Noname manuscript No.
(will be inserted by the editor)

Fluid Polling Systems

Omer Czerniak · **Uri Yechiali**

Received: date / Accepted: date

Abstract We study N-queues single-server fluid polling systems, where a fluid is continuously flowing into the queues at queue-dependent rates. When visiting and serving a queue, the server reduces the amount of fluid in that queue at a queue-dependent rate. Switching from queue i to queue j requires two random-duration steps: (i) departing queue i, and (ii) reaching queue j. The length of time the server resides in a queue depends on the service regime. We consider three main regimes: Exhaustive, Gated and Globally-Gated. Two polling procedures are analyzed: (i) cyclic and (ii) probabilistic. Under steady-state, we derive the Laplace–Stieltjes transform (LST), mean and second moment of the amount of flow at each queue at polling instants, as well as at an arbitrary moment. We further calculate the LST and mean of the 'waiting time' of a drop at each queue and derive expressions for the mean total load in the system for the various service regimes. Finally, we explore optimal switching procedures.

Keywords Polling models · Fluid · Cyclic · Probabilistic · Workload · Waiting times

Mathematics Subject Classification (2000) Primary 60K25 · Secondary 68M20 · 90B22

1 Introduction

We consider fluid polling systems comprised on N queues and a single server, where a fluid is continuously flowing to queue i at a constant rate α_i $(i = 1, 2, \ldots, N)$. When visiting and serving queue i the server reduces the amount of fluid there at a constant rate μ_i $(\mu_i > \alpha_i)$. Switching from queue i to queue j takes two independent random steps: First it takes D_i

O. Czerniak
Department of Statistics and Operations Research, School of Mathematical Sciences, Tel Aviv University, Israel.
Tel.: +972-54-4671988
E-mail: omercz@gmail.com

U. Yechiali
Department of Statistics and Operations Research, School of Mathematical Sciences, Tel Aviv University, Israel, and Afeka College of Engineering, Tel Aviv, Israel
Tel.: +972-3-6409357
E-mail: uriy@post.tau.ac.il

units of time to depart queue i, and then it takes additional R_j units to reach queue j. The duration of time the server resides in queue i depends on the service regime. We consider three main regimes: Exhaustive, Gated and Globally-Gated. Two polling procedures are studied: (i) cyclic, where the server visits the queues in the order $1,2,\ldots,N-1,N,1,2,\ldots$ and, (ii) probabilistic, where after exiting queue i, the server switches over to queue j with probability p_j, $\sum_{j=1}^{N} p_j = 1$.

Discrete arrival M/G/1-type polling models have been studied extensively in the literature (see e.g. [10], [11], [1], [2], [4], [5], [6], [12], [13] and references there). The main services regimes are Exhaustive, Gated, Globally-Gated and K_i-limited, while the polling procedures vary from pure cyclic through Markovian routing, to general polling tables.

Under steady-state we derive the LST, mean and second moment of the amount of fluid in the queues at a polling instant, as well as at an arbitrary moment. We further calculate the LST and mean of the 'waiting time' of a drop at each queue and derive expressions for the mean total load in the system for the various service regimes. Finally, for the probabilistic polling, we explore optimal switching schemes of the server so as to (i) minimize the sum of fluid amounts at polling instants at the various queues; (ii) minimize the maximum value of the mean amount of fluid in the queues at polling instants, and (iii) minimize the total workload in the system. We compare our results with those obtained for the classical discrete-arrival M/G/1-type polling models, and indicate the differences between the discrete and fluid systems.

The paper is structured as follows. In section 2 we study the cyclic polling procedure and analyze each of the three service regimes: Exhaustive, Gated and Globally-Gated. In section 3 we analyze the corresponding probabilistic polling scheme. Workloads are treated in section 4 and optimal switching policies are discussed in section 5.

Notation: For a continuous random variable X we denote its mean by $E[X] = x$, its second moment by $E[X^2] = x^{(2)}$, and its LST by $\tilde{X}(\cdot)$.

2 Cyclic Switching

Under cyclic switching, the server visits the queues in the order $1,2,\ldots,N-1,N,1,2,\ldots$. Let X_i^j denote the amount of fluid found in queue j $(j = 1,2,\ldots,N)$ when the server arrives at (polls) queue i $(i = 1,2,\ldots,N)$. $X_i = (X_i^1, X_i^2, \ldots, X_i^N)$ is the state of the system at that instant. As described in the Introduction, moving from queue i to queue $i+1$ takes $D_i + R_{i+1}$ units of time. Let $M_i = D_i + R_{i+1}$.

2.1 Cyclic Exhaustive

Under the Exhaustive regime, during a visit, the server clears all the fluid in the queue until it is empty and then moves on to the next queue. Let S_i be the visit time of the server in queue i. Since the server clears all the fluid found at polling instant, as well as all the additional fluid flowing during its visit time, then

$$S_i = \frac{X_i^i}{\mu_i} + \frac{S_i \alpha_i}{\mu_i}, \tag{1}$$

implying that,

$$S_i = \frac{X_i^i}{\mu_i - \alpha_i}. \tag{2}$$

The evolution of the state of the system at polling instants is given by

$$X^j_{i+1} = \begin{cases} X^j_i + \alpha_j(\frac{X^i_i}{\mu_i-\alpha_i} + M_i) & \text{if } j \neq i \\ \alpha_i M_i & \text{if } j = i. \end{cases} \tag{3}$$

That is, the amount of fluid in queue j at a polling instant of queue $i+1$ ($j \neq i$) is composed of: (i) the amount in queue j at a polling instant of queue i, and (ii) the fluid that was accumulated in queue j during the service time of queue i and the switchover time from i to $i+1$. In the case where $i = j$ the fluid in queue i is only the new accumulated fluid during the switchover time (the server leaves queue i with zero fluid).
Our first goal is to derive the multidimensional LST $L_i(\underline{\theta})$ of the state of the system at a polling instant of queue i ($i = 1,2,\ldots,N$). This transform is defined as

$$L_i(\underline{\theta}) = L_i(\theta_1,\ldots,\theta_{i-1},\theta_i,\theta_{i+1},\ldots,\theta_N) = E[e^{-\sum_{j=1}^N \theta_j X^j_i}]. \tag{4}$$

Using (3) and (4) we obtain $L_{i+1}(\underline{\theta})$ in terms of $L_i(\cdot)$, namely, for $i = 1,2,\ldots,N$,

$$\begin{aligned} L_{i+1}(\underline{\theta}) &= E[e^{-\sum_{j=1}^N \theta_j X^j_{i+1}}] = E[e^{-\sum_{\substack{j=1\\ j\neq i}}^N \theta_j (X^j_i + \frac{\alpha_j X^i_i}{\mu_i-\alpha_i})}]E[e^{-\sum_{j=1}^N \theta_j \alpha_j M_i}] \\ &= L_i(\theta_1,\ldots,\theta_{i-1},\frac{1}{\mu_i-\alpha_i}\sum_{\substack{j=1\\ j\neq i}}^N \theta_j\alpha_j,\theta_{i+1},\ldots,\theta_N)\cdot \tilde{M}_i(\sum_{j=1}^N \theta_j\alpha_j). \end{aligned} \tag{5}$$

Equations (5) are now used to derive moments of the variables X^j_i.

Fluid Amount at Polling Instants
First Moments
The mean fluid amount, $f_i(j) \triangleq E[X^j_i]$, at queue j when the server polls queue i is given by

$$f_i(j) \triangleq E[X^j_i] = -\frac{\partial L_i(\underline{\theta})}{\partial \theta_j}\Big|_{\underline{\theta}=\underline{0}}. \tag{6}$$

This leads to

$$f_{i+1}(j) = \begin{cases} f_i(j) + \frac{\alpha_j}{\mu_i-\alpha_i} f_i(i) + \alpha_j m_i & \text{if } j \neq i \\ \alpha_i m_i & \text{if } j = i. \end{cases} \tag{7}$$

Clearly, equation (7) can be obtained directly from (3) by taking expectations.
Set $\rho_k = \frac{\alpha_k}{\mu_k}$, $\rho = \sum_{k=1}^N \rho_k$, $m = \sum_{i=1}^N m_i$, where ρ_k is the rate of work flowing into queue k, and ρ is the total rate of work flowing into the system. The solution of (7) is

$$f_i(j) = \begin{cases} \alpha_j\left(\sum_{k=j}^{i-1} m_k + \sum_{k=j+1}^{i-1} \rho_k \frac{m}{1-\rho}\right) & \text{if } j \neq i \\ \alpha_i(1-\rho_i)\frac{m}{1-\rho} & \text{if } j = i, \end{cases} \tag{8}$$

where the summation is cyclic. As in the regular M/G/1-type polling systems, a necessary and sufficient condition for stability is $\rho < 1$. Note also the similarity between $f_i(j)$ given by (8) and the corresponding $f_i(j)$ of the regular (discrete arrivals) M/G/1-type model (see for example Takagi [10] or Yechicali [12]).

Second Moments
The second and mixed moments of the X_i^j are given by

$$f_i(j,k) \triangleq E[X_i^j X_i^k] = \frac{\partial^2 L_i(\underline{\theta})}{\partial \theta_j \partial \theta_k}\Big|_{\underline{\theta}=\underline{0}}. \tag{9}$$

Then,

$$Var[X_i^i] = f_i(i,i) - f_i^2(i) = f_i(i,i) - \frac{\alpha_i^2(1-\rho_i)^2 m^2}{(1-\rho)^2}. \tag{10}$$

Differentiating (5) with respect to θ_j and θ_k, we get N^3 linear equations (see Appendix A.1). Apparently there is no closed-form solution to equations (167)-(169), they may be solved numerically as a system of linear equations. For the special case of identical queues, setting $\rho_i = \rho_1, m_i = m_1, \alpha_i = \alpha_1$ and $\mu_i = \mu_1$, we obtain, after algebraic manipulations,

$$f_i(i) = \frac{(1-\rho_1)\alpha_1 m}{1-\rho}, \tag{11}$$

$$f_i(i,i) = \frac{N(1-\rho_1)\alpha_1^2 m_1^{(2)}}{1-\rho} + \frac{N(N-1)(1-\rho_1)\alpha_1^2 m_1^2}{(1-\rho)^2}, \tag{12}$$

$$Var[X_i^i] = \frac{N(1-\rho_1)\alpha_1^2(m_1^{(2)} - m_1^2)}{1-\rho}, \tag{13}$$

where $N\rho_1 = \rho$ and $Nm_1 = m$.

Visit Time, Intervisit Time and Cycle Time
We set

$$\tilde{X}_i^i(s) \triangleq E[e^{-sX_i^i}] = L_i(0,0,0,\ldots,0,s,0,\ldots,0). \tag{14}$$

Visit Time
Let S_i denote the visit time (within a cycle) of the server at queue i. From (2) we have,

$$\tilde{S}_i(s) = \tilde{X}_i^i(\frac{s}{\mu_i - \alpha_i}), \tag{15}$$

$$E[S_i] = \frac{E[X_i^i]}{\mu_i - \alpha_i} = \frac{\rho_i m}{1-\rho}, \tag{16}$$

$$Var[S_i] = \frac{Var[X_i^i]}{(\mu_i - \alpha_i)^2} = \frac{f_i(i,i)}{(\mu_i - \alpha_i)^2} - \frac{\rho_i^2 m^2}{(1-\rho)^2}. \tag{17}$$

In the case of identical queues we have,

$$Var[S_i] = \frac{N\rho_1^2(m_1^{(2)} - m_1^2)}{(1-\rho)(1-\rho_1)}. \tag{18}$$

Intervisit Time
Let I_i be the intervisit period of queue i. That is, the period within a cycle beginning at the time when service is completed at queue i, and ending at the instant when queue i is polled

again in the next cycle. Note that the fluid amount in queue i at polling instant of that queue is equal to the amount accumulated there during the preceding intervisit time. Then

$$X_i^i = \alpha_i I_i, \tag{19}$$

from which we derive the LST and moments of S_i,

$$\tilde{I}_i(s) = \tilde{X}_i^i(\frac{s}{\alpha_i}), \tag{20}$$

$$E[I_i] = \frac{E[X_i^i]}{\alpha_i} = \frac{(1-\rho_i)m}{1-\rho}, \tag{21}$$

$$Var[I_i] = \frac{Var[X_i^i]}{\alpha_i^2} = \frac{f_i(i,i)}{\alpha_i^2} - \frac{(1-\rho_i)^2 m^2}{(1-\rho)^2}, \tag{22}$$

In the case of identical queues we have,

$$Var[I_i] = \frac{N(1-\rho_1)(m_1^{(2)} - m_1^2)}{1-\rho}. \tag{23}$$

Cycle Time
Let C_i be the cycle time for queue i. C_i spans the period beginning at the time when queue i is polled and ending at the time when it is polled again in the next cycle. In steady state, C_i may also be defined as the period beginning at the time when service is completed in queue i at a given cycle and ending at the time of its next service completion,

$$C_i = I_i + S_i = I_i + \frac{\alpha_i I_i}{\mu_i - \alpha_i} = \frac{I_i}{1-\rho_i} = \frac{S_i}{\rho_i} = \frac{X_i^i}{\alpha_i(1-\rho_i)}. \tag{24}$$

Hence,

$$X_i^i = \alpha_i(1-\rho_i)C_i, \tag{25}$$

meaning, indeed, that X_i^i is equal to the accumulated fluid during the intervisit time of queue i. Hence,

$$\tilde{C}_i(s) = \tilde{X}_i^i(\frac{s}{\alpha_i(1-\rho_i)}), \tag{26}$$

$$E[C_i] = \frac{E[X_i^i]}{\alpha_i(1-\rho_i)} = \frac{m}{1-\rho}, \tag{27}$$

$$Var[C_i] = \frac{Var[X_i^i]}{\alpha_i^2(1-\rho_i)^2} = \frac{Var[S_i]}{\rho_i^2} = \frac{f_i(i,i)}{\alpha_i^2(1-\rho_i)^2} - \frac{m^2}{(1-\rho)^2}. \tag{28}$$

In the case of identical queues we have,

$$Var[C_i] = \frac{N(m_1^{(2)} - m_1^2)}{(1-\rho_1)(1-\rho)}. \tag{29}$$

Clearly, as in many other polling models, the mean cycle time can be obtained when using the following balance equation $E[C_i] = \rho E[C_i] + m$.

Now we can explain (8). The mean cycle time is $E[C] = \frac{m}{1-\rho}$, and the mean time spent in each queue is $E[S_i] = \rho_i E[C]$. For queue $i = j$, the elapsed time since last leaving queue i until the next polling instant of queue i is $(1-\rho_i)E[C]$. During that time the amount of fluid in queue i is constantly increasing by a rate of α_i and the total fluid accumulated at queue i

during that time is $\alpha_i(1-\rho_i)E[C]$. For queue $i \neq j$, the elapsed time since leaving queue j, until the next polling instant of queue i, is the switchover times from queue j to queue i and the time spent in each queue between $j+1$ to $i-1$.

Fluid Amount at Arbitrary Times
Let L_i denote the fluid amount in queue i at an arbitrary moment, and let $L_i(t)$ be the fluid in queue i at time t within C_i. The LST of L_i is calculated by dividing the expected area of the function $e^{-sL_i(t)}$ over an arbitrary cycle, by the expected cycle time. That is,

$$\tilde{L}_i(s) = E[e^{-sL_i}] = \frac{E[\int_0^{C_i} e^{-sL_i(t)}dt]}{E[C_i]}. \tag{30}$$

Figure 1 shows the fluid amount in queue i during a full cycle. During the intervisit time,

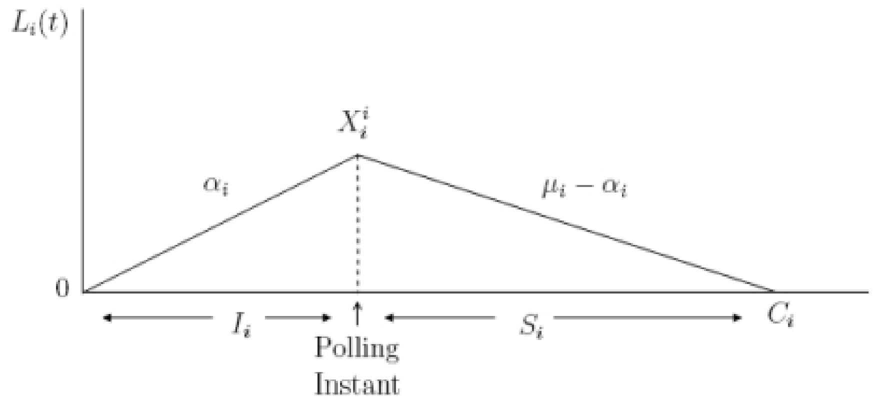

Fig. 1 Exhaustive: Changes of fluid in queue i during a cycle.

the amount of fluid is constantly increasing at a rate of α_i, while during the visit time, the amount of fluid is constantly decreasing at a rate of $\mu_i - \alpha_i$. Then,

$$\tilde{L}_i(s) = \frac{E[\int_0^{I_i} e^{-sL_i(t)}dt + \int_{I_i}^{C_i} e^{-sL_i(t)}dt]}{E[C_i]}. \tag{31}$$

We have,

$$E[\int_0^{I_i} e^{-sL_i(t)}dt] = E[\int_0^{I_i} e^{-s\alpha_i t}dt] = \frac{1}{\alpha_i s}(1-\tilde{I}_i(\alpha_i s)) = \frac{1}{\alpha_i s}(1-\tilde{X}_i^i(s)). \tag{32}$$

Since $S_i = \frac{\alpha_i I_i}{\mu_i - \alpha_i}$ we get,

$$\begin{aligned} E[\int_{I_i}^{C_i} e^{-sL_i(t)}dt] &= E[\int_0^{S_i} e^{-s(\mu_i-\alpha_i)t}dt] = \frac{1}{(\mu_i-\alpha_i)s}\Big(1-\tilde{S}_i\big((\mu_i-\alpha_i)s\big)\Big) \\ &= \frac{1}{(\mu_i-\alpha_i)s}(1-\tilde{X}_i^i(s)). \end{aligned} \tag{33}$$

From (27), (30), (32) and (33) we obtain

$$\tilde{L}_i(s) = \frac{1-\tilde{X}_i^i(s)}{sE[X_i^i]}, \tag{34}$$

which is the LST of the residual value of X_i^i. From (34) we get,

$$E[L_i] = \frac{E[X_i^{i2}]}{2E[X_i^i]}. \tag{35}$$

Note that in the regular Exhaustive M/G/1 polling systems, $E[L_i]$ (number of customers in queue i) is equal to our result plus additional two terms: (i) ρ_i - which is equal to the mean number of customers from queue i that are being served, and (ii) a function of the second moment of the service time. Term (i) is missing from our system since in our case the instantaneous amount of fluid that is being served equals zero; term (ii) is missing since in our case the 'service time' equals zero as well.
In the case of identical queues we have,

$$E[L_i] = \alpha_1\Big(\frac{m_1^{(2)}-m_1^2}{2m_1} + \frac{m(1-\rho_1)}{2(1-\rho)}\Big) \tag{36}$$

Waiting Times
Calculating the LST of the 'waiting time' of an arbitrary drop of fluid is similar to the calculation of the fluid amount at an arbitrary time. Let $W_i(t)$ be the waiting time of a drop that arrives at time t to queue i (within C_i), and let W_i denote the waiting time of an arbitrary drop. We express $\tilde{W}_i(\cdot)$, the LST of W_i, as the time average of $e^{-sW_i(t)}$ over a polling cycle,

$$\tilde{W}_i(s) = E[e^{-sW_i}] = \frac{E[\int_0^{C_i} e^{-sW_i(t)}dt]}{E[C_i]}. \tag{37}$$

During the intervisit time, a drop arriving at time t ($0 < t < I_i$) waits until the next polling instant of queue i plus additional time of clearing all the drops that arrived before that drop

$$W_i(t) = I_i - t + \frac{\alpha_i t}{\mu_i}. \tag{38}$$

Then,

$$E[\int_0^{I_i} e^{-sW_i(t)}dt] = E[\int_0^{I_i} e^{-s(I_i-t+\frac{\alpha_i t}{\mu_i})}dt] = \frac{1}{s(1-\rho_i)}(\tilde{I}_i(s\rho_i)-\tilde{I}_i(s)). \tag{39}$$

During the service time where $I_i < t < C_i$, a drop waits $\frac{L_i(t)}{\mu_i}$ units of time, then

$$W_i(t) = \frac{L_i(t)}{\mu_i} = \frac{\alpha_i I_i - (\mu_i-\alpha_i)(t-I_i)}{\mu_i} = \rho_i t + (I_i - t), \tag{40}$$

i.e.

$$E[\int_{I_i}^{C_i} e^{-sW_i(t)}dt] = E[\int_{I_i}^{C_i} e^{-s(\rho_i t+(I_i-t))}dt] = \frac{1}{s(1-\rho_i)}\big(1-\tilde{I}_i(s\rho_i)\big), \tag{41}$$

where we used $C_i = \frac{I_i}{1-\rho_i}$ (see equation (24)).
Combining (37), (39) and (41) we have,

$$\tilde{W}_i(s) = \frac{1-\rho}{m(1-\rho_i)} \cdot \frac{1-\tilde{I}_i(s)}{s} = \frac{1-\tilde{I}_i(s)}{sE[I_i]}, \tag{42}$$

and

$$E[W_i] = \frac{1-\rho}{m(1-\rho_i)} \cdot \frac{E[I_i^2]}{2} = \frac{E[I_i^2]}{2E[I_i]} = \frac{E[X_i^{i2}]}{2\alpha_i E[X_i^i]}. \tag{43}$$

Indeed, Little's law applies: $E[L_i] = \alpha_i E[W_i]$. Note also the analogy of (43) to the result for the waiting time in the M/G/1 queue with multiple server vacations [9], where the waiting time is the sum of the waiting time in a regular M/G/1 queue (with no vacations) and the residual time of the vacation period. In our case, since there is no waiting time in a non-vacation fluid queue (since $\mu_i > \alpha_i$), we are left with the residual time of the 'vacation' period. In fact, the right-hand side of (42) is the LST of the residual time of the intervisit time I_i.
In the case of identical queues we have

$$E[W_i] = \frac{m_1^{(2)} - m_1^2}{2m_1} + \frac{m(1-\rho_1)}{2(1-\rho)}. \tag{44}$$

2.2 Cyclic Gated

Under the Gated regime, during a visit to queue i the server clears all and only the quantity X_i^i and then moves to the next queue. Thus, the "clearing" (visit) time of queue i takes $\frac{X_i^i}{\mu_i}$ units of time, and moving from queue i to queue $i+1$ takes M_i units of time. Then, for the cyclic Gated regime, the evolution of the state of the system is given by

$$X_{i+1}^j = \begin{cases} X_i^j + \alpha_j(\frac{X_i^i}{\mu_i} + M_i) & \text{if } j \neq i \\ \alpha_i(\frac{X_i^i}{\mu_i} + M_i) & \text{if } j = i. \end{cases} \tag{45}$$

That is, the amount of fluid in queue j at a polling instant of queue $i+1$ ($j \neq i$) is composed of: (i) the amount of fluid in queue j at a polling instant of queue i, and (ii) the fluid that was accumulated in queue j during the service time of queue i and the switching time from queue i to queue $i+1$. In the case where $i = j$ the fluid in queue i is only the new accumulated fluid.
Then, for $i = 1, 2, \ldots, N$,

$$\begin{aligned} L_{i+1}(\underline{\theta}) &= E[e^{-\sum_{j=1}^N \theta_j X_{i+1}^j}] = E[e^{-\sum_{j=1, j\neq i}^N \theta_j X_i^j} e^{-\sum_{j=1}^N \theta_j \alpha_j \frac{X_i^i}{\mu_i}}] E[e^{-\sum_{j=1}^N \theta_j \alpha_j M_i}] \\ &= L_i(\theta_1, \ldots, \theta_{i-1}, \frac{1}{\mu_i} \sum_{j=1}^N \theta_j \alpha_j, \theta_{i+1}, \ldots, \theta_N) \cdot \tilde{M}_i(\sum_{j=1}^N \theta_j \alpha_j). \end{aligned} \tag{46}$$

Fluid Amount at Polling Instants: Moments
First Moments
Taking the derivative of (46) with respect to θ_j and then letting $\underline{\theta} = \underline{0}$, leads to the following N^2 linear equations:

$$f_{i+1}(j) = \begin{cases} f_i(j) + \alpha_j m_i + \frac{\alpha_j}{\mu_i} f_i(i) & \text{if } j \neq i \\ \alpha_i m_i + \frac{\alpha_i}{\mu_i} f_i(i) & \text{if } j = i. \end{cases} \tag{47}$$

The solution of (47) is given by

$$f_i(j) = \begin{cases} \alpha_j \left(\sum_{k=j}^{i-1} (m_k + \frac{\rho_k m}{1-\rho}) \right) & \text{if } j \neq i \\ \frac{\alpha_i m}{1-\rho} & \text{if } j = i. \end{cases} \tag{48}$$

Second Moments
The mixed moments of the X_i^j are also derived from the set of LSTs by taking derivatives. As in the Exhaustive regime we get the N^3 linear equations (see Appendix A.2). In the special case of identical queues we obtain,

$$f_i(i,i) = \frac{N\alpha_1^2 m_1^{(2)}}{(1-\rho)(1+\rho_1)} + \frac{N(N-1)\alpha_1^2 m_1^2}{(1-\rho)^2(1+\rho_1)} + \frac{2N^2\alpha_1^2 m_1^2 \rho_1}{(1-\rho)^2(1+\rho_1)}, \tag{49}$$

$$Var[X_i^i] = \frac{N\alpha_1^2(m_1^{(2)} - m_1^2)}{(1-\rho)(1+\rho_1)}. \tag{50}$$

Note that $Var_{\text{Gated}}[X_i^i] = \frac{Var_{\text{Exhaustive}}[X_i^i]}{1-\rho_1^2}$. Clearly, as in the classical discrete-arrival M/G/1-type polling systems, $Var_{\text{Gated}}[X_i^i] > Var_{\text{Exhaustive}}[X_i^i]$.

Visit Time, Intervisit Time and Cycle Time
Visit Time
Since $S_i = \frac{X_i^i}{\mu_i}$, then

$$\tilde{S}_i(s) = \tilde{X}_i^i(\frac{s}{\mu_i}), \tag{51}$$

$$E[S_i] = \frac{E[X_i^i]}{\mu_i} = \frac{\rho_i m}{1-\rho}, \tag{52}$$

$$Var[S_i] = \frac{Var[X_i^i]}{\mu_i^2} = \frac{f_i(i,i)}{\mu_i^2} - \frac{\rho_i^2 m^2}{(1-\rho)^2}. \tag{53}$$

Note that (52) coincides with (16). In the case of identical queues we have,

$$Var[S_i] = \frac{N\rho_1^2(m_1^{(2)} - m_1^2)}{(1-\rho)(1+\rho_1)} = \frac{1-\rho_1}{1+\rho_1} \cdot Var_{\text{Exhaustive}}[S_i]. \tag{54}$$

Cycle and Intervisit Times
Since the fluid in queue i at its polling instant is the accumulated fluid during the pervious cycle, we have $X_i^i = \alpha_i C_i$. Thus

$$\tilde{C}_i(s) = \tilde{X}_i^i(\frac{s}{\alpha_i}), \tag{55}$$

$$E[C_i] = \frac{m}{1-\rho}, \tag{56}$$

$$Var[C_i] = \frac{Var[X_i^i]}{\alpha_i^2} = \frac{f_i(i,i)}{\alpha_i^2} - \frac{m^2}{(1-\rho)^2}. \tag{57}$$

In the case of identical queues we have,

$$Var[C_i] = \frac{N(m_1^{(2)} - m_1^2)}{(1-\rho)(1+\rho_1)} = \frac{1-\rho_1}{1+\rho_1} \cdot Var_{\text{Exhaustive}}[C_i]. \tag{58}$$

The mean intervisit time is given by

$$E[I_i] = E[C_i] - E[S_i] = \frac{m(1-\rho_i)}{1-\rho}. \tag{59}$$

Now we can explain (48). For queue $i = j$, the elapsed time since the last gating instant of queue i until the next gating (polling) instant of that queue is the cycle time. During that time the arrival rate to queue i is α_i and the total fluid accumulated in queue i during that time is $\alpha_i E[C]$. In the case where $i \neq j$, we know that the mean time spent in queue k is $\rho_k E[C]$ and the elapsed time since the last gating instant of queue j, until the next polling instant of queue i, is the total time spent in all queues between j to $i-1$ plus the switchover times from queue to queue.

Fluid Amount at Arbitrary Times

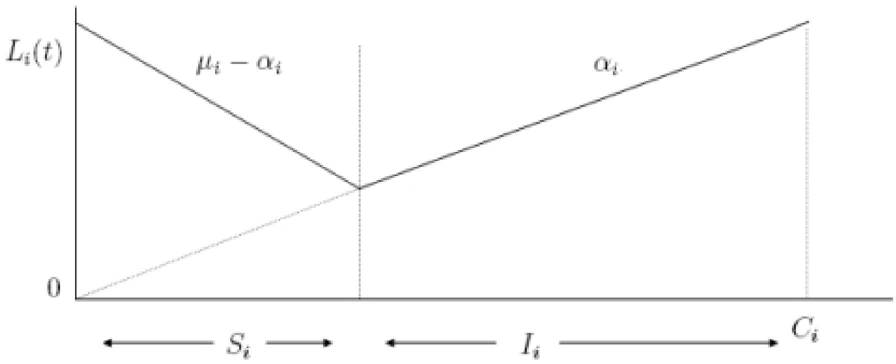

Fig. 2 Gated: Changes of the fluid during a cycle.

Equation (30) is valid for the Gated regime as well. Unlike the Exhaustive regime, we examine the cycle time first during the service time and then during the intervisit time. From Figure 2 we see that the fluid amount at time t $(0 < t < S_i)$, during the service time of length S_i, is $\alpha_i S_i + (S_i - t)(\mu_i - \alpha_i)$ or equivalently $\alpha_i t + (S_i - t)\mu_i$, i.e. the fluid amount at time t is the total accumulated fluid $\alpha_i t$ plus the fluid that needs to be 'served' during the remaining time of the service period $(S_i - t)\mu_i$. The fluid amount at time t $(t > S_i)$ during the intervisit time is simply $\alpha_i t$. Since

$$\tilde{L}_i(s) = \frac{E[\int_0^{S_i} e^{-sL_i(t)}dt] + E[\int_{S_i}^{C_i} e^{-sL_i(t)}dt]}{E[C_i]}, \tag{60}$$

we have,

$$\begin{aligned} E[\int_{0_i}^{s_i} e^{-sL_i(t)}dt] = E[\int_0^{S_i} e^{-s(\alpha_i t + (S_i - t)\mu_i)}dt] &= \frac{1}{(\mu_i - \alpha_i)s}\left(\tilde{S}_i(\alpha_i s) - \tilde{S}_i(\mu_i s)\right) \\ = \frac{1}{s(\mu_i - \alpha_i)}\left(\tilde{X}_i^i(\rho_i s) - \tilde{X}_i^i(s)\right), \end{aligned} \tag{61}$$

and

$$E[\int_{S_i}^{C_i} e^{-sL_i(t)}dt] = E[\int_{S_i}^{C_i} e^{-s\alpha_i t}dt] = \frac{1}{\alpha_i s}\left(\tilde{S}_i(\alpha_i s) - \tilde{C}_i(\alpha_i s)\right) = \frac{1}{\alpha_i s}\left(\tilde{X}_i^i(\rho_i s) - \tilde{X}_i^i(s)\right). \tag{62}$$

From (60), (61) and (62),

$$\tilde{L}_i(s) = \frac{\tilde{X}_i^i(\rho_i s) - \tilde{X}_i^i(s)}{s(1-\rho_i)E[X_i^i]}, \tag{63}$$

then

$$E[L_i] = \frac{(1+\rho_i)E[X_i^{i2}]}{2E[X_i^i]}. \tag{64}$$

Note that in the regular Exhaustive M/G/1 polling systems $E[L_i]$ is equal to our result plus additional term: ρ_i, which is equal to the mean number of customers from queue i that are being served (unlike the Exhaustive regime, in the M/G/1-type Gated regime $E[L_i]$ doesn't depend on the second moment of the service time).
In the case of identical queues we have,

$$E[L_i] = \alpha_1\Big(\frac{m_1^{(2)} - m_1^2}{2m_1} + \frac{m(1+\rho_1)}{2(1-\rho)}\Big) > E_{\text{Exhaustive}}[L_i]. \tag{65}$$

Waiting Times
A drop that arrives at queue i at time t ($0 < t < C_i$) waits until the next polling instant of that queue plus additional time of clearing all the fluid accumulated before time t. Then

$$W_i(t) = C_i - t + \frac{\alpha_i t}{\mu_i}. \tag{66}$$

We therefore have,

$$E[\int_0^{C_i} e^{-sW_i(t)}dt] = E[\int_0^{C_i} e^{-s(C_i - t + \frac{\alpha_i t}{\mu_i})}dt] = \frac{1}{s(1-\rho_i)}\left(\tilde{C}_i(s\rho_i) - \tilde{C}_i(s)\right). \tag{67}$$

Combining (37) and (67) leads to,

$$\tilde{W}_i(s) = \frac{\tilde{C}_i(\rho_i s) - \tilde{C}_i(s)}{s(1-\rho_i)E[C_i]}, \tag{68}$$

from which we get,

$$E[W_i] = \frac{(1+\rho_i)E[C_i^2]}{2E[C_i]} = \frac{1+\rho_i}{\alpha_i}\cdot\frac{E[X_i^{i2}]}{2E[X_i^i]}. \tag{69}$$

In the case of identical queues we have,

$$E[W_i] = \frac{m_1^{(2)} - m_1^2}{2m_1} + \frac{m(1+\rho_1)}{2(1-\rho)} > E_{\text{Exhaustive}}[W_i]. \tag{70}$$

2.3 Cyclic Globally-Gated

Globally-Gated service scheme [5] uses a time-stamp mechanism for its operation: the server moves cyclically among the queues and uses the instant of cycle-beginning as a reference point of time; when it reaches a queue it clears all (and only) the fluid that was present at that queue at the moment of cycle-beginning.

'Gated' Cycle Time
Assume, without loss of generality, that a cycle starts at queue 1, then $(X_1^1, \ldots, X_1^N)$ is the state of the system at the beginning of a cycle. For simplicity we write $X_j = X_1^j$, then, the cycle duration is

$$C = M + \sum_{j=1}^{N} \frac{X_j}{\mu_j}, \tag{71}$$

where $M = \sum_{i=1}^{N} M_i$ is the total switchover time in a cycle.
Since $\alpha_j C = X_j$, the LST of C is derived as follows

$$\tilde{C}(s) = E[e^{-sC}] = E[e^{-sM - s\sum_{j=1}^{N} \frac{X_j}{\mu_j}}] = E[e^{-sM - s\sum_{j=1}^{N} \rho_j C}] = \tilde{M}(s) \cdot \tilde{C}(s \sum_{j=1}^{N} \rho_j). \tag{72}$$

The last equality is due to independence of the switch-over times of the current cycle and the length of previous cycle. By differentiating (72), we have

$$E[C] = \frac{m}{1-\rho} \tag{73}$$

We see, again, that the expected cycle time is identical to the expected cycle time for both the Exhaustive and the Gated regime.
The second moment of C is computed through (72)

$$E[C^2] = \frac{1}{1-\rho^2}\Big(m^{(2)} + 2m\rho E[C]\Big), \tag{74}$$

and

$$Var[C] = \frac{m^{(2)} - m^2}{1-\rho^2}. \tag{75}$$

Note that the moments of C can be computed directly from (71) as well. Comparing (74) with equation (2.9) in [5], we observe, again, that the term involving the second moment of the service time, vanishes.

Visit Time and Cycle Time
Visit Time
Since $X_i = \alpha_i C$, we have

$$S_i = \rho_i C. \tag{76}$$

Hence,

$$E[S_i] = \frac{\rho_i m}{1-\rho}, \tag{77}$$

$$Var[S_i] = \frac{\rho_i^2 (m^{(2)} - m^2)}{1-\rho^2}. \tag{78}$$

Note that (77) coincides with (16) and (52).

Cycle Time
Let C_i be the cycle time for queue i. C_i is composed of: (i) the period from polling instant of queue i until the end of the current (gated) cycle, $\sum_{j=i}^{N}\frac{X_j}{\mu_j}+\sum_{j=i}^{N}M_j$ and (ii) the time elapsing from the beginning of the next cycle until the next polling instant of queue i, $\rho_1(\sum_{j=1}^{N}\frac{X_j}{\mu_j}+\sum_{j=1}^{N}M_j)+M_1^*+\cdots+\rho_{i-1}(\sum_{j=1}^{N}\frac{X_j}{\mu_j}+\sum_{j=1}^{N}M_j)+M_{i-1}^*$ (where M_j^* is the switching time from queue j in the next cycle and $M_j^*\sim M_j$), then, by using $X_i=\alpha_i C$, we get

$$C_i = C(\rho-\rho^{(i)}+\rho\rho^{(i)})+\rho^{(i)}\sum_{j=1}^{i-1}M_j+(1+\rho^{(i)})\sum_{j=i}^{N}M_j+\sum_{j=1}^{i-1}M_j^* \tag{79}$$

where $\rho^{(i)}=\sum_{j=1}^{i-1}\rho_j$. Hence

$$\begin{aligned}\tilde{C}_i(s) &= E[e^{-sC_i}] = E[e^{-s\left(C(\rho-\rho^{(i)}+\rho\rho^{(i)})+\rho^{(i)}\sum_{j=1}^{i-1}M_j+(1+\rho^{(i)})\sum_{j=i}^{N}M_j+\sum_{j=1}^{i-1}M_j^*\right)}] \\ &= \tilde{C}\big(s(\rho-\rho^{(i)}+\rho\rho^{(i)})\big)\cdot\prod_{j=1}^{i-1}\tilde{M}_j(s\rho^{(i)})\cdot\prod_{j=i}^{N}\tilde{M}_j\big(s(1+\rho^{(i)})\big)\cdot\prod_{j=1}^{i-1}\tilde{M}_j(s).\end{aligned} \tag{80}$$

Differentiation (80) at $s=0$ leads to $E[C_i]=E[C]$. In addition we get that

$$Var[C_i] = Var[C]\Big((\rho-\rho^{(i)}+\rho\rho^{(i)})^2+(1+{\rho^{(i)}}^2)(1-\rho^2)\Big)+2\rho^{(i)}\sum_{j=i}^{N}(m_j^{(2)}-m_j^2). \tag{81}$$

In the case of identical queues

$$Var[C_i] = \frac{(m_1^{(2)}-m_1^2)}{1+\rho}\Big(\frac{N}{(1-\rho)}+2\rho_1(i-1)(N-i+1)\Big). \tag{82}$$

Fluid Amount at Arbitrary Times
We derive the LST of the fluid amount at arbitrary times in queue i, $\tilde{L}_i(s)$, by the same technique that has been already used in the cyclic Gated regime,

$$\tilde{L}_i(s) = E[e^{-sL_i}] = \frac{E[\int_0^C e^{-sL_i(t)}dt]}{E[C]}. \tag{83}$$

First we evaluate the expected area of $e^{-sL_i(t)}$ over an arbitrary cycle of length C, or equivalently of length $\sum_{k=1}^{N}\big(\frac{X_k}{\mu_k}+M_k\big)$. We split the cycle time into (i) the time from the start of the cycle at queue 1 until polling instant of queue i, (ii) the visit time of queue i, and (iii) the time from finishing service at queue i until the beginning of the next cycle. Define the polling instant of queue i as

$$F_i \triangleq \sum_{k=1}^{i-1}\big(\frac{X_k}{\mu_k}+M_k\big), \tag{84}$$

hence,

$$\begin{aligned}E[\int_0^C e^{-sL_i(t)}dt] &= E[\int_0^{\sum_{k=1}^{N}(\frac{X_k}{\mu_k}+M_k)} e^{-sL_i(t)}dt] \\ &= E[\int_0^{F_i} e^{-sL_i(t)}dt]+E[\int_{F_i}^{F_i+\frac{X_i}{\mu_i}} e^{-sL_i(t)}dt]+E[\int_{F_i+\frac{X_i}{\mu_i}}^{\sum_{k=1}^{N}\left(\frac{X_k}{\mu_k}+M_k\right)} e^{-sL_i(t)}dt].\end{aligned} \tag{85}$$

For (i) the fluid amount at queue i at time t is $X_i+\alpha_i t$, where $0<t\le F_i$, hence

$$E\Big[\int_0^{F_i} e^{-s(X_i+\alpha_i t)}dt\Big] = -\frac{\tilde{C}\big(s\alpha_i(1+\sum_{k=1}^{i-1}\rho_k)\big)\prod_{k=1}^{i-1}\tilde{M}_k(s\alpha_i)}{s\alpha_i} + \frac{\tilde{C}(s\alpha_i)}{s\alpha_i}. \tag{86}$$

For (ii) the fluid amount at queue i at t is $X_i+\alpha_i t-\mu_i(t-F_i)$, where $F_i<t\le F_i+\frac{X_i}{\mu_i}$. Thus,

$$\begin{aligned} E\Big[\int_{F_i}^{F_i+\frac{X_i}{\mu_i}} e^{-s\left(X_i+\alpha_i t-\mu_i(t-F_i)\right)}dt\Big] &= \frac{\tilde{C}\big(s\alpha_i\sum_{k=1}^{i}\rho_k\big)\prod_{k=1}^{i-1}\tilde{M}_k(s\alpha_i)}{s(\mu_i-\alpha_i)} \\ &\quad - \frac{\tilde{C}\big(s\alpha_i(1+\sum_{k=1}^{i-1}\rho_k)\big)\prod_{k=1}^{i-1}\tilde{M}_k(s\alpha_i)}{s(\mu_i-\alpha_i)}. \end{aligned} \tag{87}$$

For (iii) the fluid amount at queue i at t is $\alpha_i t$, where $F_i+\frac{X_i}{\mu_i}<t\le\sum_{k=1}^{N}\big(\frac{X_k}{\mu_k}+M_k\big)$. Then,

$$E\Big[\int_{F_i+\frac{X_i}{\mu_i}}^{\sum_{k=1}^{N}\left(\frac{X_k}{\mu_k}+M_k\right)} e^{-s\alpha_i t}dt\Big] = -\frac{\tilde{C}(s\alpha_i)}{s\alpha_i} + \frac{\tilde{C}\big(s\alpha_i\sum_{k=1}^{i}\rho_k\big)\prod_{k=1}^{i-1}\tilde{M}_k(s\alpha_i)}{s\alpha_i}. \tag{88}$$

Summing (86)-(88) and dividing by $E[C]$, we have

$$\tilde{L}_i(s) = \frac{\prod_{k=1}^{i-1}\tilde{M}_k(s\alpha_i)\cdot\Big(\tilde{C}(s\alpha_i\sum_{k=1}^{i}\rho_k)-\tilde{C}\big(s\alpha_i(1+\sum_{k=1}^{i-1}\rho_k)\big)\Big)}{s\alpha_i(1-\rho_i)E[C]}. \tag{89}$$

Finally, the mean amount of flow is obtained by taking derivative of (89)

$$E[L_i] = \alpha_i\Big(\frac{E[C^2]}{2E[C]}+\frac{E[C^2]}{E[C]}\sum_{k=1}^{i-1}\rho_k+\sum_{k=1}^{i-1}m_k+\frac{\rho_i E[C^2]}{2E[C]}\Big). \tag{90}$$

Waiting Times

Computing the LST of the 'waiting time' for an arbitrary drop of fluid in queue i, $\tilde{W}_i(s)$, is calculated similarly as before

$$\tilde{W}_i(s) = E[e^{-sW_i}] = \frac{E[\int_0^C e^{-sW_i(t)}dt]}{E[C]}. \tag{91}$$

During a cycle time of length C a drop that arrives at time t ($0<t<C$) waits until
(i) the end of the current cycle, i.e. $C-t$,
(ii) the service time of all drops that arrived to queues 1 to $i-1$ during the cycle in which the tagged drop arrives, i.e. $\sum_{k=1}^{i-1}\rho_k C$,
(iii) the switchover times of the server between queues 1 to i, i.e. $\sum_{k=1}^{i-1}M_k$ and
(iv) the service times $\rho_i t$ of all drops that arrived to queue i before the tagged drop (during the cycle in which the tagged drop arrives).
Hence,

$$E\Big[\int_0^C e^{-sW_i(t)}dt\Big] = E\Big[\int_0^C e^{-s\left(C-t+\sum_{k=1}^{i-1}\rho_k C+\sum_{k=1}^{i-1}M_k+\rho_i t\right)}dt\Big]. \tag{92}$$

From (91) and (92), we get

$$\tilde{W}_i(s) = \frac{\prod_{k=1}^{i-1}\tilde{M}_k(s)\cdot\Big(\tilde{C}(s\sum_{k=1}^{i}\rho_k)-\tilde{C}\big(s(1+\sum_{k=1}^{i-1}\rho_k)\big)\Big)}{s(1-\rho_i)E[C]}. \tag{93}$$

The mean waiting time is obtained by taking derivative of (93)

$$E[W_i] = \frac{E[C^2]}{2E[C]} + \frac{E[C^2]}{E[C]} \sum_{k=1}^{i-1} \rho_k + \sum_{k=1}^{i-1} m_k + \frac{\rho_i E[C^2]}{2E[C]} = \frac{1}{\alpha_i} E[L_i]. \tag{94}$$

The expression for $E[W_i]$ can be explained as follows: The first term is the mean residual time of a cycle. The second term is the amount of work flowing into queues 1 to $i-1$ during the past and residual parts of a cycle time. The third term is the switching time associated with queues 1 to $i-1$, while the fourth term is the amount of work arriving at queue i during the past part of a cycle. Note that equation (94) is identical to equation (2.17) in Boxma et al. [5] or equation (31) in Yechiali [12].

3 Probabilistic Switching

Under probabilistic switching, after finishing a visit to a queue, the server moves to queue $i, i = 1, \ldots, N$, with probability p_i, $\sum_{i=1}^{N} p_i = 1$. It takes the server R_i units of time to get into queue i, then the server clears fluid in that queue (according to the service regime) and only then the server moves out of the queue, which takes D_i units of time. Then the server has to choose again the next queue to be served. Let t_n denote the nth choosing instant.

Define:

(i) $Y_i^{(n)}$ the fluid in queue i just before the nth choosing instant, that is $Y_i^{(n)} = L_i(t_n^-)$.

(ii) $Z_i^{(n)}$ the fluid at queue i just before the start of the nth polling instant,

$Z_i^{(n)} = Y_i^{(n)} + \alpha_i R_j \cdot 1\{\text{next queue is j}\}$, where $1\{A\}$ is the indicator of the event A. That is,

$$Z_i^{(n)} = \begin{cases} Y_i^{(n)} + \alpha_i R_1 & \text{w.p } p_1 \\ \cdots & \\ Y_i^{(n)} + \alpha_i R_i & \text{w.p } p_i \\ \cdots & \\ Y_i^{(n)} + \alpha_i R_N & \text{w.p } p_N. \end{cases} \tag{95}$$

The LST of the amount of fluid found in the system just before the nth choosing instant is

$$L_n(\theta_1, \ldots, \theta_N) = E[e^{-\sum_{j=1}^{N} \theta_j Y_j^{(n)}}]. \tag{96}$$

The LST of the amount of fluid found in the system at the nth polling instant is

$$F_n(\theta_1, \ldots, \theta_N) = E[e^{-\sum_{j=1}^{N} \theta_j Z_j^{(n)}}]. \tag{97}$$

It follows that

$$F_n(\theta_1, \ldots, \theta_N) = L_n(\theta_1, \ldots, \theta_N) \sum_{k=1}^{N} p_k \tilde{R}_k(\sum_{j=1}^{N} \theta_j \alpha_j). \tag{98}$$

We consider a stable system where $Y_i^{(n)}$ and $Z_i^{(n)}$ converge to Y_i and Z_i respectively. Assuming stability, we define the limiting LSTs as

$$L(\theta_1, \ldots, \theta_N) = \lim_{n\to\infty} L_n(\theta_1, \ldots, \theta_N), \tag{99}$$

$$F(\theta_1, \ldots, \theta_N) = \lim_{n\to\infty} F_n(\theta_1, \ldots, \theta_N). \tag{100}$$

Define $X_i^{(m)}$ the fluid in queue i just before the start of the mth polling instant to queue i. Then, for some $m \le n$, $X_i^{(m)} = Y_i^{(n)} + \alpha_i R_i$, if the nth polling instant is the mth selection of queue i. Assuming convergence of $X_i^{(m)}$ and $Y_i^{(n)}$ we write

$$X_i = Y_i + \alpha_i R_i, \tag{101}$$

and

$$\tilde{X}_i(s) = E[e^{-sX_i}] = E[e^{-sY_i}]E[e^{-s\alpha_i R_i}] = L(0,\dots,0,s,\dots,0)\cdot \tilde{R}_i(s\alpha_i). \tag{102}$$

3.1 Probabilistic Exhaustive

Recall, under the Exhaustive regime, the server clears all the fluid in a queue until it is empty and only then it leaves the queue. The total amount of fluid in queue i just before the $n+1$st choosing instant is composed of: The amount just before the nth choosing instant plus the fluid that was accumulated during (i) moving in to the nth served queue, (ii) the nth serving time and (iii) moving out from the nth served queue. If queue i is the one that was selected during the nth choosing instant the total amount of fluid in queue i is composed only of (iii). Hence, the evolution of the state of the system at the choosing instants is

$$Y_i^{(n+1)} = \begin{cases} Y_i^{(n)} + \alpha_i(R_1 + \frac{Y_1^{(n)}+\alpha_1 R_1}{\mu_1-\alpha_1} + D_1) & \text{w.p } p_1 \\ \dots & \\ \alpha_i D_i & \text{w.p } p_i \\ \dots & \\ Y_i^{(n)} + \alpha_i(R_N + \frac{Y_N^{(n)}+\alpha_N R_N}{\mu_N-\alpha_N} + D_N) & \text{w.p } p_N. \end{cases} \tag{103}$$

To calculate $L(\theta_1,\dots,\theta_N)$, we express $L_{n+1}(\theta_1,\dots,\theta_N)$ in terms of $L_n(\theta_1,\dots,\theta_N)$. This is done by conditioning on the specific queue being served during the nth service period,

$$\begin{aligned} &L_{n+1}(\theta_1,\dots,\theta_{i-1},\theta_i,\theta_{i+1},\dots,\theta_N|A_i) = \\ &L_n(\theta_1,\dots,\theta_{i-1},\frac{\sum_{j=1, j\neq i}^N \theta_j\alpha_j}{\mu_i-\alpha_i},\theta_{i+1},\dots,\theta_N)\cdot \tilde{D}_i(\sum_{j=1}^N \theta_j\alpha_j)\cdot \tilde{R}_i(\sum_{\substack{j=1\\ j\neq i}}^N \theta_j\alpha_j(\frac{1}{1-\rho_i})), \end{aligned} \tag{104}$$

where A_i is the event that queue i was polled at the nth service period.
By unconditioning on A_i and letting n approach infinity we obtain from (104)

$$\begin{aligned} &L(\theta_1,\dots,\theta_{i-1},\theta_i,\theta_{i+1},\dots,\theta_N) = \\ &p_1 L(\frac{\sum_{j=1, j\neq 1}^N \theta_j\alpha_j}{\mu_1-\alpha_1},\theta_2,\dots,\theta_N)\cdot \tilde{D}_1(\sum_{j=1}^N \theta_j\alpha_j)\cdot \tilde{R}_1(\sum_{\substack{j=1\\ j\neq 1}}^N \theta_j\alpha_j(\frac{1}{1-\rho_1})) \\ &+\dots+p_i L(\theta_1,\dots,\theta_{i-1},\frac{\sum_{j=1, j\neq i}^N \theta_j\alpha_j}{\mu_i-\alpha_i},\theta_{i+1},\dots,\theta_N)\cdot \tilde{D}_i(\sum_{j=1}^N \theta_j\alpha_j)\cdot \tilde{R}_i(\sum_{\substack{j=1\\ j\neq i}}^N \theta_j\alpha_j(\frac{1}{1-\rho_i})) \\ &+\dots+p_N L(\theta_1,\theta_2,\dots,\frac{\sum_{j=1, j\neq N}^N \theta_j\alpha_j}{\mu_N-\alpha_N})\cdot \tilde{D}_N(\sum_{j=1}^N \theta_j\alpha_j)\cdot \tilde{R}_N(\sum_{\substack{j=1\\ j\neq N}}^N \theta_j\alpha_j(\frac{1}{1-\rho_N})). \end{aligned} \tag{105}$$

Fluid Amount at Polling or Choosing Instants
First Moment
We compute from (105) the moments of the fluid amount in the system at choosing instants,

$$E[Y_i] = -\frac{\partial L(\theta_1,\ldots,\theta_N)}{\partial \theta_i}\Big|_{\underline{\theta}=\underline{0}} \qquad 1 \le i \le N. \tag{106}$$

Then we have,

$$E[Y_i] = \frac{\alpha_i(1-\rho_i)\sum_{j=1}^{N} p_j(r_j+d_j)}{p_i(1-\rho)} - \alpha_i r_i. \tag{107}$$

From (95),

$$E[Z_i] = E[Y_i] + \alpha_i \sum_{j=1}^{N} p_j r_j \tag{108}$$

and

$$E[X_i] = E[Y_i] + \alpha_i r_i = \frac{\alpha_i(1-\rho_i)\sum_{j=1}^{N} p_j(r_j+d_j)}{p_i(1-\rho)}. \tag{109}$$

For the special case where $p_j = \frac{1}{N}$ for every j, we find that (109) is equal to the equivalent expression for $f_i(i)$ in the Exhaustive service cyclic polling system (see equation (8)).

Second Moments
To find $E[Y_iY_j]$ we differentiate (105) with respect to θ_i and θ_j and then set $\underline{\theta} = \underline{0}$. This yields a set of N^2 linear equations (see Appendix B.1). In the case of identical queues, they can be solved analytically in a closed form. We have,

$$E[Y_i^2] = \frac{2N(N-1)(1-\rho_1)\alpha_1^2 m_1^2}{(1-\rho)^2} + \frac{(N-1)\alpha_1^2(r_1^{(2)}-2r_1)}{1-\rho} + \frac{N\alpha_1^2 d_1^{(2)}(1-\rho_1)}{1-\rho}, \tag{110}$$

and by using (95) and (101) we get that $E[X_i^2] = E[Z_i^2]$, and

$$E[X_i^2] = E[Z_i^2] = \frac{N(1-\rho_1)\alpha_1^2 m_1^{(2)}}{1-\rho} + \frac{2N(N-1)(1-\rho_1)\alpha_1^2 m_1^2}{(1-\rho)^2}, \tag{111}$$

$$Var[X_i] = \frac{N(1-\rho_1)\alpha_1^2(m_1^{(2)}-m_1^2)}{1-\rho} + \frac{N(N-1)(1-\rho_1)\alpha_1^2 m_1^2}{(1-\rho)^2}. \tag{112}$$

Visit Time, Intervisit Time and Cycle Time
Let S_i be defined as before. The intervisit period I_i of queue i is defined to be the period beginning at the time of its service completion and ending at the time when it is polled next. Let C_i be the cycle time for queue i. C_i consists of a service period followed by an intervisit period (or an intervisit period followed by a service period). The behavior of these periods in the probabilistic switching system is very similar to their behavior in the cyclic switching system. In both systems, a service period of queue i is followed by an intervisit period of queue i, and this is followed by another service period of queue i, etc. Hence, the following relations are valid for the probabilistic system,

$$S_i = \frac{X_i}{\mu_i - \alpha_i}, \tag{113}$$

$$X_i = \alpha_i I_i, \tag{114}$$

$$X_i = \alpha_i(1-\rho_i)C_i. \tag{115}$$

Using (113), (114) and (115) we get the LSTs and moments of these periods.

	LST	First Moment	Variance
Visit Time	$\tilde{S}_i(s) = \tilde{X}_i(\frac{s}{\mu_i-\alpha_i})$	$E[S_i] = \frac{\rho_i \sum_{j=1}^N p_j(r_j+d_j)}{p_i(1-\rho)}$	$Var[S_i] = \frac{E[X_i^2]}{(\mu_i-\alpha_i)^2} - \frac{\rho_i^2\left(\sum_{j=1}^N p_j(r_j+d_j)\right)^2}{p_i^2(1-\rho)^2}$
Intervisit Time	$\tilde{I}_i(s) = \tilde{X}_i(\frac{s}{\alpha_i})$	$E[I_i] = \frac{(1-\rho_i)\sum_{j=1}^N p_j(r_j+d_j)}{p_i(1-\rho)}$	$Var[I_i] = \frac{E[X_i^2]}{\alpha_i^2} - \frac{(1-\rho_i)^2\left(\sum_{j=1}^N p_j(r_j+d_j)\right)^2}{p_i^2(1-\rho)^2}$
Cycle Time	$\tilde{C}_i(s) = \tilde{X}_i(\frac{s}{\alpha_i(1-\rho_i)})$	$E[C_i] = \frac{\sum_{j=1}^N p_j(r_j+d_j)}{p_i(1-\rho)}$	$Var[C_i] = \frac{E[X_i^2]}{\alpha_i^2(1-\rho_i)^2} - \frac{\left(\sum_{j=1}^N p_j(r_j+d_j)\right)^2}{p_i^2(1-\rho)^2}$

In the case of identical queues we have

	First Moment	Variance
Visit Time	$E[S_i] = \frac{\rho_1 m}{1-\rho}$	$Var[S_i] = \frac{N\rho_1^2(m_1^{(2)}-m_1^2)}{(1-\rho)(1-\rho_1)} + \frac{N(N-1)\rho_1^2 m_1^2}{(1-\rho)^2(1-\rho_1)}$
Intervisit Time	$E[I_i] = \frac{(1-\rho_1)m}{1-\rho}$	$Var[I_i] = \frac{N(1-\rho_1)(m_1^{(2)}-m_1^2)}{1-\rho} + \frac{N(N-1)(1-\rho_1)m_1^2}{(1-\rho)^2}$
Cycle Time	$E[C_i] = \frac{m}{1-\rho}$	$Var[C_i] = \frac{N(m_1^{(2)}-m_1^2)}{(1-\rho)(1-\rho_1)} + \frac{N(N-1)m_1^2}{(1-\rho)^2(1-\rho_1)}$

Fluid Amount at Arbitrary Times and Waiting Times
Let L_i and W_i be defined as before. Using the same analysis as for the cyclic Exhaustive we get, where X_i replaces X_i^i,

$$E[L_i] = \frac{E[X_i^2]}{2E[X_i]}, \tag{116}$$

$$E[W_i] = \frac{E[X_i^2]}{2\alpha_i E[X_i]}. \tag{117}$$

In the case of identical queues we have,

$$E[L_i] = \alpha\Big(\frac{m_1^{(2)} - m_1^2}{2m_1} + \frac{m(1-\rho_1)}{2(1-\rho)} + \frac{m_1(N-1)}{2(1-\rho)}\Big), \tag{118}$$

$$E[W_i] = \frac{m_1^{(2)} - m_1^2}{2m_1} + \frac{m(1-\rho_1)}{2(1-\rho)} + \frac{m_1(N-1)}{2(1-\rho)}. \tag{119}$$

The difference between the mean waiting time of the probabilistic system and that of the cyclic corresponding system (equation (44)) is $\frac{m_1(N-1)}{2(1-\rho)}$ as in the result of Kleinrock and Levi [7]. We know that in the case of identical channels the expected time of a switchover duration plus a visit time equals $\frac{m_1}{1-\rho}$, and the expected number of queue changes (i.e. moving to a queue and serving it) from instant of arrival of a drop to a queue, until the server arrives at that queue is $N-1$ in the probabilistic system and $\frac{N-1}{2}$ in the cyclic system. Hence the difference is $\frac{N-1}{2} \cdot \frac{m_1}{1-\rho}$.

3.2 Probabilistic Gated

Under the Gated regime, the total amount of fluid in queue i just before the $n+1$th choosing instant is composed of: The amount just before the nth choosing instant plus the fluid that was accumulated during (i) moving into the nth served queue, (ii) the nth serving time and (iii) moving out from the nth served queue. If queue i is the one that was selected during the nth choosing instant the total amount of fluid in queue i is composed only of (ii) and (iii). Hence, the evolution of the state of the system is

$$Y_i^{(n+1)} = \begin{cases} Y_i^{(n)} + \alpha_i(R_1 + \frac{Y_1^{(n)}+\alpha_1 R_1}{\mu_1} + D_1) & \text{w.p } p_1 \\ \dots \\ \alpha_i(\frac{Y_i^{(n)}+\alpha_i R_i}{\mu_i} + D_i) & \text{w.p } p_i \\ \dots \\ Y_i^{(n)} + \alpha_i(R_N + \frac{Y_N^{(n)}+\alpha_N R_N}{\mu_N} + D_N) & \text{w.p } p_N. \end{cases} \tag{120}$$

Thus, the LST of the fluid amount at choosing instant, conditioned on queue i being served during the previous service period, is

$$L_{n+1}(\theta_1,\dots,\theta_{i-1},\theta_i,\theta_{i+1},\dots,\theta_N|A_i) = \\ L_n(\theta_1,\dots,\theta_{i-1},\frac{\sum_{j=1}^N \theta_j\alpha_j}{\mu_i},\theta_{i+1},\dots,\theta_N)\cdot\tilde{D}_i(\sum_{j=1}^N \theta_j\alpha_j)\cdot\tilde{R}_i(\sum_{\substack{j=1\\ j\neq i}}^N \theta_j\alpha_j + \rho_i\sum_{j=1}^N \theta_j\alpha_j). \tag{121}$$

By unconditioning (121), and letting n approach infinity we obtain

$$L(\theta_1,\dots,\theta_{i-1},\theta_i,\theta_{i+1},\dots,\theta_N) = \\ p_1 L(\frac{\sum_{j=1}^N \theta_j\alpha_j}{\mu_1},\theta_2,\dots,\theta_N)\cdot\tilde{D}_1(\sum_{j=1}^N \theta_j\alpha_j)\cdot\tilde{R}_1(\sum_{\substack{j=1\\ j\neq 1}}^N \theta_j\alpha_j + \rho_1\sum_{j=1}^N \theta_j\alpha_j) \\ +\dots+p_i L(\theta_1,\dots,\theta_{i-1},\frac{\sum_{j=1}^N \theta_j\alpha_j}{\mu_i},\theta_{i+1},\dots,\theta_N)\cdot\tilde{D}_i(\sum_{j=1}^N \theta_j\alpha_j)\cdot\tilde{R}_i(\sum_{\substack{j=1\\ j\neq i}}^N \theta_j\alpha_j + \rho_i\sum_{j=1}^N \theta_j\alpha_j) \\ +\dots+p_N L(\theta_1,\theta_2,\dots,\frac{\sum_{j=1}^N \theta_j\alpha_j}{\mu_N})\cdot\tilde{D}_N(\sum_{j=1}^N \theta_j\alpha_j)\cdot\tilde{R}_N(\sum_{\substack{j=1\\ j\neq N}}^N \theta_j\alpha_j + \rho_N\sum_{j=1}^N \theta_j\alpha_j). \tag{122}$$

Fluid Amount at Polling or Choosing Instants
First Moment
We obtain from (122)

$$E[Y_i] = \frac{\alpha_i\sum_{j=1}^N p_j(d_j+r_j)}{p_i(1-\rho)} - \alpha_i r_i, \tag{123}$$

and

$$E[X_i] = \frac{\alpha_i\sum_{j=1}^N p_j(d_j+r_j)}{p_i(1-\rho)} = E_{\text{Exhaustive}}[X_i]\setminus(1-\rho_i). \tag{124}$$

Note again, for the special case $p_j = \frac{1}{N}$ for each j, we find that (124) is equal to the equivalent expression for $f_i(i)$ in the Gated service cyclic polling systems (48).

Second Moments
To find $E[Y_iY_j]$ we differentiate (122) with respect to θ_i and θ_j and then let $\underline{\theta}=\underline{0}$, which yield a set of N^2 linear equations (see Appendix B.2). In the case of identical queues,

$$E[X_i^2]=\frac{N\alpha_1^2m_1^{(2)}}{(1-\rho)(1+\rho_1)}+\frac{2N(N-1)\alpha_1^2m_1^2}{(1-\rho)^2(1+\rho_1)}+\frac{2N^2\alpha_1^2m_1^2\rho_1}{(1-\rho)^2(1+\rho_1)}, \tag{125}$$

$$Var[X_i]=\frac{N\alpha_1^2(m_1^{(2)}-m_1^2)}{(1-\rho)(1+\rho_1)}+\frac{N(N-1)\alpha_1^2m_1^2}{(1-\rho)^2(1+\rho_1)}. \tag{126}$$

Note that for the probabilistic switching we get that $Var_{\text{Gated}}[X_i]=(1-\rho_1^2)Var_{\text{Exhaustive}}[X_i]$, same as the corresponding relation for the cyclic polling.

Visit Time, Intervisit Time and Cycle Time
Denoting by C_i the length of time between two consecutive polling instants to queue i, we have,

$$X_i=\alpha_iC_i, \tag{127}$$

and

$$S_i=\frac{X_i}{\mu_i}. \tag{128}$$

Using (127) and (128) we get the LSTs and moments of these periods.

	LST	First Moment	Variance
Visit Time	$\tilde{S}_i(s)=\tilde{X}_i(\frac{s}{\mu_i})$	$E[S_i]=\frac{\rho_i\sum_{j=1}^N p_j(d_j+r_j)}{p_i(1-\rho)}$	$Var[S_i]=\frac{E[X_i^2]}{\mu_i^2}-\frac{\rho_i^2\left(\sum_{j=1}^N p_j(d_j+r_j)\right)^2}{p_i^2(1-\rho)^2}$
Cycle Time	$\tilde{C}_i(s)=\tilde{X}_i(\frac{s}{\alpha_i})$	$E[C_i]=\frac{\sum_{j=1}^N p_j(d_j+r_j)}{p_i(1-\rho)}$	$Var[C_i]=\frac{E[X_i^2]}{\alpha_i^2}-\frac{\left(\sum_{j=1}^N p_j(d_j+r_j)\right)^2}{p_i^2(1-\rho)^2}$

In the case of identical queues we have

	First Moment	Variance
Visit Time	$E[S_i]=\frac{N\rho_1m_1}{1-\rho}$	$Var[S_i]=\frac{N\rho_1^2(m_1^{(2)}-m_1^2)}{(1-\rho)(1+\rho_1)}+\frac{N(N-1)\rho_1^2m_1^2}{(1-\rho)^2(1+\rho_1)}$
Cycle Time	$E[C_i]=\frac{m}{1-\rho}$	$Var[C_i]=\frac{N(m_1^{(2)}-m_1^2)}{(1-\rho)(1+\rho_1)}+\frac{N(N-1)m_1^2}{(1-\rho)^2(1+\rho_1)}$

As for the intervisit period, its mean is $E[I_i]=E[C_i]-E[S_i]$.
Fluid Amount at Arbitrary Times and Waiting Times
The means of L_i and W_i have the same expressions as for the cyclic Gated ((64) and (69)), where X_i replaces X_i^i,

$$E[L_i]=\frac{(1+\rho_i)E[X_i^2]}{2E[X_i]}, \tag{129}$$

$$E[W_i]=\frac{1+\rho_i}{\alpha_i}\cdot\frac{E[X_i^2]}{2E[X_i]}. \tag{130}$$

In the case of identical queues we have,

$$E[L_i]=\alpha_1\Big(\frac{m_1^{(2)}-m_1^2}{2m_1}+\frac{m(1+\rho_1)}{2(1-\rho)}+\frac{m_1(N-1)}{2(1-\rho)}\Big), \tag{131}$$

$$E[W_i] = \frac{m_1^{(2)} - m_1^2}{2m_1} + \frac{m(1+\rho_1)}{2(1-\rho)} + \frac{m_1(N-1)}{2(1-\rho)}. \quad (132)$$

Again, the difference in $E[W_i]$ between the probabilistic and cyclic polling scheme is the last term in equation (132).

4 Workloads

In this section we compute the mean value of the total amount of work (workload) in the system. Let $\hat{V}$ be the total amount of workload in our system (cyclic or probabilistic). We use a method similar to the one used by Boxma and Groenendijk [2] to compute the exact value of $E[\hat{V}]$. It has been shown [2] that in a M/G/1 polling system,

$$\hat{V} \triangleq V + Y, \quad (133)$$

where $\hat{V}$ is the steady-state amount of work in a system with switchover times, Y is the steady-state amount of work in that system at an arbitrary switchover epoch, and V is the steady-state amount of work in the corresponding system without switchover times ($\triangleq$ stands for equality in distribution). In our fluid system one would expect that $V = 0$ meaning that there is no waiting in a corresponding system with no switchover times. We show that in the fluid system

$$\hat{V} \triangleq Y. \quad (134)$$

Figure 3 shows the amount of work in front of the server as a function of time. During a

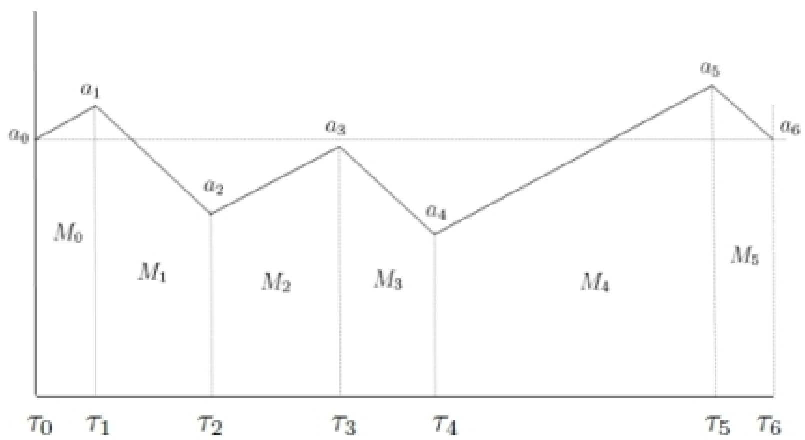

Fig. 3 Amount of work in the system.

switchover period (idle-time) it constantly increases at a rate of ρ, while during a service period it constantly decreases at a rate of $1-\rho$. This follows since the rate of work flowing into queue i is ρ_i, while the total rate of work flowing into the entire system is ρ. When busy, the server works at a rate 1 so that the net total amount of work decreases at a rate of $1-\rho$.

Lemma 1 *Consider a positive continuous piecewise linear function $f(t)$ with up slopes $\alpha > 0$ and down slopes $\beta < 0$ (as shown in Figure 3). Looking over a finite interval, τ_0 to τ_m ($\tau_m > \tau_0$), where $f(\tau_0) = f(\tau_m)$, the average value of the function $f(t)$ at points where $f'(t) = \alpha$ (up slope) equals the average value of the function $f(t)$ at point where $f'(t) = \beta$ (down slope).*

Proof:
The proof is done by geometrical considerations. Let $f(\tau_0) = f(\tau_m) = a_0$ and let $\tau_1, \tau_2, \ldots, \tau_{m-1}$ be the slope changing points; let $f(\tau_j) = a_j$ ($1 \le j \le m-1$). We assume that m is even (when m is odd a similar proof applies). Let M_k be the area bounded by $a_k, a_{k+1}, \tau_k, \tau_{k+1}$. Then, for M_ks where $a_k < a_{k+1}$,

$$M_k = \frac{a_{k+1}^2 - a_k^2}{2\alpha}. \tag{135}$$

Then the total area under all $\alpha > 0$ slopes is

$$M_0 + M_2 + \cdots + M_{m-2} = \frac{a_1^2 - a_0^2 + a_3^2 - a_2^2 + \cdots + a_{m-1}^2 - a_{m-2}^2}{2\alpha}, \tag{136}$$

while its length is

$$(\tau_1 - \tau_0) + (\tau_3 - \tau_2) + \cdots + (\tau_{m-1} - \tau_{m-2}) = \frac{a_1 - a_0 + a_3 - a_2 + \cdots + a_{m-1} - a_{m-2}}{\alpha}. \tag{137}$$

Hence,

$$\begin{aligned} \frac{\int_{\tau_0}^{\tau_m} f(t) 1\{f'(t^+) = \alpha\} dt}{\int_{\tau_0}^{\tau_m} 1\{f'(t^+) = \alpha\} dt} &= \frac{M_0 + M_2 + \cdots + M_{m-2}}{(\tau_1 - \tau_0) + (\tau_3 - \tau_2) + \cdots + (\tau_{m-1} - \tau_{m-2})} \\ &= \frac{a_1^2 - a_0^2 + a_3^2 - a_2^2 + \cdots + a_{n-1}^2 - a_{n-2}^2}{2(a_1 - a_0 + a_3 - a_2 + \cdots + a_{m-1} - a_{m-2})}. \end{aligned} \tag{138}$$

Using the same consideration and taking into account that $a_m = a_0$ we have that

$$\frac{\int_{\tau_0}^{\tau_m} f(t) 1\{f'(t^+) = \alpha\} dt}{\int_{\tau_0}^{\tau_m} 1\{f'(t^+) = \alpha\} dt} = \frac{\int_{\tau_0}^{\tau_m} f(t) 1\{f'(t^+) = \beta\} dt}{\int_{\tau_0}^{\tau_m} 1\{f'(t^+) = \beta\} dt} = \frac{\int_{\tau_0}^{\tau_m} f(t) dt}{\tau_m - \tau_0}. \tag{139}$$

Now we prove the following:

Theorem 1 *Suppose our system is ergodic and stationary. Then the expected amount of work in the system at an arbitrary epoch, $\hat{V}$, is equal to the expected amount of work in the system at an arbitrary epoch in a switching interval, i.e. $E[\hat{V}] = E[Y]$.*

Proof:
Let $\hat{V}(t)$ be the amount of work in the system at time t ($t > 0$). Since the system is ergodic and stationary then

$$\frac{\int_0^T \hat{V}(t) dt}{T} \to E[\hat{V}] \qquad \text{as } T \to \infty, \tag{140}$$

and

$$\frac{\int_0^T \hat{V}(t) 1\{t \text{ is within switching period}\} dt}{\int_0^T 1\{t \text{ is within switching period}\} dt} \to E[Y] \qquad \text{as } T \to \infty. \tag{141}$$

Assume, without loss of generality, that the process starts with $\hat{V}(0) = E[Y]$. Since $E[Y]$ is the expected amount of work during a switching interval, there exists an unbounded subsequence $t_0, t_1, \ldots$ where $\hat{V}(t_n) = E[Y]$. The convergence in (140) and (141) holds for the

above unbounded subsequence. Using the lemma we know that in those points equations (140) and (141) converge to the same value, hence, $E[\hat{V}] = E[Y]$. □
Arguing similarly to the lemma, it can be shown that

$$\frac{\int_{\tau_0}^{\tau_m} e^{-sf(t)} 1\{f'(t^+) = \alpha\} dt}{\int_{\tau_0}^{\tau_m} 1\{f'(t^+) = \alpha\} dt} = \frac{\int_{\tau_0}^{\tau_m} e^{-sf(t)} dt}{\tau_m - \tau_0}, \tag{142}$$

by noticing that the corresponding area of M_k which is bounded by $\tau_k, \tau_{k+1}, e^{-sa_k}, e^{-sa_{k+1}}$ is equal to M_k^* where,

$$M_k^* = \int_0^{\tau_{k+1}-\tau_k} e^{-s(a_k+\alpha t)} dt = -\frac{1}{s\alpha}\Big(e^{-sa_{k+1}} - e^{-sa_k}\Big). \tag{143}$$

Using that and by arguing as in Theorem 1, it can be shown that $E[e^{-s\hat{V}}] = E[e^{-sY}]$, hence $\hat{V} \overset{\circ}{=} Y$.

Cyclic Polling
Computing $E[Y]$ for the fluid system is similar to the computation for the discrete-arrival M/G/1 polling systems. In the cyclic polling, Boxma and Groenendijk [2] showed that for an arbitrary cyclic polling system with mixed service regimes

$$E[Y] = \rho \frac{m^{(2)}}{2m} + \frac{m}{2(1-\rho)}\big(\rho^2 - \sum_{i=1}^N \rho_i^2\big) + \sum_{i=1}^N E[M_i^{(1)}], \tag{144}$$

where $E[M_i^{(1)}]$ is the expected unfinished work at the ith queue at an instant of departure of the server from that queue. Equation (144) holds for our fluid cyclic systems as well, and $E[M_i^{(1)}]$ depends only on the service discipline in queue i. For the Exhaustive regime $E[M_i^{(1)}] = 0$ for every i, so that

$$E[\hat{V}] = \sum_{i=1}^N \rho_i E[W_i] = \rho \frac{m^{(2)}}{2m} + \frac{m}{2(1-\rho)}\big(\rho^2 - \sum_{i=1}^N \rho_i^2\big). \tag{145}$$

For the Gated regime, $E[M_i^{(1)}] = \rho_i E[S_i]$, hence

$$\sum_{i=1}^N \rho_i E[W_i] = \rho \frac{m^{(2)}}{2m} + \frac{m}{2(1-\rho)}\big(\rho^2 + \sum_{i=1}^N \rho_i^2\big). \tag{146}$$

For the Globally-Gated,

$$E[M_i^{(1)}] = \rho_i \Big(\sum_{j=1}^{i-1} \big(E[S_j] + m_j\big) + E[S_i]\Big), \tag{147}$$

hence

$$\sum_{i=1}^N \rho_i E[W_i] = \rho \frac{m^{(2)}}{2m} + \frac{m}{1-\rho}\rho^2 + \sum_{j=2}^N \rho_j \sum_{i=1}^{j-1} m_i. \tag{148}$$

Probabilistic Polling
Regarding probabilistic polling with mixed service regimes (Exhaustive or Gated), Boxma and Weststrate [3] showed that

$$E[Y] = \frac{\rho}{2\sigma}\sum_{i=1}^{N} q_i \sum_{j=1}^{N} p_{ij} s_{ij}^{(2)} + \frac{1}{\sigma}\sum_{i=1}^{N} q_i \sum_{j=1}^{N} p_{ij} s_{ij} \sum_{\substack{k=1\\k\neq i}}^{N} \rho_k E[T_{ki}] + \frac{\sigma}{1-\rho}\sum_{k\in \text{gated}} \frac{\rho_k^2}{q_k}, \tag{149}$$

where (i) p_{ij} is the stationary transition probability

$$p_{ij} = P(\text{the n+1}st \text{ served queue is } j | \text{the n}th \text{ served queue was } i), \tag{150}$$

where in our case $p_{ij} = p_j$,
(ii) q_i is the limiting stationary distribution,

$$q_i = lim_{n\to\infty} P(\text{the n}th \text{ served queue is } i), \tag{151}$$

where in our case $q_i = p_i$,
(iii) S_{ij} is the switchover time from queue i to queue j, where in our case $s_{ij} = d_i + r_j$, $s_{ij}^{(2)} = d_i^{(2)} + 2d_i r_j + r_j^{(2)}$,
(iv) $\sigma = \sum_{i=1}^{N} q_i \sum_{j=1}^{N} p_{ij} s_{ij}$ where in our case $\sigma = \sum_{i=1}^{N} p_i m_i$,
and (v) T_{ki} is the time between a departure of the server from queue i and the last previous departure from queue k. In our case

$$E[T_{ki}] = f(i) + \sum_{\substack{j=1\\j\neq i}}^{N} f(j)\frac{p_j}{p_k} = \frac{\rho_i \sigma}{p_i(1-\rho)} + \frac{(\rho-\rho_k)\sigma}{p_k(1-\rho)} + \frac{r_i}{p_k} + \frac{\sum_{i=1}^{N} p_i d_i}{p_k} \tag{152}$$

where $f(i) = \sum_{j=1}^{N} p_j d_j + r_i + E[S_i]$ (see [3]). When $\forall j$, $r_j = r_1$ and $d_j = d_1$ equation (152) reduces to

$$E[T_{ki}] = \frac{m_1}{1-\rho}\left(\frac{\rho_i}{p_i} - \frac{\rho_k}{p_k} + \frac{1}{p_k}\right). \tag{153}$$

Finally we get that for the Gated regime equation (149) reduces to

$$\sum_{i=1}^{N} \rho_i E[W_i] = \frac{m_1}{1-\rho}\sum_{k=1}^{N}\frac{\rho_k}{p_k} - \rho m_1 + \frac{\rho m_1^{(2)}}{2m_1}, \tag{154}$$

and for the Exhaustive regime,

$$\sum_{i=1}^{N} \rho_i E[W_i] = -\frac{m_1}{1-\rho}\sum_{k=1}^{N}\frac{\rho_k^2}{p_k} + \frac{m_1}{1-\rho}\sum_{k=1}^{N}\frac{\rho_k}{p_k} - \rho m_1 + \frac{\rho m_1^{(2)}}{2m_1}. \tag{155}$$

5 Optimization

Optimal server's switching procedures have been discussed by various authors (see e.g. [4], [6], [12] and [13]). In our study a question that arises is how to choose 'best' p_i's for the Probabilistic Switching. We consider three objective functions: (i) minimizing $\sum_{j=1}^{N} E[X_j]$; (ii) minimizing $\max_j\{E[X_j]\}$; and (iii) minimizing the mean workload in the system. For the Gated regime and objective (i) we wish to minimize

$$\sum_{j=1}^{N} E[X_j] = \sum_{j=1}^{N} \frac{\alpha_j \sum_{k=1}^{N} p_k(d_k + r_k)}{p_j(1-\rho)} \tag{156}$$

subject to the constraint

$$\sum_{j=1}^{N} p_j = 1. \tag{157}$$

The general solution of (156) and (157) leads to a set of equations in the values of p_j's, that can be solved numerically. For the special case where for $\forall j$, $r_j = r_1$ and $d_j = d_1$, equation (156) reduces to

$$\sum_{j=1}^{N} E[X_j] = \frac{(d_1 + r_1)}{(1-\rho)} \sum_{j=1}^{N} \frac{\alpha_j}{p_j}. \tag{158}$$

By using Lagrange Multipliers we get the optimal value of p_j, denoted p_j^* as

$$p_j^* = \frac{\sqrt{\alpha_j}}{\sum_{k=1}^{N} \sqrt{\alpha_k}}. \tag{159}$$

For the Exhaustive regime, objective (i) function gets the value

$$\sum_{j=1}^{N} E[X_j] = \sum_{j=1}^{N} \frac{\alpha_j(1-\rho_j)\sum_{k=1}^{N} p_k(r_k + d_k)}{p_j(1-\rho)}. \tag{160}$$

If $\forall j$, $r_j = r_1$ and $d_j = d_1$ then

$$\sum_{j=1}^{N} E[X_j] = \frac{(d_1 + r_1)}{(1-\rho)} \sum_{j=1}^{N} \frac{\alpha_j(1-\rho_j)}{p_j}, \tag{161}$$

and we have

$$p_j^* = \frac{\sqrt{\alpha_j(1-\rho_j)}}{\sum_{k=1}^{N} \sqrt{\alpha_k(1-\rho_k)}}. \tag{162}$$

Objective (ii) function leads to all $E[X_j]$ being equal. Then, for the Gated regime, we have $\frac{\alpha_i}{p_i^*} = \frac{\alpha_j}{p_j^*} =$ constant. Hence, for the non-symmetric Gated regime,

$$p_j^* = \frac{\alpha_j}{\sum_{k=1}^{N} \alpha_k}. \tag{163}$$

Similarly, for the Exhaustive regime, $\max_j\{E[X_j]\}$ is minimized when

$$p_j^* = \frac{\alpha_j(1-\rho_j)}{\sum_{k=1}^{N} \alpha_k(1-\rho_k)}. \tag{164}$$

Regarding objective (iii), when for $\forall j$, $r_j = r_1$ and $d_j = d_1$, minimizing (154) for the Gated regime we get

$$p_j^* = \frac{\sqrt{\rho_j}}{\sum_{k=1}^{N} \sqrt{\rho_k}}, \tag{165}$$

while minimizing (155) for the Exhaustive regime leads to

$$p_j^* = \frac{\sqrt{\rho_j(1-\rho_j)}}{\sum_{k=1}^{N} \sqrt{\rho_k(1-\rho_k)}}. \tag{166}$$

That is, for the fluid polling systems we obtain the same optimization results as in [4] for the classical discrete-arrival M/G/1-type polling models.

6 Summary

This paper studies the fluid analog of discrete-arrival, M/G/1-type polling systems. We show that the mean amount of fluid in polling instant of a queue is the same in both models. However, the corresponding second moments differ and the second moment of the service time in the discrete system vanishes. Moreover, $E[L_i]$, the mean amount of fluid in queue i at an arbitrary moment, as well as $E[W_i]$, the mean waiting time of a drop, are smaller in the fluid models. Finally, the distribution of the workload in the system is the same at visit period and during switching duration and hence it represents the workload at any arbitrary instant.

References

1. Armony, R., Yechiali, U.: Polling systems with permanent and transient customers. Stochastic Models, Vol. 15, No. 3, pp. 395-427 (1999)
2. Boxma, O.J., Groenendijk, W.P.: Pseudo-Conservation law in cyclic-service systems. J. Appl. Prob. 24, pp. 949-964 (1987)
3. Boxma, O.J., Weststrate, J.A.: Waiting times in polling systems with markovian server routing. In: G. Stiege and J.S lie (eds.), Messung, Modellierung und Bewertung von Rechensystemen und Netzen, pp. 89-104 Springer-Verlag, Berlin (1989)
4. Boxma, O.J., Levi, H., Weststrate, J.A.: Efficient visit orders for polling systems. Report Centre for Mathematics and Computer Science, Amsterdam (1990)
5. Boxma, O.J., Levi, H., Yechiali, U.: Cyclic reservation schemes for efficient operation of multiple-queue single-server systems. Annals of Operation Research 35, pp. 187-208 (1992)
6. Browne, S., Yechiali, U.: Dynamic priority rules for cyclic-type queues. Advances in Applied Probabilty. Vol 21, No.2, pp. 432-450 (1990)
7. Kleinrock, L., Levi, H.: The analysis of random polling systems. Operations Research, Vol. 36, No. 5, pp. 716-732 (1988)
8. Kofman, D., Yechiali, U.: Polling systems with station breakdowns. Performance Evaluation, Vols. 27 & 28, pp. 647-672 (1996)
9. Levi, Y., Yechiali, U.: Utilization of idle time in an M/G/1 Queuing System. Management Science, Vol. 22, No. 2, pp. 202-211 (1975)
10. Takagi, H.: Analysis of polling systems. MIT Press, Cambridge, Mass (1986)
11. Takagi, H.: Queueing analysis of polling models: an update. Stochastic Analysis of Computer and Communication Systems, pp. 267-318. North Holland (1990)
12. Yechiali, U.: Analysis and control of polling systems. In: Performance Evaluation of Computer and Communication System, pp. 630-650. Springer-Verlag (1993)
13. Wal, J. v.d., Yechiali, U.: Dynamic visit-order rules for batch-service polling. Probability in the Engineering and Informational Sciences, Vol. 17, No. 3, pp. 351-367 (2003)

A Appendices - Second Moments of Fluid Amount at Polling Instant

A.1 Cyclic Exhaustive

$$\begin{aligned} f_{i+1}(j,k) &= \alpha_j\alpha_k m_i^{(2)} + \alpha_k m_i(f_i(j) + f_i(i)\frac{\alpha_j}{\mu_i-\alpha_i}) + \alpha_j m_i(f_i(k) + f_i(i)\frac{\alpha_k}{\mu_i-\alpha_i}) \\ &+ f_i(j,k) + \frac{\alpha_j}{\mu_i-\alpha_i}f_i(i,k) + \frac{\alpha_k}{\mu_i-\alpha_i}f_i(i,j) + \frac{\alpha_j\alpha_k}{(\mu_i-\alpha_i)^2}f_i(i,i) \qquad i\neq j\ i\neq k, \end{aligned} \tag{167}$$

$$f_{i+1}(i,k) = \alpha_i\alpha_k m_i^{(2)} + \alpha_i m_i(f_i(k) + f_i(i)\frac{\alpha_k}{\mu_i-\alpha_i}) \qquad i\neq k, \tag{168}$$

$$f_{i+1}(i,i) = \alpha_i^2 m_i^{(2)}. \tag{169}$$

A.2 Cyclic Gated

$$\begin{aligned} f_{i+1}(j,k) &= \alpha_j\alpha_k m_i^{(2)} + \alpha_j m_i(f_i(k) + f_i(i)\frac{\alpha_k}{\mu_i}) + \alpha_k m_i(f_i(j) + f_i(i)\frac{\alpha_j}{\mu_i}) \\ &+ f_i(j,k) + f_i(i,j)\frac{\alpha_k}{\mu_i} + f_i(i,k)\frac{\alpha_j}{\mu_i} + f_i(i,i)\frac{\alpha_j\alpha_k}{\mu_i^2} \qquad k,j\neq i \end{aligned} \tag{170}$$

$$\begin{aligned} f_{i+1}(i,k) &= \alpha_i\alpha_k m_i^{(2)} + \alpha_i m_i(f_i(k) + \frac{\alpha_k}{\mu_i}f_i(i)) + \alpha_k\rho_i m_i f_i(i) \\ &+ \rho_i(\frac{\alpha_k}{\mu_i}f_i(i,i) + f_i(i,k)) \qquad k\neq i \end{aligned} \tag{171}$$

$$f_{i+1}(i,i) = \alpha_i^2 m_i^{(2)} + 2\alpha_i\rho_i m_i f_i(i) + \rho_i^2 f_i(i,i). \tag{172}$$

B Appendices - Second Moments of Fluid Amount at Choosing Instant

B.1 Probabilistic Exhaustive

$$\begin{aligned} \frac{\partial^2 L}{\partial\theta_i\partial\theta_i}\Big|_{\underline{\theta}=\underline{0}} = E[Y_i^2] = \sum_{\substack{j=1\\ j\neq i}}^{N} p_j \Bigg(& \alpha_i d_j(\frac{\alpha_i}{\mu_j-\alpha_j}E[Y_j] + E[Y_i]) + \alpha_i\frac{1}{1-\rho_j}r_j(\frac{\alpha_i}{\mu_j-\alpha_j}E[Y_j] + E[Y_i]) + \\ & \frac{2\alpha_i}{\mu_j-\alpha_j}E[Y_jY_i] + (\frac{\alpha_i}{\mu_j-\alpha_j})^2 E[Y_j^2] + E[Y_iY_i] + \\ & \alpha_i^2 d_j^{(2)} + \alpha_i^2\frac{1}{1-\rho_j}d_j r_j + \alpha_i d_j(\frac{\alpha_i}{\mu_j-\alpha_j}E[Y_j] + E[Y_i]) + \\ & \alpha_i^2(\frac{1}{1-\rho_j})^2 r_j^{(2)} + \alpha_i^2\frac{1}{1-\rho_j}d_j r_j + \alpha_i\frac{1}{1-\rho_j}r_j(\frac{\alpha_i}{\mu_j-\alpha_j}E[Y_j] + E[Y_i])\Bigg) + p_i\alpha_i^2 d_i^{(2)}, \end{aligned} \tag{173}$$

$$\frac{\partial^2 L}{\partial\theta_i\partial\theta_j}\Big|_{\underline{\theta}=\underline{0}} = E[Y_iY_j] = \sum_{\substack{k=1\\k\neq i\\k\neq j}}^{N} p_k\Bigg(\alpha_j d_k(\frac{\alpha_i}{\mu_k-\alpha_k}E[Y_k]+E[Y_i]) + \alpha_j\frac{1}{1-\rho_k}r_k(\frac{\alpha_i}{\mu_k-\alpha_k}E[Y_k]+E[Y_i])+$$
$$\frac{\alpha_i}{\mu_k-\alpha_k}E[Y_kY_j] + \frac{\alpha_i\alpha_j}{(\mu_k-\alpha_k)^2}E[Y_k^2] + E[Y_iY_j] + \frac{\alpha_j}{\mu_k-\alpha_k}E[Y_kY_i]+$$
$$\alpha_i\alpha_j d_k^{(2)} + \alpha_i\alpha_j\frac{1}{1-\rho_k}d_kr_k + \alpha_i d_k(\frac{\alpha_j}{\mu_k-\alpha_k}E[Y_k]+E[Y_j])+$$
$$\alpha_i\alpha_j(\frac{1}{1-\rho_k})^2 r_k^{(2)} + \alpha_i\alpha_j\frac{1}{1-\rho_k}d_kr_k + \alpha_i\frac{1}{1-\rho_k}r_k(\frac{\alpha_j}{\mu_k-\alpha_k}E[Y_k]+E[Y_j])\Bigg)+$$
$$p_i\Big(\alpha_i\alpha_j d_i^{(2)} + \alpha_i\alpha_j\frac{1}{1-\rho_i}d_ir_i + \alpha_i d_i(\frac{\alpha_j}{\mu_i-\alpha_i}E[Y_i]+E[Y_j])\Big)+$$
$$p_j\Big(\alpha_i\alpha_j d_j^{(2)} + \alpha_i\alpha_j\frac{1}{1-\rho_j}d_jr_j + \alpha_j d_j(\frac{\alpha_i}{\mu_j-\alpha_j}E[Y_j]+E[Y_i])\Big), \qquad i\neq j \tag{174}$$

B.2 Probabilistic Gated

$$\frac{\partial^2 L}{\partial\theta_i\partial\theta_i}\Big|_{\underline{\theta}=\underline{0}} = E[Y_i^2] = \sum_{\substack{j=1\\j\neq i}}^{N} p_j\Bigg(2\alpha_i d_j(\frac{\alpha_i}{\mu_j}E[Y_j]+E[Y_i]) + 2\alpha_i(1+\rho_i)r_j(\frac{\alpha_i}{\mu_j}E[Y_j]+E[Y_i]) + \frac{\alpha_i^2}{\mu_j^2}E[Y_jY_j]+$$
$$\frac{2\alpha_i}{\mu_j}E[Y_jY_i] + E[Y_iY_i] + \alpha_i^2 d_j^{(2)} + 2\alpha_i^2(1+\rho_j)d_jr_j + \alpha_i^2(1+\rho_j)^2 r_j^{(2)}\Bigg)+$$
$$p_i\Big(2\alpha_i\rho_i d_iE[Y_i] + \rho_i^2E[Y_iY_i] + 2\alpha_i\rho_i^2 r_iE[Y_i] + \alpha_i^2 d_i^{(2)} + 2\alpha_i^2\rho_i d_ir_i + \rho_i^2\alpha_i^2 r_i^{(2)}\Big) \tag{175}$$

$$\frac{\partial^2 L}{\partial\theta_i\partial\theta_j}\Big|_{\underline{\theta}=\underline{0}} = E[Y_iY_j] = \sum_{\substack{k=1\\k\neq i\\k\neq j}}^{N} p_k\Bigg(\alpha_j d_k(\frac{\alpha_i}{\mu_k}E[Y_k]+E[Y_i]) + \alpha_j(1+\rho_k)r_k(\frac{\alpha_i}{\mu_k}E[Y_k]+E[Y_i])+$$
$$\frac{\alpha_i\alpha_j}{\mu_k^2}E[Y_kY_k] + \frac{\alpha_i}{\mu_k}E[Y_kY_j] + \frac{\alpha_j}{\mu_k}E[Y_kY_i] + E[Y_iY_j] + \alpha_i\alpha_j d_k^{(2)} + 2\alpha_i\alpha_j(1+\rho_k)d_kr_k+$$
$$\alpha_i d_k(\frac{\alpha_j}{\mu_k}E[Y_k]+E[Y_j]) + \alpha_i\alpha_j(1+\rho_k)^2 r_k^{(2)} + \alpha_i(1+\rho_k)r_k(\frac{\alpha_j}{\mu_k}E[Y_k]+E[Y_j])\Bigg)+$$
$$p_i\Big(\alpha_j\rho_i d_iE[Y_i] + \alpha_j(1+\rho_i)\rho_i r_iE[Y_i] + \rho_i(\frac{\alpha_j}{\mu_i}E[Y_iY_i]+E[Y_iY_j]) + \alpha_i\alpha_j d_i^{(2)}+$$
$$\alpha_i\alpha_j(1+\rho_i)d_ir_i + \alpha_i d_i(\frac{\alpha_j}{\mu_i}E[Y_i]+E[Y_j]) + \rho_i\alpha_i\alpha_j(1+\rho_i)r_i^{(2)} + \rho_i\alpha_i\alpha_j d_ir_i+$$
$$\rho_i\alpha_i r_i(\frac{\alpha_j}{\mu_i}E[Y_i]+E[Y_j])\Big) + p_j\Big(\alpha_i\rho_j d_jE[Y_j] + \alpha_i(1+\rho_j)\rho_j r_jE[Y_j] + \rho_j(\frac{\alpha_i}{\mu_j}E[Y_jY_j]+$$
$$E[Y_jY_i]) + \alpha_j\alpha_i d_j^{(2)} + \alpha_j\alpha_i(1+\rho_j)d_jr_j + \alpha_j d_j(\frac{\alpha_i}{\mu_j}E[Y_j]+E[Y_i]) + \rho_j\alpha_j\alpha_i(1+\rho_j)r_j^{(2)}+$$
$$\rho_j\alpha_j\alpha_i d_jr_j + \rho_j\alpha_j r_j(\frac{\alpha_i}{\mu_j}E[Y_j]+E[Y_i])\Big) \qquad i\neq j \tag{176}$$