{M}\frac{s_m^{(2)}}{2s} - \frac{1}{2}\rho. \tag{3.9}$$

Note that the form of Formula (3.9) is still independent of the service strategies at the various pseudostations. Only the $E\mathbf{M}_m^{(1)}$ and $E\mathbf{V}_m$ depend on the choice of service strategies.

The $E\mathbf{M}_m^{(1)}$ are readily found for an exhaustive, gated or 1-limited strategy at PS_m:
PS_m has an exhaustive service strategy:

$$E\mathbf{M}_m^{(1)} = 0. \tag{3.10}$$

PS_m has a gated service strategy:

$$E\mathbf{M}_m^{(1)} = \rho_{T(m)}E\mathbf{V}_m, \tag{3.11}$$

with $E\mathbf{V}_m$ as given in (2.17).
PS_m has a 1-limited service strategy: a similar derivation as in [4] leads to

$$E\mathbf{M}_m^{(1)} = \frac{\lambda_{T(m)}s}{1-\rho}\rho_{T(m)}E\mathbf{W}_{T(m)} + \rho_{T(m)}^2\frac{s}{1-\rho} + \frac{\rho_{T(m)}s}{1-\rho}\frac{\lambda_{T(m)}^{(2)}-\lambda_{T(m)}}{2\lambda_{T(m)}}. \tag{3.12}$$

Substitution of (3.10), (3.11) and (3.12) into (3.9) gives our main result, which is formulated in Theorem 3.1 below. Denote by
e: *the group of exhaustive stations,*
g: *the group of gated stations,*
$\tilde{g}$ *the group of gated pseudostations,*
1l: *the group of 1-limited stations.*

THEOREM 3.1
Consider an ergodic and stationary single-server system with a polling table and mixed service strategies as described in Section 2. Then

$$\sum_{n\in e}\rho_n E\mathbf{W}_n + \sum_{n\in g}\rho_n E\mathbf{W}_n + \sum_{n\in 1l}\rho_n[1-\frac{\lambda_n s}{1-\rho}]E\mathbf{W}_n =$$
$$\frac{\lambda\beta^{(2)}}{2(1-\rho)}\rho + \frac{(\lambda^{(2)}-\lambda^2-\lambda)\beta}{2\lambda(1-\rho)}\rho + \rho\sum_{m=1}^{M}\frac{s_m^{(2)}}{2s} - \frac{1}{2}\rho +$$
$$\frac{s}{(1-\rho)}\sum_{n\in 1l}(\rho_n^2 + \rho_n\frac{\lambda_n^{(2)}-\lambda_n}{2\lambda_n}) +$$
$$\sum_{k=1}^{M}\rho_{T(k)}\sum_{m=1}^{M}\frac{s_m}{s}z_{km}\sum_{j=k}^{m-1}{}^{\bullet}(s_j+E\mathbf{V}_{j+1}) + \sum_{j\in\tilde{g}}\rho_{T(j)}E\mathbf{V}_j\sum_{m=1}^{M}\frac{s_m}{s}z_{jm} + \sum_{m\in\tilde{g}}\rho_{T(m)}\frac{s_m}{s}E\mathbf{V}_m. \tag{3.13}$$

Note that, for this complex polling system, the right-hand side of (3.13) can be easily evaluated for each given set of parameter values, polling order and service strategies.

SPECIAL CASES
(i) When each station is polled only once during a cycle, Theorem 3.1 reduces to Theorem 2 of [4].
(ii) If the numbers of message arrivals at the queues per slot are independent Poisson distributed random variables, then $\lambda_n^{(2)}=\lambda_n^2+\lambda_n$ and $\lambda^{(2)}=\lambda^2+\lambda$; this leads to minor simplifications in the right-

hand side of (3.13).
(iii) If messages arrive in batches with γ_n the arrival rate of a batch at Q_n and $G_n(z)$ the generating function of the batch size, and if the numbers of batch arrivals at the queues are independent Poisson distributed random variables, then $\lambda_n^{(2)}-\lambda_n^2-\lambda_n=\gamma_n G_n^{(2)}(1)$ and $\lambda^{(2)}-\lambda^2-\lambda=\sum_{n=1}^N \gamma_n G_n^{(2)}(1)$.

REMARK 3.1
Giannakouros & Laloux [8] also derive a pseudoconservation law for a polling system with general service order table. However, they do not specify the $EM_m^{(1)}$ terms for particular service strategies. In their model, for stations occurring more than once in the table, the work in each of its pseudostations can only arrive during part of a cycle, and is not shifted to other connected pseudostations (cf. the convention introduced in Section 2).

THE CONTINUOUS-TIME CASE
In the present paper we have expressed all quantities involved, including waiting times, in slots with the slot length equal to the time unit. If instead we assume a slot to be of length Δ we are able, by taking the limit $\Delta \to 0$, to pass the results over to continuous time.

We first translate (3.13), using a tilde to indicate that a quantity is expressed in time units instead of slots. Introduce

$$\tilde\lambda_n=\lambda_n/\Delta,\quad \tilde\lambda=\lambda/\Delta,\quad \tilde\beta_n=\beta_n\Delta,\quad \tilde\beta=\beta\Delta,\quad \tilde s_n=s_n\Delta,\quad \tilde s=s\Delta,\quad E\tilde{\mathbf{W}}_n=E\mathbf{W}_n\Delta;$$

furthermore, cf. [4],

$$\tilde\lambda_n^{(2)}=\frac{\lambda_n^{(2)}}{\Delta}+\frac{1}{\Delta}\Big(\frac{1}{\Delta}-1\Big)\lambda_n^2,\quad \tilde\lambda^{(2)}-\tilde\lambda^2=\sum_{n=1}^N(\lambda_n^{(2)}-\lambda_n^2)/\Delta,\quad \tilde\beta_n^{(2)}=\beta_n^{(2)}\Delta^2,\quad \tilde\beta^{(2)}=\beta^{(2)}\Delta^2,\quad \tilde s_n^{(2)}=s_n^{(2)}\Delta^2.$$

Of course $\rho_n=\lambda_n\beta_n=\tilde\lambda_n\tilde\beta_n$. Formula (3.13) now becomes:

$$\sum_{n\in e}\rho_n E\tilde{\mathbf{W}}_n\frac{1}{\Delta}+\sum_{n\in g}\rho_n E\tilde{\mathbf{W}}_n\frac{1}{\Delta}+\sum_{n\in 1l}\rho_n\Big[1-\frac{\tilde\lambda_n\tilde s}{1-\rho}\Big]E\tilde{\mathbf{W}}_n\frac{1}{\Delta}=$$

$$\frac{\tilde\lambda\tilde\beta^{(2)}}{2(1-\rho)}\rho\frac{1}{\Delta}+\frac{(\tilde\lambda^{(2)}-\tilde\lambda^2-\tilde\lambda)\tilde\beta}{2\tilde\lambda(1-\rho)}\rho\frac{1}{\Delta}+\rho\frac{\sum_{m=1}^M\tilde s_m^{(2)}}{2\tilde s}\frac{1}{\Delta}-\frac{1}{2}\rho+$$

$$\frac{\tilde s}{(1-\rho)}\sum_{n\in 1l}\Big(\rho_n^2+\rho_n\frac{\tilde\lambda_n^{(2)}-(1-\Delta)\tilde\lambda_n^2-\tilde\lambda_n}{2\tilde\lambda_n}\Big)\frac{1}{\Delta}+$$

$$\sum_{k=1}^M\rho_{T(k)}\sum_{m=1}^M\frac{\tilde s_m}{\tilde s}z_{km}\sum_{j=k}^{m-1}{}^{\bullet}(\tilde s_j+E\tilde{\mathbf{V}}_{j+1})\frac{1}{\Delta}+\sum_{j\in\tilde g}\rho_{T(j)}E\tilde{\mathbf{V}}_j\sum_{m=1}^M\frac{\tilde s_m}{\tilde s}z_{jm}\frac{1}{\Delta}+$$

$$\sum_{m\in\tilde g}\rho_{T(m)}\frac{\tilde s_m}{\tilde s}E\tilde{\mathbf{V}}_m\frac{1}{\Delta}. \tag{3.14}$$

In (3.14) we can take the limit for $\Delta\to 0$ by multiplying the left- and right-hand side with Δ and substituting $\Delta=0$. If we do so, we obtain

$$\sum_{n\in e}\rho_n E\tilde{\mathbf{W}}_n+\sum_{n\in g}\rho_n E\tilde{\mathbf{W}}_n+\sum_{n\in 1l}\rho_n\Big[1-\frac{\tilde\lambda_n\tilde s}{1-\rho}\Big]E\tilde{\mathbf{W}}_n=$$

$$\frac{\tilde\lambda\tilde\beta^{(2)}}{2(1-\rho)}\rho+\frac{(\tilde\lambda^{(2)}-\tilde\lambda^2-\tilde\lambda)\tilde\beta}{2\tilde\lambda(1-\rho)}\rho+\rho\frac{\sum_{m=1}^M\tilde s_m^{(2)}}{2\tilde s}+$$

$$\frac{\tilde{s}}{(1-\rho)}\sum_{n\in \tilde{1l}}(\rho_n^2 + \rho_n\frac{\tilde{\lambda}_n^{(2)}-\tilde{\lambda}_n^2-\tilde{\lambda}_n}{2\tilde{\lambda}_n}) +$$

$$\sum_{k=1}^{M}\rho_{T(k)}\sum_{m=1}^{M}\frac{\tilde{s}_m}{\tilde{s}}z_{km}\sum_{j=k}^{m-1}{}^{\bullet}(\tilde{s}_j+E\tilde{V}_{j+1}) + \sum_{j\in\tilde{g}}\rho_{T(j)}E\tilde{V}_j\sum_{m=1}^{M}\frac{\tilde{s}_m}{\tilde{s}}z_{jm} + \sum_{m\in\tilde{g}}\rho_{T(m)}\frac{\tilde{s}_m}{\tilde{s}}E\tilde{V}_m. \tag{3.15}$$

Formula (3.15) is the generalization of the continuous-time pseudoconservation law of [3]. In fact, it also extends the results of [3] to the case in which the message arrival process at each queue is allowed to be a Poisson process with batch arrivals.

4. Example: the star network

In this section we evaluate the pseudoconservation law (3.13) for a network with a star configuration, and we make a comparison with the network with corresponding stations and strictly cyclic service order. A polling network with a star configuration represents, e.g., a computer with multidrop terminals in which the computer, after polling a terminal, transmits its outbound traffic and then polls the next terminal. Two cases are considered. In Case A the central station Q_1 receives exhaustive service, whereas in Case B it receives gated service; in both cases $Q_2, \ldots, Q_N$, $N \geq 2$, receive 1-limited service. It will appear that in both cases the mean workload in the star system is smaller than the mean workload in the corresponding cyclic-service system.

The polling table is: $T=[1,2,1,3,...,1,N]$. There are $M=2(N-1)$ pseudostations. Denote by

$\tilde{e}=\{1,3,...,2N-3\}$: *the group of exhaustive pseudostations (Case A),*

$\tilde{g}=\{1,3,...,2N-3\}$: *the group of gated pseudostations (Case B),*

$\tilde{1l}=\{2,4,...,2N-2\}$: *the group of 1-limited pseudostations.*

Definition (3.6) of the matrix Z implies that

when $i\in\tilde{e}$ $(\tilde{g})$: $z_{i,i+1}=1$; $z_{ij}=0$ otherwise;
when $i\in\tilde{1l}$: $z_{ij}\equiv 1$, $i\neq j$.

Case A: Q_1 exhaustive service
Introducing

$$C := \frac{\lambda\beta^{(2)}}{2(1-\rho)}\rho + \frac{(\lambda^{(2)}-\lambda^2-\lambda)\beta}{2\lambda(1-\rho)}\rho - \frac{1}{2}\rho + \frac{s}{1-\rho}\sum_{n\in 1l}(\rho_n^2 + \rho_n\frac{\lambda_n^{(2)}-\lambda_n}{2\lambda_n}), \tag{4.1}$$

and substituting the z_{ij} values calculated above into (3.13) gives the pseudoconservation law for a star network in discrete time ($\mathbf{W}_n^{star}$ denotes the waiting time at Q_n in the star network):

$$\rho_1 E\mathbf{W}_1^{star} + \sum_{n\in 1l}\rho_n[1-\frac{\lambda_n s}{1-\rho}]E\mathbf{W}_n^{star} =$$

$$C + \rho\sum_{m=1}^{M}\frac{s_m^{(2)}}{2s} + \rho_1\sum_{k\in\tilde{e}}\frac{s_{k+1}}{s}(s_k+E\mathbf{V}_{k+1}) + \sum_{k\in\tilde{1l}}\rho_{T(k)}\sum_{m\neq k}\frac{s_m}{s}\sum_{j=k}^{m-1}{}^{\bullet}(s_j+E\mathbf{V}_{j+1}). \tag{4.2}$$

The mean visit times $E\mathbf{V}_k$ can be specified using (2.13) and (2.15):

$$k \in \tilde{1l}: \quad EV_k = \rho_{T(k)} \frac{s}{(1-\rho)}; \tag{4.3}$$

$$k \in \tilde{e}: \quad EV_k = \frac{\rho_1}{1-\rho_1}[s_{k-1}+s_{k-2}+EV_{k-1}] = \frac{\rho_1}{1-\rho_1}[s_{k-1}+s_{k-2}+\rho_{T(k-1)}\frac{s}{1-\rho}]. \tag{4.4}$$

Note that the sum of the visit times in (4.4) satisfies (2.18). To simplify the following calculations, it is assumed that all M switch-over times are equal to the constant r. Substitution of (4.3) and (4.4) into (4.2) yields, after a tedious but straightforward calculation:

$$\rho_1 E\mathbf{W}_1^{star} + \sum_{n \in 1l} \rho_n [1 - \frac{\lambda_n Mr}{1-\rho}] E\mathbf{W}_n^{star} =$$

$$C + \rho_1 r + \frac{1}{2}(\rho-\rho_1)Mr + \rho_1 \frac{\rho-\rho_1}{1-\rho} r + \frac{1}{2}\rho_1 \frac{\rho-\rho_1}{1-\rho_1} Mr +$$

$$\frac{\rho_1}{(1-\rho)(1-\rho_1)}(M-1)r \sum_{n=2}^{N} \rho_n^2 + \frac{\rho_1}{(1-\rho)(1-\rho_1)}(M-2)r \sum_{j=2}^{N-1} \sum_{i=j+1}^{N} \rho_i \rho_j +$$

$$\frac{1}{1-\rho} Mr \sum_{j=2}^{N-1} \sum_{i=j+1}^{N} \rho_i \rho_j. \tag{4.5}$$

We now compare the star network with the 'corresponding' strictly cyclic service system; this is a system with N queues $Q_1, \ldots, Q_N$ with cyclic service in this order, where Q_1 receives exhaustive service and $Q_2, \ldots, Q_N$ receive 1-limited service, and where each queue has exactly the same traffic characteristics as its counterpart in the star network. Furthermore, the total switch-over times in both systems correspond; so in the strictly cyclic system, $s = Mr$ and $s^{(2)} = M^2 r^2$.

A comparison between the mean workloads in both models amounts to a comparison between the expressions in the right-hand sides of the pseudoconservation laws for both models. The pseudoconservation law for the 'corresponding' strictly cyclic service system reads in this case (cf. [4]; $\mathbf{W}_n^{cycl}$ denotes the waiting time at Q_n in the cyclic network):

$$\rho_1 E\mathbf{W}_1^{cycl} + \sum_{n \in 1l} \rho_n [1 - \frac{\lambda_n Mr}{1-\rho}] E\mathbf{W}_n^{cycl} =$$

$$C + \frac{1}{2}\rho Mr + \frac{Mr}{2(1-\rho)}[\rho^2 - \sum_{n=1}^{N} \rho_n^2]. \tag{4.6}$$

Subtract the right-hand side of (4.5) from the right-hand side of (4.6), and call the difference $Diff_E$. Note that, with an obvious notation,

$$Diff_E = E\mathbf{V}^{cycl} - E\mathbf{V}^{star} = E\mathbf{Y}^{cycl} - E\mathbf{Y}^{star}.$$

From (4.5) and (4.6),

$$Diff_E = \rho_1 (N-2) r [1 + \frac{\rho-\rho_1}{1-\rho_1}] + \frac{\rho_1}{(1-\rho)(1-\rho_1)}(2N-2) r \sum_{j=2}^{N-1} \sum_{i=j+1}^{N} \rho_i \rho_j] \geq 0. \tag{4.7}$$

This result leads to the following observations:

- The mean workload in the star system is at most equal to the mean workload in the corresponding cyclic system, and the difference increases roughly linearly in N.
- $Diff_E = 0$ when $N = 2$, and when $\rho_1 = 0$; indeed, in those cases the two systems coincide.
- $Diff_E$ approaches zero when $r \rightarrow 0$.
- $Diff_E$ depends on λ_n and β_n only via their product ρ_n.
- When $\rho = \rho_1$, the star and cyclic systems reduce to vacation queues with vacation $2r$ respectively Mr. A standard decomposition result for vacation queues [7] learns that the workload difference equals $\rho_1 \frac{1}{2} Mr - \rho_1 \frac{1}{2}(2r)$; this is confirmed by (4.7).

In our model of star polling, Q_1 is served exhaustively N times during a cycle. Generally speaking, exhaustive service minimizes workload in polling systems with switch-over times, cf. Takagi [13]; therefore it is not surprising that $Diff_E \geq 0$, and that the difference increases roughly linearly in N.

REMARK 4.1
As in Manfield [12], the mean waiting time for the exhaustive station can be explicitly calculated. Assume that all 1-limited stations in the network have the same traffic characteristics and all switch-over times are equal to r; then

$$EW_1^{star} = \frac{\lambda\beta^{(2)}}{2(1-\rho_1)} + \frac{\lambda_1^{(2)}-\lambda_1^2-\lambda_1}{2\lambda_1(1-\rho_1)}\beta_1 + \frac{(\rho-\rho_1)r}{1-\rho_1} + (N-1)\frac{1-\rho_1}{1-\rho}r + \frac{1}{2}.$$

We can substitute this result into (4.5); since in this case all mean waiting times at the 1-limited queues are equal, we obtain the exact mean waiting time at the 1-limited queues.

Case B: Q_1 gated service
In this case the pseudoconservation law reduces to (cf. (4.2)):

$$\rho_1 EW_1^{star} + \sum_{n\in 1l}\rho_n[1-\frac{\lambda_n s}{1-\rho}]EW_n^{star} =$$
$$C + \rho\sum_{m=1}^{M}\frac{s_m^{(2)}}{2s} + \rho_1{\sum_{k\in\tilde{g}}}^{\bullet}\frac{s_{k+1}}{s}(s_k+EV_{k+1}) +$$
$$\sum_{k\in\tilde{1l}}\rho_{T(k)}\sum_{m\neq k}\frac{s_m}{s}{\sum_{j=k}^{m-1}}^{\bullet}(s_j+EV_{j+1}) + \rho_1\sum_{j\in\tilde{g}}EV_j\sum_{m=1}^{M}\frac{s_m}{s}z_{jm} + \rho_1\sum_{m\in\tilde{g}}\frac{s_m}{s}EV_m. \quad (4.8)$$

The mean visit times for the 1-limited pseudostations are again given by (4.3). It follows from (2.17) that

$$k\in\tilde{g}:\quad EV_k = \rho_1[s_{k-1} + EV_{k-1} + \sum_{j=1}^{M}h_{jk}(s_{j-1}+EV_{j-1})] =$$
$$\rho_1[s_{k-2}+s_{k-1}+EV_{k-2}+EV_{k-1}] = \rho_1[s_{k-2}+s_{k-1}+\rho_{T(k-1)}\frac{s}{1-\rho}] + \rho_1 EV_{k-2} =$$
$$\rho_1[s_{k-2}+s_{k-1}+\rho_{T(k-1)}\frac{s}{1-\rho}] + \rho_1^2[s_{k-4}+s_{k-3}+\rho_{T(k-3)}\frac{s}{1-\rho}] +$$
$$\cdots + \rho_1^{\frac{1}{2}M}[s_k+s_{k+1}+\rho_{T(k+1)}\frac{s}{1-\rho}] + \rho_1^{\frac{1}{2}M}EV_k, \quad (4.9)$$

leading to an explicit expression for EV_k.
To simplify the calculations it is again assumed that all M switch-over times are equal to the constant r. Formula (4.9) now reduces to:

$$k\in\tilde{g}:\quad EV_k = \frac{\rho_1}{1-\rho_1}2r + \frac{\rho_1}{1-\rho}R_k Mr, \quad (4.10)$$

with

$$R_k := \frac{1}{1-\rho_1^{\frac{1}{2}M}}[\rho_{T(k-1)}+\rho_1\rho_{T(k-3)}+\rho_1^2\rho_{T(k-5)}+\cdots+\rho_1^{\frac{1}{2}M-1}\rho_{T(k+1)}]. \quad (4.11)$$

Substitution of (4.3) and (4.10) into (4.8) yields:

$$\rho_1 EW_1^{star} + \sum_{n\in 1l}\rho_n[1-\frac{\lambda_n Mr}{1-\rho}]EW_n^{star} =$$

$$C + \rho_1 r + \tfrac{1}{2}(\rho-\rho_1)Mr + \rho_1\frac{\rho-\rho_1}{1-\rho}r + 2\frac{\rho_1^2}{1-\rho}r + \tfrac{1}{2}\rho_1\frac{\rho-\rho_1}{1-\rho_1}Mr +$$

$$\frac{1}{1-\rho}Mr\sum_{j=2}^{N-1}{}'\sum_{i=j+1}^{N}\rho_i\rho_j + \frac{\rho_1}{1-\rho}r\sum_{k\in\tilde{1}l}\rho_{T(k)}[(M-1)R_{k+1}+(M-3)R_{k+3}+\cdots+R_{k-1}]. \qquad (4.12)$$

Again we make a comparison with the corresponding strictly cyclic service system. The pseudoconservation law reads in this case (cf. [4], and compare with (4.6)):

$$\rho_1 EW_1^{cycl} + \sum_{n\in 1l}\rho_n[1-\frac{\lambda_n Mr}{1-\rho}]EW_n^{cycl} =$$

$$C + \tfrac{1}{2}\rho Mr + \frac{Mr}{2(1-\rho)}[\rho^2 - \sum_{n=1}^{N}\rho_n^2] + \frac{\rho_1^2}{1-\rho}Mr. \qquad (4.13)$$

Subtract the right-hand side of (4.12) from the right-hand side of (4.13), and call the difference $Diff_G$. Then

$$Diff_G = \rho_1(N-2)r[1+\frac{\rho-\rho_1}{1-\rho_1}] + \frac{\rho_1(\rho-\rho_1)^2}{(1-\rho)(1-\rho_1)}(2N-3)r +$$

$$\frac{\rho_1^2}{1-\rho}2(N-2)r - \frac{\rho_1}{1-\rho}r\sum_{k\in\tilde{1}l}\rho_{T(k)}[(M-1)R_{k+1}+(M-3)R_{k+3}+\cdots+R_{k-1}]. \qquad (4.14)$$

This result leads to similar observations as the ones for Case A (cf. below (4.7)). To see that $Diff_G \geqslant 0$, consider the last term, LT, in the right-hand side of (4.14). The coefficient of $\rho_{T(k)}^2$ in LT is:

$$\frac{\rho_1}{1-\rho}r\frac{1}{1-\rho_1^{\frac{1}{2}M}}[(M-1)+(M-3)\rho_1+\cdots+3\rho_1^{\frac{1}{2}M-2}+\rho_1^{\frac{1}{2}M-1}] \leqslant$$

$$\frac{\rho_1}{1-\rho}r\frac{1}{1-\rho_1^{\frac{1}{2}M}}(M-1)[1+\rho_1+\ldots+\rho_1^{\frac{1}{2}M-2}+\rho_1^{\frac{1}{2}M-1}] = \frac{\rho_1}{(1-\rho)(1-\rho_1)}(M-1)r.$$

The coefficient of $\rho_{T(k)}\rho_{T(k-2)}$ in LT is:

$$\frac{\rho_1}{1-\rho}r\frac{1}{1-\rho_1^{\frac{1}{2}M}}([(M-1)\rho_1+(M-3)\rho_1^2+\cdots+3\rho_1^{\frac{1}{2}M-1}+1] +$$

$$[(M-1)\rho_1^{\frac{1}{2}M-1}+(M-3)+\ldots+3\rho_1^{\frac{1}{2}M-3}+\rho_1^{\frac{1}{2}M-2}]) \leqslant$$

$$\frac{\rho_1}{(1-\rho)(1-\rho_1)}2(M-1)r.$$

Similarly for the other products. Summing all the upper bounds yields:

$$LT \leqslant \frac{\rho_1}{(1-\rho)(1-\rho_1)}[\rho_{T(2)}+\ldots+\rho_{T(M)}]^2(M-1)r = \frac{\rho_1}{(1-\rho)(1-\rho_1)}(\rho-\rho_1)^2(M-1)r.$$

Hence

$$Diff_G \geqslant \rho_1(N-2)r[1+\frac{\rho-\rho_1}{1-\rho_1}] + \frac{\rho_1^2}{1-\rho}2(N-2)r \geqslant 0.$$

Remark 4.2

Translation of the results of this section to the continuous-time case is almost immediate. In particular, the expressions for $Diff_E$ and $Diff_G$ are not affected.

REMARK 4.3

In the symmetric case $\rho_2 = \cdots = \rho_N = (\rho - \rho_1)/(N-1)$, the formulas (4.7) for $Diff_E$ and (4.14) for $Diff_G$ become very simple:

$$Diff_E = (N-2) r \rho_1 \frac{1-\rho_1}{1-\rho}, \tag{4.15}$$

and

$$Diff_G = (N-2) r \rho_1 \frac{1+\rho_1}{1-\rho}. \tag{4.16}$$

We have also evaluated (3.13) for a network with scan polling (polling table $T = [1,2,...,N-1,N,N,N-1,...,2,1]$, cf. also Takagi & Murata [15]), and we have again compared the result with the network with corresponding stations and strictly cyclic service. Because of the complexity of the calculations, we have restricted ourselves to the case of exhaustive service at all (pseudo-)stations, constant switch-over times r between all pseudostations in the scan network, and equal traffic intensities at all stations: $\rho_1 = \cdots = \rho_N = \rho/N$. If the switch-over times in the cyclic system equal $2r$ (so that the mean cycle times in both systems are the same), then

$$Diff = E\mathbf{V}^{cycl} - E\mathbf{V}^{scan} = \frac{\rho r}{2} + \frac{\rho r(N-1)}{6(1-\rho)} \frac{2N-3\rho-1}{N-\rho} > 0; \tag{4.17}$$

this is not surprising, as the queues in the scan system are visited twice as often as in the cyclic system. However, it seems more realistic to choose the switch-over times in the cyclic system equal to r, just as in the scan system; then

$$Diff = E\mathbf{V}^{cycl} - E\mathbf{V}^{scan} = -\frac{\rho r(N-1)}{6(1-\rho)} \frac{N+1}{N-\rho} \leqslant 0. \tag{4.18}$$

Again, we might have expected this, because of the inefficient visiting pattern of scan polling.

5. CONCLUSIONS

This paper has been devoted to the waiting-time analysis of a polling system in which the stations are visited according to a general service order table. Non-zero switch-over times of the server between the queues are assumed. We have formulated a decomposition for the amount of work in this polling system, into the amount of work in the system *without* switch-over times and an additional term. This decomposition has led to a pseudoconservation law for the mean waiting times, i.e., an exact expression for a weighted sum of the mean waiting times. The pseudoconservation law can be used to obtain or test approximations for individual mean waiting times, and generally to provide insight into the behaviour of polling systems. Accurate yet simple mean waiting-time approximations in polling systems with service order tables would be extremely useful. For example, they could enable some form of optimization, by determining how often a station should be visited compared to other stations.

REFERENCES

[1] BAKER, J.E., RUBIN, I. (1987). Polling with a general-service order table. *IEEE Trans. Commun.*, Vol. COM-35, 283-288.

[2] BOXMA, O.J. (1988). Workloads and waiting times in single-server systems with multiple customer classes. *Report Centre for Mathematics and Computer Science;* invited paper for *Queueing Systems.*

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

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

[5] BROWNE, S., YECHIALI, U. (1987). Dynamic priority rules for cyclic-type queues. *Report Graduate School of Business, Columbia University.*

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

[7] FUHRMANN, S.W., COOPER, R.B. (1985). Stochastic decompositions in the M/G/1 queue with generalized vacations. *Oper. Res.* 33, 1117-1129.

[8] GIANNAKOUROS, N.P., LALOUX, A. (1988). A general conservation law for a priority polling system. *Report Telecommunications Laboratory, Leuven University.*

[9] HUNTER, J.J. (1983). *Mathematical Techniques of Applied Probability, Vol. 2.* Academic Press, New York.

[10] KOBAYASHI, H., KONHEIM, A.G. (1977). Queueing models for computer communications systems analysis. *IEEE Trans. Commun.*, Vol. COM-25, 2-29.

[11] LEVY, H. (1988). Optimization of polling systems via binomial service. *Report Department of Computer Science, Tel Aviv University.*

[12] MANFIELD, D.R. (1985). Analysis of a priority polling system for two-way traffic. *IEEE Trans. Commun.*, Vol. COM-33, 1001-1006.

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

[14] 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.

[15] TAKAGI, H., MURATA, M. (1986). Queueing analysis of scan-type TDM and polling systems. In: *Computer networking and performance evaluation,* eds. T. Hasegawa, H. Takagi, Y. Takahashi, North-Holland Publ. Cy., Amsterdam, 199-211.

[16] 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.