

Waiting Times in Priority Polling Systems with Batch Poisson Arrivals

W. Ryu¹⁾, K.P. Jun²⁾, D.W. Kim³⁾ and B.U. Park⁴⁾

Abstract

In this paper we consider a polling system where the token is passed according to a general service order table. We derive an exact and explicit formula to compute the mean waiting time for a message when the arrivals of messages are modeled by batch Poisson processes.

1. Introduction.

A priority polling system, like a cyclic polling system, consists of a single server shared by multiple queues (or stations), say N . In a priority polling system, however, each queue is served in an order specified in a polling table of length $M(\geq N)$. Queues are given higher priority by being listed more frequently in the polling table. After serving the station at the end of the table, the server restarts at the first station in the table and the ordering is repeated. Queues can be listed in an arbitrary order and an arbitrary number of times.

In this paper, we consider a priority polling system with infinite capacity where messages arrive in batches. The batch arrival processes are assumed to be independent Poisson with different rates and the numbers of messages in a batch are assumed to have arbitrary distributions. The service time distributions are arbitrary and independent across queues. The switch-over times, which begin at the completion of serving one station and end at the polling instance to the very next one, are also arbitrary and independent. The service policy is exhaustive, which means that once the server starts to serve a queue it continues until the queue is empty. We derive an explicit formula to compute the mean waiting time for a message. The results presented here relieve one from depending on expensive simulations.

The analysis of cyclic polling systems can be found in Ferguson and Aminetzah (1985), and Takagi (1986) for example. Priority polling systems with Poisson message arrival processes have been considered by Eisenberg (1972), and Baker and Rubin (1987). This paper basically extends the results of the latter two works to the case of batch Poisson arrival processes.

1) Electronics and Telecommunications Research Institute.
2) Electronics and Telecommunications Research Institute.
3) Department of Information Communications, Okchon College.
4) Department of Statistics, Seoul National University.

In the next section, we consider single queue systems with vacation periods which provide basic building blocks for the analysis of polling systems. The main results are contained in Section 3.

2. Single Queue System with Vacation Period.

In a polling system with several stations, a station may be thought to have a vacation period when the server is polling one of the other stations. Hence, we first consider a system with a single station and a single server with vacation periods. A single server queue system with vacation periods has been considered by Eisenberg (1972), and Levy and Yechiali (1975) for Poisson arrival processes. Here, we extend their results to batch Poisson arrival processes.

There are several models for single server queue systems with vacation periods. For direct application to polling systems we consider the following model. Upon termination of a single vacation, the server returns to the queue and begins to serve those messages, if any, which have arrived during vacation, and continues to serve the queue as long as there is at least one message in the system. If the server finds the system empty at the end of a vacation, it immediately takes another vacation, and continues in this manner until it finds at least one waiting message upon return from a vacation. In what follows, a vacation period means the time interval from the point when the server leaves the queue to the point when it starts to serve the queue again. It should be differentiated from a single vacation.

2.1. Vacation and busy period. Let v denote a single vacation period, and write V for a vacation period. Denote by b a busy period initiated by 1 message. We say that a busy period begins when the server starts to serve the messages in the system and ends when there are no more messages left. A busy period initiated by 1 message is the busy period in the case that there is only 1 message waiting for service when the server returns from a vacation. Let B be a busy period (or an occupation period). Denote, by \mathcal{L}_X , the Laplace-Stieltjes transform (LST for short) of the distribution function of a random variable X , and by p_X , the probability generating function of a discrete random variable X .

Suppose the queue has independent Poisson arrivals of batches with arrival rate λ . Then following Levy and Yechiali (1975) we can deduce

$$\mathcal{L}_v(z) = \{1 - c_0 \mathcal{L}_v(z)\}^{-1} (1 - c_0) \mathcal{L}_v(z) \quad (2.1)$$

where $c_0 = \mathcal{L}_v(\lambda)$.

To find a formula for the LST of B , let $N(t)$ and $M(t)$ denote the number of batches and messages, respectively, during the time period t , and m denote the number of messages contained in an arbitrary batch. Given $N(t) = k$ ($k \geq 1$), the distribution of $M(t)$ is a k -fold convolution of the distribution of m . Furthermore, given $M(t) = j$ ($j \geq 1$) the distribution of B

is a j -fold convolution of the distribution of b . These entail, for $k \geq 1$,

$$E \{ e^{-zB} | N(v) = k \} = \{ p_m(\mathcal{L}_b(z)) \}^k.$$

In the case that $N(v) = 0$, the server takes another single vacation so that the conditional distribution of B given $N(v) = 0$ is the same as the unconditional distribution of B . The LST of the conditional distribution of B given $v = v_0$ is then given by

$$E \{ e^{-zB} | v = v_0 \} = \{ E(e^{-zB}) - 1 \} e^{-\lambda v_0} + e^{-\lambda(1 - p_m(\mathcal{L}_b(z)))v_0}.$$

From this we get the following equation for the unconditional LST of B :

$$\mathcal{L}_B(z) = \{ \mathcal{L}_B(z) - 1 \} c_0 + \mathcal{L}_v \{ \lambda(1 - p_m(\mathcal{L}_b(z))) \}$$

Solving this equation for $\mathcal{L}_B(z)$ we obtain

$$\mathcal{L}_B(z) = (1 - c_0)^{-1} [\mathcal{L}_v \{ \lambda(1 - p_m(\mathcal{L}_b(z))) \} - c_0]. \tag{2.2}$$

2.2. An Extended Markov-Chain. The method of finding the LST of the waiting time distribution relies on a Markov chain representation for the number of messages in the system. We define an embedded Markov chain with transitions occurring at epochs of service completion or vacation termination, and to distinguish between these two types of transition instants we consider an extended state space $\{ (i, j) : i = 0, 1 ; j = 0, 1, 2, \dots \}$. Here, i represents each type of transition instants, 0 for vacation termination and 1 for service completion of a message, and j counts the number of messages in the system.

Let S denote the service time of a message. If we write (i_n, j_n) for the state of the system at the n -th transition moment, then the sequence $\{ (i_n, j_n) ; n \geq 1 \}$ determines a semi-Markov chain with a transition law given by

$$\begin{aligned} (i_{n+1}, j_{n+1}) &= (1, j_n + M(S) - 1), \quad j_n \geq 1 \\ &= (0, M(V)), \quad (i_n, j_n) = (1, 0) \end{aligned}$$

The equilibrium state probabilities, $\pi_{ij} \equiv \lim_{n \rightarrow \infty} P(i_n = i, j_n = j) ; i = 0, 1 ; j = 0, 1, 2, \dots$, satisfy

$$\pi_{0j} = \pi_{10} P \{ M(V) = j \}, \quad j = 1, 2, \dots, \tag{2.3}$$

$$\pi_{1j} = \sum_{k=1}^{j+1} \pi_{.k} P \{ M(S) = j - k + 1 \}, \quad j = 0, 1, \dots \tag{2.4}$$

where $\pi_{.k} = \pi_{0k} + \pi_{1k}$.

We will express the generating functions of these state probabilities

$$\pi_0(z) = \sum_{j=0}^{\infty} z^j \pi_{0j}, \quad \pi_1(z) = \sum_{j=0}^{\infty} z^j \pi_{1j}, \quad \pi(z) = \pi_0(z) + \pi_1(z)$$

in terms of the probability generating functions of v and S . First, we can easily deduce

$$P \{ M(V) = j \} = (1 - c_0)^{-1} P \{ M(v) = j \}$$

Using (2.3) and (2.4) we obtain

$$\pi_0(z) = (1 - c_0)^{-1} \pi_{10} \{p_{M(v)}(z) - c_0\}, \tag{2.5}$$

$$\pi_1(z) = z^{-1} \{\pi(z) - \pi_{10}\} p_{M(S)}(z). \tag{2.6}$$

Applying (2.5) and using $\pi(z) = \pi_0(z) + \pi_1(z)$ we can deduce

$$\pi_1(z) = \pi_{10} (1 - c_0)^{-1} \{z - p_{M(S)}(z)\}^{-1} p_{M(S)}(z) \{p_{M(v)}(z) - 1\}. \tag{2.7}$$

Both the equations (2.5) and (2.7) involve π_{10} . Since $\pi_0(1) = \pi_{10}$ from (2.5), it follows that $\pi_{10} = 1 - \pi_1(1)$. Now applying l'Hopital's rule on (2.7) we find

$$\pi_1(1) = (1 - c_0)^{-1} \{1 - \lambda E(m) E(S)\}^{-1} \lambda \pi_{10} E(m) E(v). \tag{2.8}$$

This entails

$$\pi_{10} = \frac{(1 - c_0) \{1 - \lambda E(m) E(S)\}}{(1 - c_0) \{1 - \lambda E(m) E(S)\} + \lambda E(m) E(v)}.$$

2.3. Waiting Time. Now we find the LST of the distribution of the waiting time W of an arbitrary arrival. Let C be a sojourn time of an arbitrary message, defined by $W + S$. The number of messages that this message leaves behind it at its service completion equals to that of arrivals during its sojourn time C . This implies $P\{M(C) = j\} = \pi_{1j} / \pi_1(1)$. Thus we can write $E\{z^{M(C)}\} = \pi_1(z) / \pi_1(1)$. Since it also follows that $E\{z^{M(C)}\} = \mathcal{L}_C\{\lambda(1 - p_m(z))\}$, we can deduce utilising (2.7) and (2.8)

$$\mathcal{L}_C\{\lambda(1 - p_m(z))\} = \frac{\{1 - \lambda E(m) E(S)\} p_{M(S)}(z) \{p_{M(v)}(z) - 1\}}{\lambda E(m) E(v) \{z - p_{M(S)}(z)\}}.$$

If we let $\alpha = \lambda(1 - p_m(z))$, this results in

$$\mathcal{L}_C(\alpha) = \frac{\{1 - \lambda E(m) E(S)\} \mathcal{L}_S(\alpha) \{1 - \mathcal{L}_v(\alpha)\}}{\lambda E(m) E(v) \{\mathcal{L}_S(\alpha) - p_m^{-1}(1 - \alpha/\lambda)\}}.$$

Since $C = W + S$, the LST of the waiting time distribution is given by

$$\mathcal{L}_W(z) = \frac{\{1 - \lambda E(m) E(S)\} \{1 - \mathcal{L}_v(z)\}}{\lambda E(m) E(v) \{\mathcal{L}_S(z) - p_m^{-1}(1 - z/\lambda)\}} \tag{2.9}$$

By applying l'Hopital's rule on (2.9) successively and noting that the second derivative of the inverse function of p_m is given by $(d^2/dz^2)p_m^{-1}(z) = -p_m''(p_m^{-1}(z)) / \{p_m'(p_m^{-1}(z))\}^3$, the mean wating time is expressed by

$$E(W) = \frac{E(v^2)}{2 E(v)} + \frac{\lambda E(m) E(S^2)}{2 \{1 - \lambda E(m) E(S)\}} + \frac{E\{m(m-1)\}}{2 \lambda \{E(m)\}^2 \{1 - \lambda E(m) E(S)\}} \tag{2.10}$$

3. Priority Polling System.

In a priority polling system each queue (station) is served in turn according to a polling table T . For example, suppose there are 3 queues and $T = [1, 2, 1, 3]$. Then in a cycle station 1 is polled first and third while station 2 and station 3 are polled second and last. After station 3 is polled, the next cycle starts with station 1 again. Thus station 1 is polled twice as frequently as station 2 and 3.

A priority polling system has been considered by Baker and Rubin (1987) for Poisson arrival processes. Here, we extend their results to batch Poisson arrival processes. Our method heavily relies on the aforementioned work, and basically uses the notion of *pseudostation* which corresponds to each entry in the polling table. In the above example where $T = [1, 2, 1, 3]$ there are 4 pseudostations.

The system consists of N queues with infinite capacity. There are M pseudostations in the polling table $T(i), i = 1, \dots, M$. The service policy is exhaustive, which means that the server continues to serve a station until it becomes empty. Batches of messages arrive at each queue according to independent Poisson processes with rates $\lambda_i, i = 1, \dots, N$, and each batch arriving at station i consists of m_i messages. These batch sizes and the service times of a message, denoted by S_i for station i , are independent across the stations and have arbitrary distributions.

Let $k(j)$ be the index of the j th visited pseudostation. We note that $k(j+M) = k(j)$. Write $q(j) = T(k(j))$. Thus $q(j)$ is the index of the j th visited underlying station. Let $I_j, j = 1, 2, \dots$ be the time during which messages are accumulated for the j th visited pseudostation. If station i corresponds to the j th visited pseudostation, i.e., $i = q(j)$, then this equals the time elapses from the last departure to the next poll for station i . If we focus on only one station, say station i , and if $i = q(j)$, then I_j may be thought to be a single vacation period for station i . Hence using the equation (2.10) we readily obtain the mean waiting time for the station if we find the first two moments of I_j . Below we derive useful formulas for computing these two moments.

3.1. The First Moment of I_j . We define *station time* at a pseudostation be the time spent for switch-over from the previous pseudostation plus the time spent for serving messages in the pseudostation during a single visit. We write T_j for the station time realized at the j th visit to a pseudostation. Let h_{ij} equal 1 if messages arriving at pseudostation i during the station time of pseudostation j are not served until the next visit to pseudostation i , and 0 otherwise. For example, if $T = [1, 2, 1, 3]$ then $(h_{11}, h_{12}, h_{13}, h_{14}) = (0, 0, 0, 1)$, $(h_{21}, h_{22}, h_{23}, h_{24}) = (1, 0, 1, 1)$, $(h_{31}, h_{32}, h_{33}, h_{34}) = (0, 1, 0, 0)$, and $(h_{41}, h_{42}, h_{43}, h_{44}) = (1, 1, 1, 0)$. Let D_j be the time spent for switch-over from pseudostation j to $j+1 \pmod{M}$. Define

$C_j = I_j + T_j - D_{j-1}$. If $i = q(j)$, then it is the time between two consecutive departures from station i .

The basic relationships between I_j, T_j, C_j are :

$$I_j = \sum_{i=1}^{M-1} T_{j-i} \cdot h_{k(j), k(j-i)} + D_{j-1} \tag{3.1}$$

$$C_j = \sum_{i=1}^{M-1} T_{j-i} \cdot h_{k(j), k(j-i)} + T_j \tag{3.2}$$

A formula for the first moment of I_j may be obtained from these and another identity which are based on a recursion formula for the LST of the distribution functions of $\mathcal{T}_j \equiv (T_j, T_{j+1}, \dots, T_{j+M-1})$.

Let b_i be the busy period initiated by 1 message for station i . Given $x = (x_1, \dots, x_M)$, write for $i = 1, \dots, M-1$

$$y_{j,i} = \lambda_{q(j)} h_{k(j), k(j+i)} \{1 - p_{m_{q(i)}}(\mathcal{L}_{b_{q(i)}}(x_M))\},$$

and $y_{j,M} = \lambda_{q(j)} \{1 - p_{m_{q(i)}}(\mathcal{L}_{b_{q(i)}}(x_M))\}$. Then, by the arguments parallel to those of Baker and Rubin (1987) we obtain

$$\mathcal{L}_{\mathcal{T}_{j-1}}(x) = \mathcal{L}_{D_{k(j-M-1)}}(x_M + y_{j,M}) \cdot \mathcal{L}_{\mathcal{T}_j}(0, x_1 + y_{j,1}, \dots, x_{M-1} + y_{j,M-1}), \tag{3.3}$$

By differentiating (3.3) with respect to x_M we get

$$\begin{aligned} E(T_{j+M}) = & E(D_{j-1}) \{1 + \lambda_{q(j)} E(b_{q(j)}) E(m_{q(j)})\} \\ & + \lambda_{q(j)} E(b_{q(j)}) E(m_{q(j)}) \sum_{\ell=1}^{M-1} E(T_{j+\ell}) h_{k(j), k(j+\ell)} \end{aligned} \tag{3.4}$$

Since $\mathcal{L}_{b_i}(z) = \mathcal{L}_{S_i}(z + \lambda_i \{1 - p_{m_i}(\mathcal{L}_{b_i}(z))\})$, we find, writing $\rho_i = \lambda_i E(m_i) E(S_i)$,

$$E(b_i) = E(S_i) / (1 - \rho_i) \tag{3.5}$$

By using (3.5) we can rewrite (3.4) as

$$E(T_{j+M}) = E(D_{j-1}) + \rho_{q(j)} \left\{ \sum_{\ell=1}^{M-1} E(T_{j+\ell}) h_{k(j), k(j+\ell)} + E(T_{j+M}) \right\}.$$

A relationship between the first moments of T_j and I_j may be established from this by using (3.2) and the fact $C_j = I_j - D_{j-1} + T_j$:

$$E(T_{j+M}) = E(D_{j-1}) + \rho_{q(j)} E(I_{j+M}) / (1 - \rho_{q(j)}) \tag{3.6}$$

We now derive a formula for the first moment of I_j in equilibrium state. We first observe that in steady state the distributions of T_j, I_j, C_j, D_j do not depend on the time index j itself, but on its corresponding pseudostation index $j(\text{mod}M)$. Thus in an abuse of notation we now index all these random variables by pseudostation in what follows. In steady state, for $j = 1, \dots, M$, we can write from (3.1) and (3.2)

$$E(I_j) = \sum_{i \neq j}^M E(T_i)h_{ji} + E(D_{j-1}), \tag{3.7}$$

$$E(C_j) = \sum_{i \neq j}^M E(T_i)h_{ji} + E(T_j), \tag{3.8}$$

Furthermore, we can rewrite (3.6) as

$$E(T_j) = E(D_{j-1}) + \rho_{T(j)} E(I_j)/(1 - \rho_{T(j)}). \tag{3.9}$$

A set of M simultaneous equations for the first moment of I_j is readily obtained from (3.7), (3.8) and (3.9) :

$$E(I_j) = \sum_{i \neq j}^M \left\{ E(D_{j-1}) + \frac{\rho_{T(i)}}{1 - \rho_{T(i)}} E(I_i) \right\} \cdot h_{ji} + E(D_{j-1}), \quad 1 \leq j \leq M. \tag{3.10}$$

3.2. The Second Moment of I_j . We show that the variances of I_j 's depend on the second moments of station time T_j 's. First, we observe that the expected value of the cross product $T_j T_\ell$ depends not only on the pseudostation indices j and ℓ but also on which pseudostation is visited first. Hence we define $r_{j\ell} = E(T_j T_\ell) - E(T_j) E(T_\ell)$ when pseudostation ℓ is visited before pseudostation j . Then, in equilibrium state we can deduce from (3.1), (3.7) and the fact that T_ℓ and D_k are independent

$$\begin{aligned} \text{var}(I_j) = & \sum_{\ell=1}^{j-1} \left\{ \sum_{k=1}^{\ell-1} r_{\ell k} h_{jk} + \sum_{k=\ell}^{j-1} r_{k\ell} h_{jk} + \sum_{k=j+1}^M r_{\ell k} h_{jk} \right\} \\ & + \sum_{\ell=j+1}^M \left\{ \sum_{k=1}^{j-1} r_{k\ell} h_{jk} + \sum_{k=j+1}^{\ell-1} r_{\ell k} h_{jk} + \sum_{k=\ell}^M r_{k\ell} h_{jk} \right\} \\ & + \text{var}(D_{j-1}). \end{aligned} \tag{3.11}$$

A system of M^2 equations regarding $r_{j\ell}, j=1, \dots, M, \ell=1, \dots, M$ can be derived by differentiating (3.3) twice and utilising (3.4) and (3.5), which is given by

$$\begin{aligned} r_{j\ell} = & \frac{\rho_{T(j)}}{1 - \rho_{T(j)}} \left\{ \sum_{k=j+1}^M r_{\ell k} h_{jk} + \sum_{k=1}^{\ell-1} r_{\ell k} h_{jk} + \sum_{k=\ell}^{j-1} r_{k\ell} h_{jk} \right\}, \quad \ell < j, \\ r_{jj} = & \frac{\text{var}(D_{j-1})}{(1 - \rho_{T(j)})^2} + \frac{\lambda_{T(j)} \{ E(S_{T(j)}) \}^2 \{ E(m_{T(j)}^2) - E(m_{T(j)}) \} E(I_j)}{(1 - \rho_{T(j)})^2} \\ & + \lambda_{T(j)} \{ E(S_{T(j)}^2) E(m_{T(j)}) + \lambda_{T(j)} \{ E(S_{T(j)}) \}^3 E(m_{T(j)}) \\ & \{ E(m_{T(j)}^2) - E(m_{T(j)}) \} \} E(I_j) \div (1 - \rho_{T(j)})^3 \\ & + \frac{\rho_{T(j)}}{1 - \rho_{T(j)}} \sum_{k \neq j}^M r_{jk} h_{jk}, \\ r_{j\ell} = & \frac{\rho_{T(j)}}{1 - \rho_{T(j)}} \left\{ \sum_{k=\ell}^M r_{k\ell} h_{jk} + \sum_{k=1}^{j-1} r_{k\ell} h_{jk} + \sum_{k=j+1}^{\ell-1} r_{\ell k} h_{jk} \right\}, \quad \ell > j. \end{aligned} \tag{3.12}$$

We note from (3.12) that the three summations inside the brackets of (3.11) can be replaced by $r_{j\ell}$. Therefore, we find that

$$\text{var}(I_j) = \frac{1 - \rho_{T(j)}}{\rho_{T(j)}} \cdot \sum_{\ell \neq j}^M r_{j\ell} h_{j\ell} + \text{var}(D_{j-1}). \quad (3.13)$$

The variances of I_j can be obtained by first solving the system of equations (3.12) to get $r_{j\ell}, j=1, \dots, M, \ell=1, \dots, M$ and then plugging them into (3.13).

3.3. Mean Waiting Time. We first consider the mean waiting times of pseudostations $W_j^*, j=1, \dots, M$. As we noted it earlier, I_j corresponds to a single vacation period for the underlying station of pseudostation j . In applying (2.10) we note that the arrival rate, the number of messages in a batch and the service time depend only on the underlying station index (not on the pseudostation index). Thus we have, for $j=1, \dots, M$,

$$\begin{aligned} E(W_j^*) &= \frac{\text{var}(I_j)}{2 E(I_j)} + \frac{E(I_j)}{2} + \frac{\rho_{T(j)}}{2(1 - \rho_{T(j)})} \cdot \frac{E(S_{T(j)}^2)}{E(S_{T(j)})} \\ &\quad + \frac{E(m_{T(j)}^2) - E(m_{T(j)})^2}{2\lambda_{T(j)}\{E(m_{T(j)})\}^2(1 - \rho_{T(j)})}, \quad j=1, \dots, M. \end{aligned} \quad (3.14)$$

The mean waiting times of the underlying stations W_i can now be obtained from $E(W_j^*)$ by the principle called PASTA (Poisson Arrivals See Time Average ; see for example Wolff, 1982). Let T be one full cycle time of the server, i.e., $T = \sum_{j=1}^M T_j$. Then it is easy to see that, independent of the indices of stations, we have $T = \sum_{j: T(j)=i} C_j$. Here and below $\sum_{j: T(j)=i}$ means summation over all pseudostation indices j such that $T(j) = i$. Hence by PASTA we find, for $i=1, \dots, N$,

$$E(W_i) = \sum_{j: T(j)=i} \frac{E(C_j)}{E(T)} \cdot E(W_j^*), \quad i=1, \dots, N. \quad (3.15)$$

Now the fact that $E(T_j) = E(D_{j-1}) + \rho_{T(j)} E(C_j)$ together with (3.9) implies $E(C_j) =$

$E(I_j)/(1 - \rho_{T(j)})$. Furthermore we can see $E(T) = E(D)/(1 - \rho)$ where $D = \sum_{j=1}^M D_j$ and $\rho = \sum_{i=1}^N \rho_i$. Plugging these two expressions for C_j and T into (3.15) we get

$$E(W_i) = \frac{1 - \rho}{E(T)(1 - \rho_i)} \cdot \sum_{j: T(j)=i} E(I_j) E(W_j^*), \quad i=1, \dots, N.$$

References

- [1] Baker, J. E. and Rubin, I. (1987). Polling with a general-service order table, *IEEE Transactions on Communications*, **Com-35**, 283-288.
- [2] Eisenberg, M. (1972). Queues with periodic service and changeover time, *Operations Research*, Vol. 20, 440-451.
- [3] Ferguson, M. J. and Aminetzah, Y. J. (1985). Exact results for nonsymmetric token ring systems, *IEEE Transactions on Communications*, **Com-33**, 223-231.
- [4] Levy, Y. and Yechiali, U. (1975). Utilization of idle time in an M/G/1 queueing system, *Management Science*, Vol. 22, 202-211.
- [5] Takagi, H. (1986). *Analysis of Polling Systems*. The MIT Press, Cambridge Massachusetts, and London England.
- [6] Wolff, R. (1982). Poisson arrivals see time averages, *Operations Research*, Vol. 30, 223-231.