

**WAITING TIME VARIANCES IN  
CONTINUOUS-TIME POLLING SYSTEMS**

Mandyam M. Srinivasan  
Department of Industrial & Operations Engineering  
University of Michigan  
Ann Arbor, MI 48109-2117

Technical Report 91-37  
December 1991

# Waiting Time Variances in Continuous-Time Polling Systems

Mandyam M. Srinivasan

Department of  
Industrial and Operations Engineering  
The University of Michigan  
Ann Arbor, MI 48109-2117

Technical Report 91-37  
December 1991

## Abstract

We present an efficient technique to compute waiting time variances in polling systems that use the exhaustive and gated service disciplines. The polling systems can have a mix of service disciplines, wherein some of the nodes can use the exhaustive service discipline while the other nodes use the gated service discipline. The technique allows the user to obtain the waiting time variances at select nodes without having to obtain waiting times variances at all nodes simultaneously. Once the waiting time variances have been evaluated for a given configuration, then the IS technique allows the user to rapidly evaluate several alternate configurations having different switchover time parameters, with just a few elementary arithmetic operations.

Keywords:

Cyclic Queues; Polling Systems; Algorithms; Waiting Time Variances

## 1. Introduction

A polling system is a queueing system in which a single server services  $M$  nodes. The server cycles around the system, visiting each node in turn, and attends to the waiting customers using some prespecified service discipline. Real-world applications which are modeled and analyzed using polling systems can be found, for example, in computer communication and material handling systems, and there is a vast body of literature devoted to the analysis of polling systems. An excellent survey of the research done on polling systems can be found in Takagi [10].

A number of service disciplines have been studied, which determine the service strategy adopted by the server at a node. Two of the most commonly studied service disciplines are i) the exhaustive service discipline, where the server continues to serve all customers at the node until there are no more customers at the node, ii) the gated service discipline, where the server only serves those customers present at the node at the instant the node is polled.

A number of techniques that have been developed for computing the mean waiting times in polling systems that use the exhaustive and gated service disciplines. However, there is very little work on computing the variances of these waiting times. To our knowledge, there are two papers by Aminetzah [1] and Ferguson [5] for computing waiting time variances. These algorithms are based on solving a set of  $O(M^3)$  equations. As observed in Takagi [10], these equations are quite complicated and require the use of a symbolic formula manipulation language. No computational results are presented.

In a recent paper [8], we developed a powerful technique, termed the Individual Station (IS) technique, to compute mean waiting times for continuous-time polling systems using the exhaustive and/or gated service disciplines. In this paper, we further develop the IS technique and present an iterative algorithm to compute the waiting time variances for these systems. Although the iterative algorithm apparently requires the computation of an infinite sequence of terms, we can show that the number of iterations it requires is logarithmic in the accuracy required, using arguments similar to those presented by Levy [7].

The IS technique allows us to analyze systems with mixed service disciplines, wherein some nodes are served using the exhaustive service discipline while the other nodes are served using the gated service discipline. The technique also allows the user to determine the waiting time variances at one or more select nodes without having to obtain the waiting time variances at all the nodes, simultaneously. Furthermore, once waiting time variances are evaluated for a configuration, then the user can rapidly evaluate several alternate configurations that have different values for the switchover time parameters, with just a few elementary arithmetic operations.

The notation involved in the intermediate stages during the development of the algorithm may seem a little complex. However, the reader will find that very little notation is involved in the final expressions for the mean and variance of waiting times, and that the resulting algorithm is very simple and easily implemented. We provide computational results to demonstrate the efficiency of the algorithm.

## 2. The Waiting Time Moments

The waiting time at node  $m$  is the time spent by an arbitrary customer, waiting for service, and is denoted by the random variable  $W_m$ . Customers are served according to the first-come-first-served rule. In this section, we derive expressions for the first and second moments of  $W_m$ , denoted by  $E(W_m)$  and  $E(W_m^2)$ . This paper is concerned with computing the second moment/variance of the waiting time. Algorithms for computing the first moment of the waiting time have already been presented in an earlier paper [8].

We first present the notation. Unless otherwise specified, it is implicit that the index for any summation is over the range 1 through  $M$ . It is also implicit that the index is i) reset to 1 if it becomes  $M+1$ , and ii) reset to  $M$  if it drops down to 0. We adopt the convention that an empty product equals 1.

A single server serves requests from customers at the  $M$  nodes. Customers arrive at each node according to independent Poisson processes with rate  $\lambda_m$  at node  $m$ . The service time,  $B_m$ , of a customer at node  $m$  is an independent random variable, with its  $i^{\text{th}}$  moment denoted by  $E(B_m^i)$ . The Laplace Stieltjes Transform (LST) of the service time distribution is denoted by  $\beta_m(\cdot)$ . The busy period,  $H_m$ , induced by a single customer at node  $m$  has its  $i^{\text{th}}$  moment denoted by  $E(H_m^i)$ , and its LST denoted by  $\eta_m(\cdot)$ . The first three moments of  $H_m$  are (Kleinrock [6]):

$$E(H_m) = \frac{E(B_m)}{1-\rho_m}, \quad E(H_m^2) = \frac{E(B_m^2)}{(1-\rho_m)^3}, \quad E(H_m^3) = \frac{E(B_m^3)}{(1-\rho_m)^4} + 3 \frac{\lambda_m(E(B_m^2))^2}{(1-\rho_m)^5}. \quad (3.1)$$

The time taken by the server to switch from node  $m$  to node  $m+1$  is an independent random variable,  $S_m$ , with its  $i^{\text{th}}$  moment denoted by  $E(S_m^i)$ . The variance and LST of  $S_m$  are denoted by  $\text{Var}(S_m)$ , and  $\sigma_m(\cdot)$ , respectively. The traffic intensity at node  $m$  is denoted by  $\rho_m = \lambda_m E(B_m)$ , and  $\rho = \sum_m \rho_m$  denotes the server utilization. For the polling system to be stable, we require  $\rho$  to be less than one. If the polling system is stable, then it can be shown that the cycle time,  $\bar{C}$ , which is the expected time between two successive visits by the server to a node is the same for all nodes, and is given by  $\bar{C} = \sum_m E(S_m)/(1-\rho)$ . We are interested in obtaining the waiting time variance,  $\text{Var}(W_m)$ , at node  $m$ , for  $m=1, \dots, M$ . Let  $\mathcal{E}$  denote the set of nodes that is served using the

exhaustive service discipline, and let  $\mathcal{G}$  denote the set of nodes that is served using the gated service discipline.

Let  $W_m^*(s)$  denote the LST of the waiting time distribution. Let  $F_m(z_1, \dots, z_M)$  denote the probability generating function (p.g.f.) for the number of customers at each node at the instant the server polls node  $m$ , and for notational ease, let  $F_m(x_m) = F_m(z_1, \dots, z_{m-1}, x_m, z_{m+1}, \dots, z_M)$ . For polling systems that use the exhaustive or gated service discipline,  $W_m^*(s)$  is expressed in terms of  $F_m(x_m)$  as follows (see, for instance, Takagi [9]):

$$W_m^*(s) = \frac{1}{\bar{C}} \frac{1 - F_m(1-s/\lambda_m)}{s - \lambda_m + \lambda_m \beta_m(s)}, \quad m \in \mathcal{E}, \quad (3.2a)$$

$$W_m^*(s) = \frac{1}{\bar{C}} \frac{F_m(\beta(s)) - F_m(1-s/\lambda_m)}{s - \lambda_m + \lambda_m \beta_m(s)}, \quad m \in \mathcal{G}. \quad (3.2b)$$

Let  $f_m^{(i)}$  denote the  $i^{\text{th}}$  factorial moment of the number of customers present at node  $m$  when the server polls node  $m$ . These terms are obtained as:

$$f_m^{(i)} = \left. \frac{\partial^i F_m(z_m)}{\partial z_m^i} \right|_{z=1}. \quad (3.3)$$

From equations (3.2a) and (3.2b), we obtain  $E(W_m) = -\partial W_m^*(s)/\partial s|_{s=0}$  as (this requires two applications of L'Hospital's rule)

$$E(W_m) = \frac{f_m^{(2)}}{2\lambda_m f_m} + \frac{\lambda_m E(B_m^2)}{2(1-\rho_m)}, \quad m \in \mathcal{E}, \quad (3.4a)$$

$$E(W_m) = \frac{f_m^{(2)}}{2\lambda_m f_m} (1 + \rho_m), \quad m \in \mathcal{G}. \quad (3.4b)$$

We obtain  $E(W_m^2) = \partial^2 W_m^*(s)/\partial s^2|_{s=0}$  as (this requires three applications of L'Hospital's rule)

$$E(W_m^2) = E(W_m) \frac{\lambda_m E(B_m^2)}{(1-\rho_m)} + \frac{f_m^{(3)}}{3\lambda_m^2 f_m} + \frac{\lambda_m (B_m^3)}{3(1-\rho_m)}, \quad m \in \mathcal{E}, \quad (3.5a)$$

$$E(W_m^2) = E(W_m) \frac{\lambda_m E(B_m^2)}{(1+\rho_m)} + \frac{f_m^{(3)}}{3\lambda_m^2 f_m} (1 + \rho_m + \rho_m^2), \quad m \in \mathcal{G}. \quad (3.5b)$$

Therefore, if we can obtain the factorial moments,  $f_m^{(1)}$ ,  $f_m^{(2)}$ , and  $f_m^{(3)}$ , then we can determine  $E(W_m)$ , and  $\text{Var}(W_m) = E(W_m^2) - (E(W_m))^2$ . However, obtaining these factorial moments is not straightforward, since the generating function  $F_m(\cdot)$  is expressed, recursively, in terms of the

generating function  $F_{m-1}(\cdot)$ . It is well known (see for instance, Takagi [9]) that the generating functions are related by the following expressions:

$$F_{m+1}(z_1, \dots, z_M) = F_m(\eta_m[\sum_{k \neq m} (\lambda_k - \lambda_k z_k)]) \sigma_m(\sum_k (\lambda_k - \lambda_k z_k)), \quad m \in \mathbf{E}, \quad (3.6a)$$

$$F_{m+1}(z_1, \dots, z_M) = F_m(\beta_m[\sum_k (\lambda_k - \lambda_k z_k)]) \sigma_m(\sum_k (\lambda_k - \lambda_k z_k)), \quad m \in \mathbf{G}. \quad (3.6b)$$

Equations (3.6a) and (3.6b) form the starting point for the IS technique, which obtains an expression for  $F_m(\cdot)$  solely in terms of functions of input parameters. We can now differentiate the resulting expression for  $F_m(\cdot)$  three times to obtain the three factorial moments, and then use equations (3.4a) through (3.5b) to obtain the first and second moments of the waiting times.

The technique we develop below computes  $E(W_i)$  and  $E(W_i^2)$ . The first and second moments of the waiting times at the other nodes are obtained in an identical manner. We first define a recursive (nested) function,  $\gamma_m(j)$ , as follows:

$$\gamma_m(0) = z_m, \quad m=1, \dots, M, \quad (3.7a)$$

$$\gamma_m(j) = \eta_m \left( \sum_{k < m} [\lambda_k - \lambda_k \gamma_k(j-1)] + \sum_{k > m} [\lambda_k - \lambda_k \gamma_k(j)] \right), \quad j > 0, \quad m \in \mathbf{E}, \quad (3.7b)$$

$$\gamma_m(j) = \beta_m \left( \sum_{k \leq m} [\lambda_k - \lambda_k \gamma_k(j-1)] + \sum_{k > m} [\lambda_k - \lambda_k \gamma_k(j)] \right), \quad j > 0, \quad m \in \mathbf{G}. \quad (3.7c)$$

Let

$$\sigma_m(j) = \sigma_m \left( \sum_{k \leq m} [\lambda_k - \lambda_k \gamma_k(j)] + \sum_{k > m} [\lambda_k - \lambda_k \gamma_k(j+1)] \right), \quad j \geq 0, \quad m = 1, \dots, M. \quad (3.8)$$

We can now rewrite both equations (3.6a) and (3.6b) in the following general form:

$$F_1(\gamma_1(0), \dots, \gamma_M(0)) = F_M(\gamma_1(0), \dots, \gamma_{M-1}(0), \gamma_M(1)) \sigma_M(0).$$

Recursively expressing  $F_M(\cdot)$  in terms of  $F_{M-1}(\cdot)$ , and  $F_{M-1}(\cdot)$  in terms of  $F_{M-2}(\cdot)$ , and so on, we obtain:

$$F_1(\gamma_1(0), \dots, \gamma_M(0)) = F_1(\gamma_1(1), \dots, \gamma_M(1)) \prod_m \sigma_m(0).$$

We now continue to recursively express  $F_1(\gamma_1(j), \dots, \gamma_M(j))$  in terms of  $F_1(\gamma_1(j+1), \dots, \gamma_M(j+1))$ , for  $j \geq 1$ , and this results in the following expression for  $F_1(\gamma_1(0), \dots, \gamma_M(0))$ :

$$F_1(\gamma_1(0), \dots, \gamma_M(0)) = F_1(\gamma_1(n), \dots, \gamma_M(n)) \prod_{j=0}^{n-1} \left( \prod_m \sigma_m(j) \right). \quad (3.9)$$

It can be shown (refer Cooper [2], and Eisenberg [4], for example), that  $\lim_{n \rightarrow \infty} \gamma_1(n) = 1$ . Hence, letting  $n \rightarrow \infty$  in equation (3.9), we obtain:

$$F_1(\gamma_1(0), \dots, \gamma_M(0)) = \prod_{j=0}^{\infty} \left( \prod_m \sigma_m(j) \right). \quad (3.10)$$

We now have an equation for  $F_1(\cdot)$  that is expressed only in terms of (complex) functions of the  $\sigma_m(j)$  values. Let  $\mathbf{z} = (z_1, \dots, z_M)$ , and  $\mathbf{1} = (1, \dots, 1)$ . Differentiating equation (3.10) once, twice, and thrice with respect to  $z_1$ , and setting  $\mathbf{z} = \mathbf{1}$ , we obtain the following expressions for the factorial moments of  $F_1(\cdot)$ :

$$f_1 = \sum_{j=0}^{\infty} \sum_m \frac{\partial \sigma_m(j)}{\partial z_1} \Big|_{\mathbf{z}=\mathbf{1}}, \quad (3.11a)$$

$$f_1^{(2)} = f_1^2 + \sum_{j=0}^{\infty} \sum_m \left( \frac{\partial^2 \sigma_m(j)}{\partial z_1^2} - \left( \frac{\partial \sigma_m(j)}{\partial z_1} \right)^2 \right) \Big|_{\mathbf{z}=\mathbf{1}}, \quad \text{and} \quad (3.11b)$$

$$f_1^{(3)} = f_1^3 + 3f_1(f_1^{(2)} - f_1^2) + \sum_{j=0}^{\infty} \sum_m \left( \frac{\partial^3 \sigma_m(j)}{\partial z_1^3} - 3 \frac{\partial \sigma_m(j)}{\partial z_1} \frac{\partial^2 \sigma_m(j)}{\partial z_1^2} + 2 \left( \frac{\partial \sigma_m(j)}{\partial z_1} \right)^3 \right) \Big|_{\mathbf{z}=\mathbf{1}}. \quad (3.11c)$$

In the above equations, for the sake of clarity, we let  $f_1$  denote the factorial moment,  $f_1^{(1)}$ , and will continue to do so henceforth. Equations (3.11a) through (3.11c) hold regardless of whether node 1 is served according to the exhaustive service discipline, or the gated service discipline.

#### 4. Expressions for the Factorial Moments of $F_m(\mathbf{z}_m)$

From equation (3.8), we note that the derivatives in equations (3.11a) through (3.11c) can be obtained in terms of  $\partial^i \gamma_m / \partial z_1^i$ , for  $i = 1, \dots, 3$ . Thus, if we define

$$\psi_m^{(i)}(j) = \lambda_m \frac{\partial^i \gamma_m(j)}{\partial z_1^i} \Big|_{\mathbf{z}=\mathbf{1}}, \quad \text{and} \quad (4.1)$$

$$X_m^{(i)}(j) = \left[ \sum_{k \leq m} \psi_k^{(i)}(j) + \sum_{k > m} \psi_k^{(i)}(j+1) \right], \quad j \geq 0, \quad (4.2)$$

then the derivatives in equations (3.11a) through (3.11c) are expressed, for  $j \geq 0$ , as:

$$\frac{\partial \sigma_m(j)}{\partial z_1} \Big|_{\mathbf{z}=\mathbf{1}} = E(S_m) X_m^{(1)}(j), \quad \frac{\partial^2 \sigma_m(j)}{\partial z_1^2} \Big|_{\mathbf{z}=\mathbf{1}} = E(S_m) X_m^{(2)}(j) + E(S_m^2) (X_m^{(1)}(j))^2, \quad (4.3a)$$

$$\frac{\partial^3 \sigma_m(j)}{\partial z_1^3} \Big|_{\mathbf{z}=\mathbf{1}} = E(S_m^3) (X_m^{(1)}(j))^3 + 3 E(S_m^2) X_m^{(1)}(j) X_m^{(2)}(j) + E(S_m) X_m^{(3)}(j). \quad (4.3b)$$

Note that equations (4.3a) and (4.3b) are valid for both  $m \in \mathfrak{E}$ , and  $m \in \mathfrak{G}$ . We can obtain a recursive expression for  $\psi_m^{(i)}(j)$ ,  $i = 1, 2, 3$ , for  $j > 0$ , by differentiating equations (3.7b) and (3.7c) with respect to  $z_1$  the appropriate number of times, and setting  $z$  equal to  $\mathbf{1}$ . The expressions for these terms will depend on whether  $m \in \mathfrak{E}$ , or  $m \in \mathfrak{G}$ .

We will first obtain the recursive expressions for  $m \in \mathfrak{E}$ . Let

$$Y_m^{(i)}(j) = \left[ \sum_{k < m} \psi_k^{(i)}(j) + \sum_{k > m} \psi_k^{(i)}(j+1) \right] = X_m^{(i)}(j) - \psi_m^{(i)}(j), \quad j \geq 0. \quad (4.4)$$

Analogous to equations (4.3a) and (4.3b) for the derivatives of  $\sigma_m(j)$  evaluated at  $z = \mathbf{1}$ , recursive expressions for  $\psi_m^{(i)}(j)$ ,  $m \in \mathfrak{E}$ , and  $j > 0$ , are obtained as

$$\psi_m^{(1)}(j) = \lambda_m E(H_m) Y_m^{(1)}(j-1), \quad \psi_m^{(2)}(j) = \lambda_m (E(H_m) Y_m^{(2)}(j-1) + E(H_m^2) (Y_m^{(1)}(j-1))^2), \quad (4.5a)$$

$$\psi_m^{(3)}(j) = \lambda_m (E(H_m^3) (Y_m^{(1)}(j-1))^3 + 3 E(H_m^2) Y_m^{(1)}(j-1) Y_m^{(2)}(j-1) + E(H_m) Y_m^{(3)}(j-1)). \quad (4.5b)$$

For  $m \in \mathfrak{G}$ , and  $j > 0$ , the corresponding expressions for  $\psi_m^{(i)}(j)$  are:

$$\psi_m^{(1)}(j) = \lambda_m E(B_m) X_m^{(1)}(j-1), \quad \psi_m^{(2)}(j) = \lambda_m (E(B_m) X_m^{(2)}(j-1) + E(B_m^2) (X_m^{(1)}(j-1))^2), \quad (4.6a)$$

$$\psi_m^{(3)}(j) = \lambda_m (E(B_m^3) (X_m^{(1)}(j-1))^3 + 3 E(B_m^2) X_m^{(1)}(j-1) X_m^{(2)}(j-1) + E(B_m) X_m^{(3)}(j-1)). \quad (4.6b)$$

Let

$$\delta_m = \frac{\rho_m}{1-\rho_m}, \quad \Delta_m = \frac{\rho_m}{1+\rho_m}, \quad \text{and} \quad \vartheta_m = \frac{\lambda_m E(B_m^2)}{\rho_m^2}. \quad (4.7)$$

We first account for all the unknowns,  $\psi_m^{(3)}(j)$ ,  $j > 0$ , using the following observation.

**Observation 1:** a) The term  $\psi_k^{(1)}(0) = \lambda_k \frac{\partial z_k}{z_1} |_{z=\mathbf{1}} = \lambda_1$  when  $k=1$ ; it is equal to zero otherwise.  
b) The term  $\psi_k^{(i)}(0) = 0$ ,  $i > 0$ , for all  $k$ . (Refer to equation 4.1.) ■

Based on Observation 1, we obtain (Srinivasan [8]):

$$\sum_m \sum_{j=0}^{\infty} \psi_m^{(1)}(j) = \lambda_1 \frac{1-\rho_1}{1-\rho}, \quad 1 \in \mathfrak{E}, \quad \sum_m \sum_{j=0}^{\infty} \psi_m^{(1)}(j) = \frac{\lambda_1}{1-\rho}, \quad 1 \in \mathfrak{G}, \quad \text{and} \quad (4.8a)$$

$$\sum_m \sum_{j=0}^{\infty} \psi_m^{(2)}(j) = \sum_m \sum_{j=1}^{\infty} \psi_m^{(2)}(j) = \sum_m \vartheta_m \sum_{j=1}^{\infty} \frac{(\psi_m^{(1)}(j))^2}{1-\rho}. \quad (4.8b)$$



We obtain an expression for  $\sum_m \sum_{j=1}^{\infty} \psi_m^{(3)}(j)$  using an approach similar to the approach used to obtain the above equations, namely, (4.8a) and (4.8b). First, note that for  $m \in \mathbf{E}$ , from equations (4.4) and (4.5a),  $Y_m^{(1)}(j-1) = \psi_m^{(1)}(j)/\lambda_m E(H_m)$ . We can use equations (3.1) and (4.7) to rewrite this as

$$Y_m^{(1)}(j-1) = \frac{\psi_m^{(1)}(j)}{\delta_m}, \quad m \in \mathbf{E}, j > 0. \quad (4.9)$$

Again, from equations (4.4) and (4.5a), for  $m \in \mathbf{E}$ , and  $j > 0$ ,

$$Y_m^{(2)}(j-1) = \frac{\psi_m^{(2)}(j) - \lambda_m E(H_m^2) Y_m^{(1)}(j-1)}{\delta_m} = \frac{\psi_m^{(2)}(j)}{\delta_m} - \frac{\lambda_m E(H_m^2) \psi_m^{(1)}(j)}{\delta_m^3}. \quad (4.10)$$

Hence, substituting these values in equation (4.5b), and expressing  $E(H_m^2)$  and  $E(H_m^3)$  using equation (3.1), we arrive at the following intermediate result after some elementary algebra:

$$\psi_m^{(3)}(j) = (\psi_m^{(1)}(j))^3 \left( \frac{\lambda_m E(B_m^3)}{\rho_m^3 (1-\rho_m)} - 3 \frac{\vartheta_m^2}{1-\rho_m} \right) + 3 \frac{\vartheta_m}{1-\rho_m} \psi_m^{(1)}(j) \psi_m^{(2)}(j) + \delta_m Y_m^{(3)}(j-1). \quad (4.11)$$

From Observation 1,  $\psi_m^{(3)}(0) = 0$ , for all  $m$ , and hence,  $\sum_{j=1}^{\infty} Y_m^{(3)}(j-1) = \sum_{j=1}^{\infty} \sum_{k \neq m} \psi_k^{(3)}(j)$ . Therefore, if we add  $\delta_m \psi_m^{(3)}(j)$  to both sides of equation (4.11), and sum both sides of the equation over  $j=1, \dots, \infty$ , we obtain:

$$\begin{aligned} & \sum_{j=1}^{\infty} \psi_m^{(3)}(j) (1+\delta_m) \\ &= \sum_{j=1}^{\infty} \frac{(\psi_m^{(1)}(j))^3}{1-\rho_m} \left( \frac{\lambda_m E(B_m^3)}{\rho_m^3} - 3 \frac{\vartheta_m^2}{1-\rho_m} \right) + 3 \frac{\vartheta_m}{1-\rho_m} \sum_{j=1}^{\infty} \psi_m^{(1)}(j) \psi_m^{(2)}(j) + \delta_m \sum_{j=1}^{\infty} \sum_k \psi_k^{(3)}(j). \end{aligned}$$

We now sum both sides of the above equation over all  $m$ , and collect terms, to get (note that  $(1+\delta_m) = 1/(1-\rho_m)$ ):

$$\sum_m \sum_{j=1}^{\infty} \psi_m^{(3)}(j) = \sum_m \left( \frac{\lambda_m E(B_m^3)}{(1-\rho)\rho_m^3} - 3 \frac{\vartheta_m^2}{1-\rho} \right) \sum_{j=1}^{\infty} (\psi_m^{(1)}(j))^3 + 3 \sum_m \frac{\vartheta_m}{1-\rho} \sum_{j=1}^{\infty} \psi_m^{(1)}(j) \psi_m^{(2)}(j). \quad (4.12)$$

Following an identical approach, we can show that equation (4.12) holds even for  $m \in \mathbf{G}$ .

From equations (3.11a), (4.2), and (4.3a), noting that  $\psi_k^{(1)}(0) = 0$  for  $k > 1$  we get:

$$f_1 = \sum_m E(S_m) \sum_{j=0}^{\infty} \left[ \sum_{k \leq m} \psi_k^{(1)}(j) + \sum_{k > m} \psi_k^{(1)}(j+1) \right] = \sum_m E(S_m) \sum_{j=0}^{\infty} \sum_k \psi_k^{(1)}(j).$$

Therefore, from equation (4.8a),  $f_1$  is obtained as:

$$f_1 = \lambda_1 (1 - \rho_1) \bar{C}, \quad 1 \in \mathbf{E}, \quad f_1 = \lambda_1 \bar{C}, \quad 1 \in \mathbf{G}. \quad (4.13)$$

In a similar manner, from equations (3.11b), (4.2), (4.3a), and (4.8b), we get:

$$f_1^{(2)} = f_1^2 + \bar{C} \sum_m \vartheta_m \sum_{j=1}^{\infty} (\psi_m^{(1)}(j))^2 + \sum_m \text{Var}(S_m) \sum_{j=0}^{\infty} (X_m^{(1)}(j))^2. \quad (4.14)$$

From equations (3.11c), (4.2), (4.3b), and (4.12), the expression for  $f_1^{(3)}$  is obtained as:

$$\begin{aligned} f_1^{(3)} = & 3f_1 f_1^{(2)} - 2f_1^3 + \sum_m T(S_m) \sum_{j=0}^{\infty} (X_m^{(1)}(j))^3 + \sum_m \text{Var}(S_m) \sum_{j=0}^{\infty} X_m^{(1)}(j) X_m^{(2)}(j) \\ & + \bar{C} \sum_m \left( \frac{\lambda_m E(B_m^3)}{\rho_m^3} - 3\vartheta_m^2 \right) \sum_{j=1}^{\infty} (\psi_m^{(1)}(j))^3 + 3\bar{C} \sum_m \vartheta_m \sum_{j=1}^{\infty} \psi_m^{(1)}(j) \psi_m^{(2)}(j), \end{aligned} \quad (4.15)$$

where  $T(S_m)$  denotes the third central moment of the random variable  $S_m$ . (Note that  $\text{Var}(S_m)$  is the second central moment of  $S_m$ .) Equations (4.14) and (4.15) are valid regardless of whether node 1 is served using the exhaustive service discipline or using the gated service discipline.

Once we compute the three factorial moments, the mean waiting time and the second moment of the waiting time at node 1 can be obtained using equations (3.4) and (3.5). While the first factorial moment is directly computed, computing  $f_1^{(2)}$  and  $f_1^{(3)}$  apparently require the computation of infinite sums. However, we can show, using arguments similar to those presented by Levy [7], that the number of iterations required to compute these moments is logarithmic in the accuracy required. Let  $\rho_{\min} = \min_m(\rho_m)$ . Then the overall complexity of the algorithm to determine the variance of the waiting time at one station can be shown to be  $O(M \log_{\mu} \epsilon)$ , where  $\epsilon$  is the accuracy required, and  $\mu = [(\rho - \rho_{\min}) / (1 - \rho_{\min})]^2$ . In section 5, we present the algorithm to compute the three factorial moments, and thereby the variance of the waiting time. In section 6, we present a number of computational results, and report on the execution time of the algorithm. In section 6, we also present a closed form expression for  $f_1^{(2)}$  and  $f_1^{(3)}$ , for the special case of a pure exhaustive service system with two nodes.

## 5. The Algorithm

In order to compute the factorial moments using equations (4.13) through (4.15), we must evaluate the terms  $\psi_m^{(1)}(j)$ ,  $\psi_m^{(2)}(j)$ ,  $X_m^{(1)}(j)$ , and  $X_m^{(2)}(j)$ ,  $m = 1, \dots, M$ , iteratively, starting at  $j = 0$ . It is convenient, at this stage, to define the terms,  $\chi_m(j)$  and  $\chi_m^{(2)}$ , as follows:

$$\chi_m(j) = \frac{\psi_m^{(1)}(j)}{\lambda_1} = \frac{\lambda_m \partial \gamma_m(j)}{\lambda_1 \partial z_1}, \quad \text{and} \quad \chi_m^{(2)}(j) = \frac{\psi_m^{(2)}(j)}{\lambda_1^2}. \quad (5.1)$$

Instead of computing  $\psi_m^{(1)}(j)$  and  $\psi_m^{(2)}(j)$ , we compute  $\chi_m(j)$  and  $\chi_m^{(2)}(j)$ , starting with  $\chi_1(0) = 1$ ,  $\chi_m(0) = 0$  for  $m \neq 1$ , and  $\chi_m^{(2)}(j) = 0$  for all  $m$ . From equations (4.1) through (4.6) we get:

$$\chi_m(j) = \delta_m \left[ \sum_{k < m} \chi_k(j-1) + \sum_{k > m} \chi_k(j) \right], \quad m \in \mathbf{E}, \quad j > 0, \quad (5.2a)$$

$$\chi_m(j) = \rho_m \left[ \sum_{k \leq m} \chi_k(j-1) + \sum_{k > m} \chi_k(j) \right], \quad m \in \mathbf{G}, \quad j > 0. \quad (5.2b)$$

From equations (4.1) through (4.6), we also get

$$\chi_m^{(2)}(j) = \delta_m \left[ \sum_{k < m} \chi_k^{(2)}(j) + \sum_{k > m} \chi_k^{(2)}(j+1) + \frac{\vartheta_m}{\rho_m} \chi_m^2(j) \right], \quad m \in \mathbf{E}, \quad j > 0. \quad (5.3a)$$

$$\chi_m^{(2)}(j) = \rho_m \left[ \sum_{k \leq m} \chi_k^{(2)}(j-1) + \sum_{k > m} \chi_k^{(2)}(j) + \frac{\vartheta_m}{\rho_m} \chi_m^2(j) \right], \quad m \in \mathbf{G}, \quad j > 0. \quad (5.3b)$$

Equations (5.2) and (5.3) suggest that the computational effort to obtain  $\chi_m(j)$  and  $\chi_m^{(2)}(j)$  for a given  $j$  requires  $O(M)$  operations. However, we can express these terms as functions of neighboring terms with  $O(1)$  arithmetic operations, independent of  $M$ . The computational formula we derive will, however, depend on the service discipline used for node  $m+1$ . We derive the computational formula for  $\chi_m(j)$ , when  $m \in \mathbf{E}$  and  $m+1 \in \mathbf{G}$ . The expressions for the other cases are obtained in a similar manner. We consider a node  $m \neq M$ . From equation (5.2a) applied at  $m$  and equation (5.2b) applied at  $m+1$ , we obtain the following relationship between  $\chi_m(j)$  and  $\chi_{m+1}(j)$  when  $m \in \mathbf{E}$  and  $m+1 \in \mathbf{G}$ :

$$\begin{aligned} \chi_m(j) &= \delta_m \left[ \sum_{k \leq m+1} \chi_k(j-1) + \sum_{k > m+1} \chi_k(j) + \chi_{m+1}(j) - \chi_m(j-1) - \chi_{m+1}(j-1) \right] \\ &= \delta_m \left[ \chi_{m+1}(j)/\rho_{m+1} + \chi_{m+1}(j) - \chi_m(j-1) - \chi_{m+1}(j-1) \right]. \\ &= \delta_m \left[ \chi_{m+1}(j)/\Delta_{m+1} - \chi_m(j-1) - \chi_{m+1}(j-1) \right]. \end{aligned}$$

The expressions for all the possible combinations are obtained in a similar manner, and presented below as equations (5.5a) through (5.9). We use the convention  $\chi_{m+1}(k) = \chi_1(k-1)$  if  $m = M$ . Similarly,  $\chi_{m+1}^{(2)}(k) = \chi_1^{(2)}(k-1)$  if  $m = M$ . There is one special case to be considered; the case where  $m = M$ , and  $j = 1$ . For this special case, we have:

$$\chi_M(1) = \delta_M, \quad \chi_M^{(2)}(1) = \delta_M \frac{\vartheta_M}{\rho_M} \chi_M^2(1), \quad M \in \mathfrak{E}; \quad (5.5a)$$

$$\chi_M(1) = \rho_M, \quad \chi_M^{(2)}(1) = \rho_M \frac{\vartheta_M}{\rho_M} \chi_M^2(1), \quad M \in \mathfrak{G}. \quad (5.5b)$$

In general, we have the following expressions. Let:

$$A_{m+1}(j) = \frac{\chi_{m+1}^{(2)}(j)}{\rho_{m+1}} - \frac{\vartheta_{m+1}}{\rho_{m+1}} \chi_{m+1}^2(j) + \frac{\vartheta_m}{\rho_m} \chi_m^2(j), \quad \text{and} \quad B_{m+1}(j) = \frac{\chi_{m+1}^{(2)}(j)}{\Delta_{m+1}} - \frac{\vartheta_{m+1}}{\rho_{m+1}} \chi_{m+1}^2(j) + \frac{\vartheta_m}{\rho_m} \chi_m^2(j).$$

Case i)  $m \in \mathfrak{E}, m+1 \in \mathfrak{E}$ :

$$\chi_m(j) = \delta_m \left[ \frac{\chi_{m+1}(j)}{\rho_{m+1}} - \chi_m(j-1) \right], \quad \chi_m^{(2)}(j) = \delta_m [A_{m+1}(j) - \chi_m^{(2)}(j-1)], \quad (5.6)$$

Case ii)  $m \in \mathfrak{E}, m+1 \in \mathfrak{G}$ :

$$\chi_m(j) = \delta_m \left[ \frac{\chi_{m+1}(j)}{\Delta_{m+1}} - \chi_m(j-1) - \chi_{m+1}(j-1) \right], \quad \chi_m^{(2)}(j) = \delta_m [B_{m+1}(j) - \chi_m^{(2)}(j-1) - \chi_{m+1}^{(2)}(j-1)], \quad (5.7)$$

Case iii)  $m \in \mathfrak{G}, m+1 \in \mathfrak{E}$ :

$$\chi_m(j) = \rho_m \left[ \frac{\chi_{m+1}(j)}{\rho_{m+1}} \right], \quad \chi_m^{(2)}(j) = \rho_m [A_{m+1}(j)], \quad (5.8)$$

Case iv)  $m \in \mathfrak{G}, m+1 \in \mathfrak{G}$ :

$$\chi_m(j) = \rho_m \left[ \frac{\chi_{m+1}(j)}{\Delta_{m+1}} - \chi_{m+1}(j-1) \right], \quad \chi_m^{(2)}(j) = \rho_m [B_{m+1}(j) - \chi_{m+1}^{(2)}(j-1)]. \quad (5.9)$$

Once we compute  $\chi_m(j+1)$  and  $\chi_m^{(2)}(j+1)$ , we quickly compute  $X_m^{(1)}(j)$ , and  $X_m^{(2)}(j)$  as follows. From equations (4.4), (4.9) and (5.1),

$$X_m^{(1)}(j) = \lambda_1 \chi_m(j) + Y_m^{(1)}(j) = \lambda_1 \left[ \chi_m(j) + \frac{\chi_m(j+1)}{\delta_m} \right], \quad m \in \mathfrak{E}, \quad (5.10a)$$

$$X_m^{(2)}(j) = \lambda_1^2 \chi_m^{(2)}(j) + Y_m^{(2)}(j) = \lambda_1^2 \left[ \chi_m^{(2)}(j) + \frac{\chi_m^{(2)}(j+1)}{\delta_m} - \frac{\vartheta_m}{\rho_m} \chi_m^2(j+1) \right], \quad m \in \mathfrak{E}. \quad (5.10b)$$

Similarly, from equations (4.2), (4.6a), and (5.1), we obtain

$$X_m^{(1)(j)} = \lambda_1 \frac{\chi_m^{(j+1)}}{\rho_m}, \quad m \in \mathcal{G}, \quad (5.11a)$$

$$X_m^{(2)(j)} = \lambda_1^2 \left[ \frac{\chi_m^{(2)(j+1)}}{\rho_m} - \frac{\vartheta_m}{\rho_m} \chi_m^{(2)(j+1)} \right], \quad m \in \mathcal{G}. \quad (5.11b)$$

The number of arithmetic operations required to compute these terms is thus independent of the number of nodes in the system. Hence, for a given  $j$ , computing these terms for all  $m$  requires  $O(M)$  operations. Let

$$\varphi_m = \sum_{j=1}^{\infty} \chi_m^{(2)(j)}, \quad \theta_m = \sum_{j=1}^{\infty} \chi_m^{(3)(j)}, \quad \omega_m = \sum_{j=1}^{\infty} \chi_m^{(j)} \chi_m^{(2)(j)}, \quad (5.12a)$$

$$\xi_m = \sum_{j=0}^{\infty} \left( \frac{X_m^{(1)(j)}}{\lambda_1} \right)^2, \quad \alpha_m = \sum_{j=0}^{\infty} \left( \frac{X_m^{(1)(j)}}{\lambda_1} \right)^3, \quad \kappa_m = \sum_{j=0}^{\infty} \left( \frac{X_m^{(1)(j)}}{\lambda_1} \right) \left( \frac{X_m^{(2)(j)}}{\lambda_1^2} \right), \quad (5.12b)$$

$$\pi_m = \frac{\lambda_m E(B_m^3)}{\rho_m^3}. \quad (5.13)$$

The expressions for  $f_1^{(2)}$  and  $f_1^{(3)}$  are now rewritten as:

$$f_1^{(2)} = f_1^2 + \lambda_1^2 \left( \bar{C} \sum_m \vartheta_m \varphi_m + \sum_m \text{Var}(S_m) \xi_m \right), \quad (5.14a)$$

$$f_1^{(3)} = 3 f_1 f_1^{(2)} - 2 f_1^3 + \lambda_1^3 \sum_m \left( T(S_m) \alpha_m + 3 \text{Var}(S_m) \kappa_m + \bar{C} ((\pi_m - 3 \vartheta_m^2) \theta_m + 3 \vartheta_m \omega_m) \right). \quad (5.14b)$$

The iterative algorithm to compute the waiting time variances is given below as Algorithm 1.

### Algorithm 1.

0. Initialize values: Set a)  $j = 0$ , b)  $\chi_1(j) = 1$ , and  $\chi_m(j) = 0$ ,  $m = 2, \dots, M$ , c)  $\chi_m^{(2)(j)} = 0$ , for all  $m$ , and d)  $\varphi_m = \theta_m = \omega_m = \xi_m = \alpha_m = \kappa_m = 0$  for all  $m$ . Choose a tolerance value  $\epsilon$ .
1. Set  $j = j+1$ , and for each  $m$ , a) compute  $\chi_m(j)$  and  $\chi_m^{(2)(j)}$  using equations (5.5) through (5.10), b) compute  $X_m^{(1)(j)}$  and  $X_m^{(2)(j)}$  using equations (5.10a) through (5.11b).
2. Update the values of  $\varphi_m$ ,  $\theta_m$ ,  $\omega_m$ ,  $\xi_m$ ,  $\alpha_m$ , and  $\kappa_m$ , for  $m = 1, \dots, M$ , using equations (5.12a) and (5.12b).

3. If  $\chi_m(j) \leq \epsilon$ , for all  $m$ , go to step 4. Otherwise, go to step 1.
4. Compute  $f_1^{(2)}$  and  $f_1^{(3)}$  using equations (5.14a) and (5.14b), and compute  $E(W_1)$  and  $E(W_1^2)$  using equations (3.4a) through (3.5b). Obtain  $\text{Var}(W_1) = E(W_1^2) - (E(W_1))^2$ .

Algorithm 1 is executed for each node whose waiting time variance is desired. Each time, a new set of terms,  $\varphi_m$ ,  $\theta_m$ ,  $\omega_m$ ,  $\xi_m$ ,  $\alpha_m$ , and  $\kappa_m$ , need to be computed, for  $m = 1, \dots, M$ . One way to compute these terms would be to renumber the nodes. Therefore, if performance measures are required for, say, node 2, then we can consider renumbering nodes so that node 2 is now node 1, node 3 is now node 2, and so on. However, note that in order to compute  $\varphi_m$ ,  $\theta_m$ ,  $\omega_m$ ,  $\xi_m$ ,  $\alpha_m$ , and  $\kappa_m$ , the only data parameters involved are  $\rho_m$ ,  $\delta_m$ ,  $\vartheta_m$  and  $\Delta_m$ . We use a temporary array variable, say  $\tilde{\rho}_m$ , to store the permuted  $\rho_m$  values for every other node whose performance measures are desired. We use an offset to index the nodes and assign the appropriate  $\rho_m$  values to the array  $\tilde{\rho}_m$ ,  $m=1, \dots, M$ . We use the same offset to recompute new  $\delta_m$ ,  $\vartheta_m$  and  $\Delta_m$  values and proceed with Algorithm 1.

### The Special Cases of Pure Exhaustive and Pure Gated Systems

If all the nodes are served using the same service discipline, then we can reduce the computational effort almost by half. For the pure exhaustive service system, the terms  $X_m^{(1)}(j)$  and  $X_m^{(2)}(j)$  are obtained from equations (5.6), (5.10a), and (5.10b) as:

$$X_m^{(1)}(j) = \lambda_1 \frac{\chi_{m+1}(j+1)}{\rho_{m+1}}, \quad X_m^{(2)}(j) = \lambda_1^2 \left( \frac{\chi_{m+1}^{(2)}(j+1)}{\rho_{m+1}} - \frac{\vartheta_{m+1}}{\rho_{m+1}} \chi_{m+1}^{(2)}(j+1) \right), \quad m \in \mathbf{E}. \quad (5.15)$$

Note that these terms are expressed solely as functions of  $\chi_{m+1}(j+1)$  and  $\chi_{m+1}^{(2)}(j+1)$ . Corresponding expressions for  $\xi_m$ ,  $\alpha_m$ , and  $\kappa_m$ , are easily seen to be:

$$\xi_m = \varphi_{m+1}/\rho_{m+1}^2, \quad \alpha_m = \theta_{m+1}/\rho_{m+1}^3, \quad m \in \mathbf{E}, \quad m \neq M, \quad (5.16a)$$

$$\xi_M = \varphi_1/\rho_1^2 + 1, \quad \alpha_M = \theta_1/\rho_1^3 + 1, \quad m \in \mathbf{E}, \quad (5.16b)$$

$$\kappa_m = (\omega_{m+1} - \vartheta_{m+1} \theta_{m+1})/\rho_{m+1}^2, \quad m \in \mathbf{E}. \quad (5.16c)$$

Therefore, if we let

$$\tau_m = \bar{C} \vartheta_m + \text{Var}(S_{m-1})/\rho_m^2, \quad (5.17)$$

then the equations for  $f_1^{(2)}$  and  $f_1^{(3)}$  for the pure exhaustive service system can be written as:

$$f_1^{(2)} = f_1^2 + \lambda_1^2 \sum_m \varphi_m \tau_m + \lambda_1^2 \text{Var}(S_M), \quad m \in \mathbf{E}, \quad (5.18a)$$

$$f_1^{(3)} = 3 f_1 f_1^{(2)} - 2 f_1^3 + \lambda_1^3 \sum_m \left\{ \theta_m \left[ \bar{C} \pi_m + \frac{T(S_{m-1})}{\rho_m^3} - 3 \vartheta_m \tau_m \right] + 3 \omega_m \tau_m \right\} + \lambda_1^3 T(S_M), \quad m \in \mathbf{E}. \quad (5.18b)$$

Unlike the case of the pure exhaustive service system, for the pure gated service system  $X_m^{(1)}(j)$  and  $X_m^{(2)}(j)$  are expressed in terms of the parameters for node  $m$  only. The expressions for  $\xi_m$ ,  $\alpha_m$  and  $\kappa_m$  in this case are:

$$\xi_m = \varphi_m / \rho_m^2, \quad \alpha_m = \theta_m / \rho_m^3, \quad \kappa_m = (\omega_m - \vartheta_m \theta_m) / \rho_m^2, \quad m \in \mathbf{G}, \quad (5.19)$$

and so, setting

$$\tau_m' = \bar{C} \vartheta_m + \text{Var}(S_m) / \rho_m^2, \quad (5.20)$$

the equations for  $f_1^{(2)}$  and  $f_1^{(3)}$  are:

$$f_1^{(2)} = f_1^2 + \lambda_1^2 \sum_m \varphi_m \tau_m', \quad m \in \mathbf{G}, \quad (5.21a)$$

$$f_1^{(3)} = 3 f_1 f_1^{(2)} - 2 f_1^3 + \lambda_1^3 \sum_m \left\{ \theta_m \left[ \bar{C} \pi_m + \frac{T(S_m)}{\rho_m^3} - 3 \vartheta_m \tau_m' \right] + 3 \omega_m \tau_m' \right\}, \quad m \in \mathbf{G}. \quad (5.21b)$$

Thus, for either the pure exhaustive service system or the pure gated service system, we only evaluate  $\varphi_m$ ,  $\theta_m$  and  $\omega_m$ . This reduces the number of computations almost by half, compared to that needed for a system with a mix of the two service disciplines.

## 6. Computational Experience

We conducted a number of experiments to determine the efficiency and speed of convergence of the algorithm, for various stopping tolerances,  $\epsilon$ . (Recall that the algorithm terminates when the largest value of  $\chi_m(j)$ , for a given  $j$ , is less than or equal to  $\epsilon$ .) We found that a stopping tolerance of 0.0001 gave very accurate results, with the mean waiting times and variances matching the exact values up to about seven decimal places for the examples we report on below. (We did not check for numerical accuracy beyond seven decimal places.) We first present several examples, to give the reader a feel for actual numerical values. The program to run the experiments was coded in Pascal. The experiments were conducted on an IBM 9021-720 running the MTS operating system.

**Example 1.** The number of nodes = 5, with arrival rates 0.7, 0.8, 1.2, 0.1, and 0.2. The service times for the 5 nodes are uniformly distributed over the following ranges: [0, 0.8], [0, 1], [0, 0.2], [0.5, 1.5], and [0.2, 0.6]. The switchover times are exponentially distributed with means 0.5, 0.5, 0.4, 1.0, and 1.2. The utilizations at the five nodes are 0.28, 0.40, 0.12, 0.10, and 0.08 respectively, and  $\rho = 0.98$ . The mean waiting times and the waiting time variances are presented below for the following cases: i) the pure exhaustive service system, ii) the pure gated service system, iii) a system with a mix of service disciplines, where nodes 2, and 5 use the gated service discipline, while nodes 1, 3 and 4 use the exhaustive service discipline.

Note that the uniform distribution with range [a,b] has 2<sup>nd</sup> and 3<sup>rd</sup> moments  $(b^3 - a^3)/3(b-a)$ , and  $(b^4 - a^4)/4(b-a)$ , respectively. The exponential distribution with parameter  $\mu$  has third moment  $6/\mu^3$  and third central moment  $2/\mu^3$ .

- i) **The Pure Exhaustive Service System:** The mean waiting times are  $E(W_1) = 46.9850$ ,  $E(W_2) = 39.0331$ ,  $E(W_3) = 57.3817$ ,  $E(W_4) = 58.6123$ , and  $E(W_5) = 59.9624$ . The waiting time variances are  $\text{Var}(W_1) = 1,640.6459$ ,  $\text{Var}(W_2) = 1,118.7125$ ,  $\text{Var}(W_3) = 2,458.9258$ ,  $\text{Var}(W_4) = 2,556.9872$ ,  $\text{Var}(W_5) = 2,683.2556$ . The algorithm took 208 iterations to converge, and required 93.03 milliseconds of CPU time.
- ii) **The Pure Gated Service System:** The mean waiting times are  $E(W_1) = 72.2632$ ,  $E(W_2) = 79.0607$ ,  $E(W_3) = 63.2353$ ,  $E(W_4) = 62.1746$ , and  $E(W_5) = 60.9668$ . The waiting time variances are  $\text{Var}(W_1) = 1,728.0505$ ,  $\text{Var}(W_2) = 1,741.9405$ ,  $\text{Var}(W_3) = 1,800.3384$ ,  $\text{Var}(W_4) = 1,829.5007$ ,  $\text{Var}(W_5) = 1,835.4479$ . The algorithm took 352 iterations to converge, and required 157.97 milliseconds of CPU time.
- iii) **The Mixed Service System:** The mean waiting times are  $E(W_1) = 42.2697$ ,  $E(W_2) = 82.1887$ ,  $E(W_3) = 51.7842$ ,  $E(W_4) = 52.8793$ , and  $E(W_5) = 63.4502$ . The waiting time variances are  $\text{Var}(W_1) = 1,142.2205$ ,  $\text{Var}(W_2) = 2,100.5381$ ,  $\text{Var}(W_3) = 1,731.4286$ ,  $\text{Var}(W_4) = 1,796.6408$ ,  $\text{Var}(W_5) = 2,143.8311$ . The algorithm took 299 iterations to converge, and required 192.66 milliseconds of CPU time.

**Example 2.** The number of nodes = 5. The arrival rates are 0.20, 0.10, 0.16, 0.04, and 0.04. The service times at the nodes are uniformly distributed over the following ranges: [0, 6], [0, 4], [0, 1], [1, 2], and [0, 2]. The switchover times are exponentially distributed with means 1.0, 1.0, 0.8, 2.0, and 2.0. The utilizations at the five nodes are 0.60, 0.20, 0.08, 0.06, and 0.04 respectively and, as in the previous example,  $\rho = 0.98$ . The mean waiting times and the waiting time variances are presented below for: i) the pure exhaustive service system, ii) the pure gated



service system, iii) a system with a mix of service disciplines, where nodes 2, and 5 use the gated service discipline, while nodes 1, 3 and 4 use the exhaustive service discipline.

- i) **The Pure Exhaustive Service System:** The mean waiting times are  $E(W_1) = 122.1462$ ,  $E(W_2) = 245.0320$ ,  $E(W_3) = 282.8141$ ,  $E(W_4) = 289.7239$ , and  $E(W_5) = 296.4669$ . The waiting time variances are  $\text{Var}(W_1) = 13,245.9090$ ,  $\text{Var}(W_2) = 56,148.0645$ ,  $\text{Var}(W_3) = 75,071.5516$ ,  $\text{Var}(W_4) = 78,832.1916$ ,  $\text{Var}(W_5) = 82,618.0849$ . The algorithm took 176 iterations to converge, and required 78.49 milliseconds of CPU time.
- ii) **The Pure Gated Service System:** The mean waiting times are  $E(W_1) = 361.2300$ ,  $E(W_2) = 270.6877$ ,  $E(W_3) = 243.5746$ ,  $E(W_4) = 239.2250$ , and  $E(W_5) = 234.7594$ . The waiting time variances are  $\text{Var}(W_1) = 36,125.1632$ ,  $\text{Var}(W_2) = 31,690.2523$ ,  $\text{Var}(W_3) = 32,496.2060$ ,  $\text{Var}(W_4) = 32,824.0134$ ,  $\text{Var}(W_5) = 33,115.3918$ . The algorithm took 403 iterations to converge, and required 179.99 milliseconds of CPU time.
- iii) **The Mixed Service System:** The mean waiting times are  $E(W_1) = 115.3107$ ,  $E(W_2) = 345.9230$ ,  $E(W_3) = 267.3364$ ,  $E(W_4) = 273.8346$ , and  $E(W_5) = 302.9340$ . The waiting time variances are  $\text{Var}(W_1) = 11,145.7619$ ,  $\text{Var}(W_2) = 75,274.2534$ ,  $\text{Var}(W_3) = 63,820.7406$ ,  $\text{Var}(W_4) = 67,018.8124$ ,  $\text{Var}(W_5) = 75,651.9396$ . The algorithm took 201 iterations to converge, and required 121.91 milliseconds of CPU time.

**Example 3.** The number of nodes = 10, with arrival rate equal to 0.8 at node 1, and equal to 0.1 at all the other nodes. The service times are exponentially distributed with means 0.5, 0.5, 0.4, 1.0, 0.4, 0.5, 0.5, 0.4, 1.0, and 0.4. The switchover times are uniformly distributed over the following ranges: [0, 1], [0, 1], [0, 0.8], [0.5, 1.5], [0, 0.8], [0, 1], [0, 1], [0, 0.8], [0.5, 1.5], and [0, 0.8]. The utilizations at the ten nodes are 0.40, 0.05, 0.04, 0.10, 0.04, 0.05, 0.05, 0.04, 0.10, and 0.04, respectively, and  $\rho = 0.91$ . The mean waiting times and the waiting time variances are presented below for: i) the pure exhaustive service system, ii) the pure gated service system, iii) a system with a mix of service disciplines, where nodes 1 through 5 use the gated service discipline, while nodes 6 through 10 use the exhaustive service discipline.

- i) **The Pure Exhaustive Service System:** The mean waiting times are  $E(W_1) = 23.2652$ ,  $E(W_2) = 36.8740$ ,  $E(W_3) = 37.3111$ ,  $E(W_4) = 34.8626$ , and  $E(W_5) = 37.2287$ ,  $E(W_6) = 36.8793$ ,  $E(W_7) = 36.9212$ ,  $E(W_8) = 37.3629$ ,  $E(W_9) = 34.9234$ , and  $E(W_{10}) = 37.2999$ . The waiting time variances are  $\text{Var}(W_1) = 313.1774$ ,  $\text{Var}(W_2) = 825.2952$ ,  $\text{Var}(W_3) = 847.6837$ ,  $\text{Var}(W_4) = 729.0181$ ,  $\text{Var}(W_5) = 840.5882$ ,  $\text{Var}(W_6) = 825.7207$ ,  $\text{Var}(W_7) = 829.3474$ ,  $\text{Var}(W_8) =$

852.1895,  $\text{Var}(W_9) = 734.0381$ ,  $\text{Var}(W_{10}) = 846.8985$ . The algorithm took 50 iterations to converge, and required 90.73 milliseconds of CPU time.

ii) The Pure Gated Service System: The mean waiting times are  $E(W_1) = 50.4670$ ,  $E(W_2) = 37.9978$ ,  $E(W_3) = 37.6446$ ,  $E(W_4) = 39.8357$ , and  $E(W_5) = 37.5621$ ,  $E(W_6) = 37.9356$ ,  $E(W_7) = 37.9392$ ,  $E(W_8) = 37.5820$ ,  $E(W_9) = 39.7653$ , and  $E(W_{10}) = 37.4834$ . The waiting time variances are  $\text{Var}(W_1) = 548.3302$ ,  $\text{Var}(W_2) = 658.8032$ ,  $\text{Var}(W_3) = 664.4823$ ,  $\text{Var}(W_4) = 640.3986$ ,  $\text{Var}(W_5) = 658.1578$ ,  $\text{Var}(W_6) = 654.1311$ ,  $\text{Var}(W_7) = 654.4568$ ,  $\text{Var}(W_8) = 659.7992$ ,  $\text{Var}(W_9) = 635.3322$ ,  $\text{Var}(W_{10}) = 652.2293$ . The algorithm took 75 iterations to converge, and required 135.62 milliseconds of CPU time.

iii) The Mixed Service System: The mean waiting times are  $E(W_1) = 50.6987$ ,  $E(W_2) = 38.1981$ ,  $E(W_3) = 37.8473$ ,  $E(W_4) = 40.0546$ , and  $E(W_5) = 37.7763$ ,  $E(W_6) = 34.5286$ ,  $E(W_7) = 34.5329$ ,  $E(W_8) = 34.9119$ ,  $E(W_9) = 32.5897$ , and  $E(W_{10}) = 34.7659$ . The waiting time variances are  $\text{Var}(W_1) = 565.0151$ ,  $\text{Var}(W_2) = 674.7923$ ,  $\text{Var}(W_3) = 680.7920$ ,  $\text{Var}(W_4) = 657.3561$ ,  $\text{Var}(W_5) = 675.3881$ ,  $\text{Var}(W_6) = 629.7939$ ,  $\text{Var}(W_7) = 630.2239$ ,  $\text{Var}(W_8) = 645.5318$ ,  $\text{Var}(W_9) = 550.7752$ ,  $\text{Var}(W_{10}) = 634.0304$ . The algorithm took 73 iterations to converge, and required 187.40 milliseconds of CPU time.

The IS technique we have developed is based on non-zero switchover times. However, the technique can be used to obtain the mean and variance of the waiting time at a node for systems in which the switchover times are arbitrarily close to zero as we show in example 4 below. The data in example 4 is the same as in example 3, except that the mean switchover times are very close to zero, and the variances are set equal to zero.

**Computational Remark:**<sup>1</sup> Note that the terms  $\varphi_m$ ,  $\theta_m$ ,  $\omega_m$ ,  $\xi_m$ ,  $\alpha_m$ , and  $\kappa_m$ , depend only on the arrival rates and the service time parameters. These terms are independent of the switchover time parameters. Hence, once these terms are computed for a given set of arrival rate and service time parameters, we can obtain the mean and variance of waiting time distributions at the various nodes, for different switchover time parameters, through just a few arithmetic operations. ■

**Example 4.** The data parameters are exactly as in example 3, except that the mean switchover time at each node are all equal to 0.000001, and the variance of the switchover time is equal to 0. The mean waiting times and the waiting time variances are presented below for: i) the pure exhaustive service system, ii) the pure gated service system, iii) a system with a mix of service

---

<sup>1</sup> The author thanks Professor Robert Cooper for a very fruitful discussion on this feature of the IS Technique.

disciplines, where nodes 1 through 5 use the gated service discipline, while nodes 6 through 10 use the exhaustive service discipline.

- i) The Pure Exhaustive Service System: The mean waiting times are  $E(W_1) = 4.5488$ ,  $E(W_2) = 7.2406$ ,  $E(W_3) = 7.3670$ ,  $E(W_4) = 6.7888$ , and  $E(W_5) = 7.2846$ ,  $E(W_6) = 7.2473$ ,  $E(W_7) = 7.2904$ ,  $E(W_8) = 7.4216$ ,  $E(W_9) = 6.8530$ , and  $E(W_{10}) = 7.3598$ . The waiting time variances are  $\text{Var}(W_1) = 37.6933$ ,  $\text{Var}(W_2) = 102.7229$ ,  $\text{Var}(W_3) = 106.2822$ ,  $\text{Var}(W_4) = 89.3060$ ,  $\text{Var}(W_5) = 104.1149$ ,  $\text{Var}(W_6) = 102.8147$ ,  $\text{Var}(W_7) = 103.9472$ ,  $\text{Var}(W_8) = 107.6527$ ,  $\text{Var}(W_9) = 90.8709$ ,  $\text{Var}(W_{10}) = 106.1094$ .
- ii) The Pure Gated Service System: The mean waiting times are  $E(W_1) = 6.8397$ ,  $E(W_2) = 5.2727$ ,  $E(W_3) = 5.2319$ ,  $E(W_4) = 5.5535$ , and  $E(W_5) = 5.1496$ ,  $E(W_6) = 5.2119$ ,  $E(W_7) = 5.2164$ ,  $E(W_8) = 5.1717$ ,  $E(W_9) = 5.4857$ , and  $E(W_{10}) = 5.0740$ . The waiting time variances are  $\text{Var}(W_1) = 63.3489$ ,  $\text{Var}(W_2) = 49.2457$ ,  $\text{Var}(W_3) = 49.0801$ ,  $\text{Var}(W_4) = 52.0203$ ,  $\text{Var}(W_5) = 47.7020$ ,  $\text{Var}(W_6) = 48.2603$ ,  $\text{Var}(W_7) = 48.3642$ ,  $\text{Var}(W_8) = 48.1301$ ,  $\text{Var}(W_9) = 50.9621$ ,  $\text{Var}(W_{10}) = 46.4690$ .
- iii) The Mixed Service System: The mean waiting times are  $E(W_1) = 7.0694$ ,  $E(W_2) = 5.4712$ ,  $E(W_3) = 5.4329$ ,  $E(W_4) = 5.7706$ , and  $E(W_5) = 5.3621$ ,  $E(W_6) = 4.9190$ ,  $E(W_7) = 4.9242$ ,  $E(W_8) = 4.9927$ ,  $E(W_9) = 4.5401$ , and  $E(W_{10}) = 4.8476$ . The waiting time variances are  $\text{Var}(W_1) = 68.0742$ ,  $\text{Var}(W_2) = 53.3218$ ,  $\text{Var}(W_3) = 53.2248$ ,  $\text{Var}(W_4) = 56.4554$ ,  $\text{Var}(W_5) = 52.0771$ ,  $\text{Var}(W_6) = 46.0762$ ,  $\text{Var}(W_7) = 46.2455$ ,  $\text{Var}(W_8) = 47.6637$ ,  $\text{Var}(W_9) = 38.5235$ ,  $\text{Var}(W_{10}) = 44.8867$ .

A few remarks are made based on the experiments. It may be observed that in all the examples, the mean waiting times in the pure gated service system are higher at the nodes with higher utilizations. For the pure exhaustive service system, however, the reverse holds: the mean waiting times are higher at nodes with lower utilizations. (For the exhaustive service system, the same pattern of behavior appears to hold even for the waiting time variances.)

It is observed that the waiting time variances at the nodes for the pure gated service system are somewhat close to one another; we observed this behavior in all the experiments conducted. (In fact, the closeness is even more pronounced if we compare the standard deviations of the waiting time distributions, instead of the variances.) It may also be observed that for the mixed service systems, the waiting time variances at the nodes served using the gated service discipline are fairly close to one another. This suggests that while the gated service system results in higher mean waiting times compared to the exhaustive service system, the waiting time variances are "smoothed

out." This is a positive factor which has to be given due consideration in the design and operation of a polling system.

The relative closeness of the waiting time variances certainly does not appear to hold for the pure exhaustive service system. On the other hand, interestingly enough, we observe that the standard deviation of the waiting time distribution tracks the mean of the distribution somewhat closely, in the case of the pure exhaustive service system. Also, as remarked upon earlier, for pure exhaustive service systems the standard deviation (at a node) is high relative to that at other nodes, when the corresponding mean waiting time is high.

### **Computational Efficiency**

In general, the computational results suggest that the algorithm requires roughly the same number of iterations for a given value of  $\rho$ , for any number of nodes. A primary advantage of the IS technique is that the user can obtain the waiting time variance at select nodes without having to evaluate waiting time variances at all nodes simultaneously.

It would be of interest to compare the relative efficiency of the IS technique with the algorithm of Ferguson [5]. We note that the algorithm of Ferguson involves the solution of  $O(M^3)$  equations (and, thereby, requires  $O(M^9)$  operations). Since the IS technique is an iterative technique, it is therefore difficult to make any immediate conclusions about the relative efficiency of either algorithm. Also, since there are no computational results reported in the paper by Ferguson, we are unable to make any comparisons on execution times. However, we can indirectly estimate the efficiency of the IS technique as follows.

First, we observe, based on the expressions for  $f_1^{(2)}$  and  $f_1^{(3)}$ , that the IS technique for computing waiting time variances requires about three times the computational effort required to just compute mean waiting times using the same approach. In an earlier paper (Srinivasan [8]), we had published extensive results on the computational effort required to obtain the mean waiting times. The numerical comparisons indicated that for  $\rho$  values of about 0.90 or less, the iterative algorithm to compute mean waiting times compared very favorably with an  $O(M^3)$  algorithm, even when the iterative algorithm was used to obtain mean waiting times at all the nodes. It is, therefore, not unreasonable to conclude that if the IS technique is used to compute waiting time variances at all the nodes, it would still compare very favorably with any algorithm involving  $O(M^3)$  computations, at least for  $\rho$  values of about 0.90 or less. Note, in addition, that the IS technique has a major advantage in that one can compute waiting time variances at select nodes, without having to compute all the waiting time variances at all nodes simultaneously.

### The special case of the system with two nodes

In this case, we can obtain the waiting time variances at the nodes in closed form, for the case of the pure exhaustive service system. It is also possible to obtain the waiting time variances for either the pure gated service system or the mixed service system with two nodes, without having to perform any iterations at all. However, the expression for the waiting time variances in these latter cases is a little more complicated, and involves the solution of a system of upto three equations; we do not present these cases here. Define

$$\Gamma = \delta_1 \delta_2, \quad \text{and} \quad \zeta_m = \frac{\delta_m \vartheta_m}{\rho_m}. \quad (6.1)$$

For the pure exhaustive service system with two nodes, the terms  $\chi_1(j)$  and  $\chi_2(j)$  are obtained very simply as:

$$\chi_1(j) = \Gamma^j, \quad \chi_2(j) = \frac{\Gamma^j}{\delta_1}, \quad j \geq 1. \quad (6.2)$$

The terms  $\chi_1^{(2)}(j)$  and  $\chi_2^{(2)}(j)$  are obtained, for  $j > 1$ , as:

$$\chi_1^{(2)}(j) = \Gamma \chi_1^{(2)}(j-1) + (\Gamma^j)^2 \left( \zeta_1 + \frac{\zeta_2}{\delta_1} \right), \quad \chi_2^{(2)}(j) = \Gamma \chi_2^{(2)}(j-1) + \frac{(\Gamma^j)^2}{\delta_1} \left( \zeta_1 + \frac{\zeta_2}{\delta_1} \right). \quad (6.3)$$

The terms  $\varphi_m$ ,  $\theta_m$ , and  $\omega_m$ ,  $m = 1, 2$ , are thus evaluated in closed form:

$$\varphi_1 = \frac{\Gamma^2}{1-\Gamma^2}, \quad \varphi_2 = \frac{\Gamma^2}{\delta_1^2 (1-\Gamma^2)}, \quad (6.4a)$$

$$\theta_1 = \frac{\Gamma^3}{1-\Gamma^3}, \quad \theta_2 = \frac{\Gamma^3}{\delta_1^3 (1-\Gamma^3)}, \quad (6.4b)$$

$$\omega_1 = \frac{\zeta_1 + \zeta_2/\delta_1}{1-\Gamma^2} \theta_1, \quad \omega_2 = \frac{\zeta_1 \delta_1 + \zeta_2 \Gamma}{\Gamma(1-\Gamma^2)} \theta_2 - \varphi_2 \zeta_1. \quad (6.4c)$$

Substituting these values in equations (5.18a) and (5.18b), we obtain  $f_1^{(2)}$  and  $f_1^{(3)}$ . The first and second moments of the waiting time at node 1 are then obtained from equations (3.4a) and (3.5a).

## 7. Summary and Conclusions

We have obtained the waiting time variances in a continuous-time polling system using a powerful technique, which we term the Individual Station (IS) technique. The polling system can adopt either the pure exhaustive service discipline, the pure gated service discipline, or a system with a mix of both disciplines.

The IS technique allows the user to obtain the waiting time variances at one or more select nodes in the polling system. Furthermore, the IS technique has the following additional advantage. The resulting expression for the second moment (or the mean) of the waiting time at a node is composed of two parts, one of which is independent of the switchover time parameters. Therefore, once the means and second moments of the waiting times have been obtained for a given set of service time and switchover time parameters, the user can compute means and second moments of the waiting times for many alternate configurations involving different switchover time parameters, with just a few elementary arithmetic operations.

The IS technique for computing waiting time variances requires about three times the computational effort required to obtain the mean waiting times using the same iterative algorithm. In a previous paper we had shown, empirically, that the IS technique using the iterative algorithm computes the mean waiting times at all the nodes in a time comparable to an  $O(M^3)$  algorithm, as long as  $\rho$  is about 0.90 or less. Therefore, this suggests that the performance of the iterative algorithm for computing waiting time variances at all nodes would also be comparable to an  $O(M^3)$  algorithm, as long as  $\rho$  is about 0.90 or less.

There are a number of topics to be explored, on further applications of the IS technique. Although we can obtain the performance measures for systems in which the switchover times are arbitrarily close to zero, the technique is based on non-zero switchover times. It would be of interest to develop the IS technique for systems in which the switchover times are all equal to zero. It would also be of interest to consider using the IS technique to analyze other polling systems.

## References

- [1] Y. Aminetzah. 1975. An exact approach to the polling system. Ph.D. dissertation, McGill University, Montreal, P.Q., Canada.
- [2] R.B. Cooper and G. Murray. 1969. Queues served in cyclic order. *Bell System Technical Journal*, v 48, pp. 675-689.
- [3] R.B. Cooper. 1970. Queues served in cyclic order: waiting times. *Bell System Technical Journal*, v 49, pp. 399-413.
- [4] M. Eisenberg. 1972. Queues with periodic service and changeover time. *Operations Research*, v 20, pp. 440-451.
- [5] M.J. Ferguson. 1986. Computation of the variance of the waiting time for token rings. *IEEE Journal on Selected Areas in Communications*, v SAC-4, pp. 775-782.
- [6] L. Kleinrock. 1975. *Queueing Systems - Volume I: Theory*, John Wiley, New York.
- [7] H. Levy. 1991. A note on the complexity of Swartz's method for calculating the expected delay in non-symmetric cyclic polling systems. *Operations Research Letters*, v 10, pp. 363-368.
- [8] M.M. Srinivasan 1991. The Individual Station Technique for the Analysis of Continuous-Time Polling Systems. Technical Report 91-36, Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, MI 48109-2117.
- [9] H. Takagi. 1986. *Analysis of polling systems*, MIT Press, Cambridge, MA.
- [10] H. Takagi. 1990. Queueing analysis of polling models: an update. *Stochastic Analysis of Computer and Communication Systems*, Elsevier Science Publishers B.V. (North-Holland), Amsterdam.