Journal of the Korean
Statistical Society
Vol. 26, No. 3, 1997

Waiting Times in Polling Systems with Markov-Modulated Poisson Process Arrival

D.W. Kim, W. Ryu, K.P. Jun [1], B.U. Park [2] and H.D. Bae [3]

Abstract

In queueing theory, polling systems have been widely studied as a way of serving several stations in cyclic order. In this paper we consider Markov-modulated Poisson process which is useful for approximating a superposition of heterogeneous arrivals. We derive the mean waiting time of each station in a polling system where the arrival process is modeled by a Markov-modulated Poisson process.

Key Words : Polling system; Markov-modulated poisson process; Exhaustive service policy.

[1]Electronics and Telecommunications Research Institute.
[2]Department of Statistics, Seoul National University, Seoul 151-742, Korea.
[3]Chungbuk National University.

1. INTRODUCTION

A polling system consists of a single server shared by multiple stations(or queues), and each station is served in cyclic order. It has been extensively studied by many researchers from various fields, and has been employed in many computer and communication systems. However, most analysis of polling systems has been focused on Poisson arrival processes. See Takagi (1986) for the literature on the analysis of polling systems in this case.

In this paper, we consider a polling system whose inputs consist of a superposition of heterogeneous arrivals. Examples of such arrivals include packetized voice sources with data traffic. This input process can possess correlations in the number of arrivals in adjacent time intervals, which can significantly affect queueing performance. These correlations result from the fact that the aggregate voice packet arrival rate itself is a stochastic process obtained by modulating the individual voice source packet rate by the number of voice sources in their talk spurt, which itself is a correlated process. Even if a component voice process is approximated as a renewal process, with deterministically spaced packets during a talk spurt followed by an exponentially distributed silence period, the superposition process is a complex nonrenewal process. Thus, an exact analysis of the system is intractable. Heffes and Lucantoni (1986), among others, introduce a way of approximating this superposition process by a simpler one, so called, 'Markov-modulated Poisson process' (MMPP for short). The parameters of the approximating MMPP is obtained by matching a few lower order moments. With this approximating tool at hand, we may focus on the analysis of the system whose inputs are modeled by a MMPP, which constitutes the main theme of this paper.

We consider a polling system with infinite capacity where any number of messages can be stored in each station without loss. In this case, there are three types of service disciplines : exhaustive, gated, and limited. In exhaustive service policy, the server continues to serve a station until it becomes empty, while in gated service policy only those that are waiting at polling instance are served. In limited service policy, the server serves a station until either it becomes empty, or a specified number of messages are served, whichever occurs first. Here we are mainly interested in exhaustive service system, and we derive the mean waiting time of each station in this service policy. The mean waiting time is one of the key elements in performance analysis of a network system. The results presented here provides new insights and can also be viewed as a step toward reducing the dependence on simulations, which can be expensive and typically used in studying performance issues for this kind of problems.

In the next section, we collect some basic facts about MMPP and polling system which will be used in deriving the mean waiting time in Section 3.

2. SOME BASIC FACTS

MMPP is an example of doubly stochastic Poisson process where the arrival rate itself is a stochastic process too. The arrival rate process is determined by the state of a continuous-time Markov chain. To be more specific, s-state MMPP means that when the chain is in state j $(j = 1, \ldots, s)$ the arrival process is Poisson with rate λ_j.

We consider an MMPP where the rate process is determined by a s-state homogeneous and continuous-time Markov chain. Let μ_j be the mean sojourn time of the Markov chain at the state j $(j = 1, \ldots, s)$. Then the MMPP is uniquely defined by μ_j and λ_j $(j = 1, \ldots, s)$. Let $N(t)$ denote the number of arrivals during the time period t. Below, we will give an explicit formula for the probability generating function of $N(t)$ in terms of μ_j and λ_j $(j = 1, \ldots, s)$. This will be one of the basic tools to be used to derive the mean waiting time in the next section.

Let G_X denote the probability generating function of a random variable X whose mass is concentrated on $\{0, 1, 2, \ldots\}$, i.e., $G_X(z) = E(z^X)$. Also, let α_j denote the sojourn rate of the state j, i.e., $\alpha_j = \mu_j/(\mu_1 + \cdots + \mu_s)$, and write $\alpha = (\alpha_1, \ldots, \alpha_s)$. Let Q denote the infinitesimal generator of the Markov chain, and write $\Lambda = Diag(\lambda_j)$, $e = (1, \ldots, 1)^T$. Then, by Neuts (1979), the probability generating function of $N(t)$ is given by

$$G_{N(t)}(z) = \alpha \exp[\{Q + (z-1)\Lambda\}t]e\,.$$

In the above expression, by $\exp(Pt)$ where P is a square matrix, we mean

$$\exp(Pt) = I + Pt + P^2t^2/2! + P^3t^3/3! + \cdots\,.$$

Now the infinitesimal generator Q satisfies the following two equations (see Kleinrock, 1975; Section 2.4, for example):

$$\alpha Q = 0, \quad Qe = 0\,.$$

Thus, it can be seen easily that

$$\alpha \exp[\{Q + (z-1)\Lambda\}t]e = \exp[(z-1)\Lambda t]e\,.$$

Since Λ is a diagonal matrix, the probability generating function $G_{N(t)}(z)$ can now be written as

$$G_{N(t)}(z) = \sum_{j=1}^{s} \alpha_j \exp\{-\lambda_j(1-z)t\}. \tag{2.1}$$

The analysis of exhaustive polling system heavily relies on the busy period analysis. We say that a busy period of a station begins when the server starts to serve the messages in the station. It ends when there are no more messages left in the station. In particular, a busy period initiated by 1 message is the busy period in the case that there is only 1 message waiting for service at the polling instance.

Let S denote a busy period initiated by 1 message, and also let J denote the number of messages served in such a busy period. We first give a useful identity regarding the Laplace transform of S. For this, let L_X denote the Laplace transform of the distribution function F_X of a positive random variable X, i.e., $L_X(t) = \int_0^\infty e^{-tx} dF_X(x)$. Furthermore, let B be the service time of a message, and $S^{(k)}$; $k = 1, 2, \ldots$, be i.i.d. S. Then one can write

$$S \stackrel{d}{=} B + S^{(1)} + \cdots + S^{(N(B))}. \tag{2.2}$$

Here and below '$=^d$' denotes that the two sides have the same distribution. Now using (2.2) one can show that

$$E(e^{-tS} | B = x) = e^{-tx} G_{N(x)}(L_S(t)).$$

Thus from (2.1) we obtain

$$L_S(t) = E(e^{-tS}) = \sum_{j=1}^{s} \alpha_j L_B\{t + \lambda_j(1 - L_S(t))\}. \tag{2.3}$$

We conclude this section by stating an identity regarding G_J. By a similar arguments for deriving (2.3), and using the fact

$$J \stackrel{d}{=} 1 + J^{(1)} + \cdots + J^{(N(B))}$$

where $J^{(k)}$; $k = 1, 2, \ldots$, be i.i.d. J, one can show that

$$G_J(z) = z \sum_{j=1}^{s} \alpha_j L_B\{\lambda_j(1 - G_J(z))\}. \tag{2.4}$$

3. MAIN RESULTS

We assume that the polling system consists of K stations and the arrival process to each station is an independent MMPP. To state the main results, let us introduce some notations first. We will append i in all the notations introduced in Section 2 when they are relevant to the i-th station. For example, λ_{ji} denotes the Poisson arrival rate of the i-th station when the Markov chain is in the state j. Let $\tilde{\lambda}_i$ be the average arrival rate of the i-th station, i.e., $\tilde{\lambda}_i = \sum_{j=1}^{s} \alpha_{ji}\lambda_{ji}$. We write $L_i(t)$ for the number of messages in the i-th station at the time t, and τ_i for the polling instance to the i-th station. Let $F_{k,l}(\cdot,\cdot;i)$ be the joint probability generating function of $L_k(\tau_i)$ and $L_k(\tau_i)$, i.e., $F_{k,l}(z_1,z_2;i) = E(z_1^{L_k(\tau_i)} z_2^{L_l(\tau_i)})$.

Write R_i for the reply(or switch-over) period which begins at the completion of serving the i-th station and ends at the polling instance to the $(i+1)$-th station. Let W_i denotes the wating time of a message in the i-th station. Write b_i and $b_i^{(2)}$ for the first two moments of B_i. Likewise, define r_i and $r_i^{(2)}$ for R_i. We are now ready to state the following theorem.

Theorem. The mean waiting time of the i-th station is given by

$$E(W_i) = \frac{b_i^{(2)} \sum_{j=1}^{s} \alpha_{ji}\lambda_{ji}^2}{2\tilde{\lambda}_i(1-\tilde{\lambda}_i b_i)} + \frac{(1-\sum_{k=1}^{K}\tilde{\lambda}_k b_k) f_{i,i}(i)}{2\tilde{\lambda}_i^2(1-\tilde{\lambda}_i b_i)\sum_{k=1}^{K} r_k}$$

where $f_{k,l}(i) = \frac{\partial^2}{\partial z_1 \partial z_2} F_{k,l}(z_1,z_2;i)|_{z_1=z_2=1}$.

Remark. In the above theorem, the value of $f_{i,i}(i)$ can be obtained by solving a system of K^3 equations regarding $\{f_{k,l}(i); i,k,l = 1,2,\ldots,K\}$. The system of the K^3 equations turns out to be of the same form as in Takagi (1986, page 73) with all λ_k replaced by $\tilde{\lambda}_k$. In particular, if all the stations are identical, i.e., when $\alpha_{ji} \equiv \alpha_j$, $\lambda_{ji} \equiv \lambda_j$ $(j = 1,\ldots,s)$, $B_i \equiv^d B$, and $R_i \equiv^d R$, then the value of $f_{i,i}(i)$ is given by

$$f_{i,i}(i) = \frac{K\delta^2\tilde{\lambda}^2(1-\tilde{\lambda}b)}{1-K\tilde{\lambda}b} + \frac{K(K-1)\tilde{\lambda}^3 r b^{(2)}}{(1-K\tilde{\lambda}b)^2} + \frac{K^2\tilde{\lambda}^2 r^2(1-\tilde{\lambda}b)^2}{(1-K\tilde{\lambda}b)^2}$$

where δ is the standard deviation of R and $\tilde{\lambda} = \sum_{j=1}^{s} \alpha_j\lambda_j$. Therefore, in the case of identical stations, we have

$$E(W) = \frac{\delta^2}{2r} + \frac{Kr(1-\tilde{\lambda}b) + K\tilde{\lambda}b^{(2)}}{2(1-K\tilde{\lambda}b)} + \frac{(\tilde{\lambda}^{(2)} - \tilde{\lambda}^2)b^{(2)}}{2\tilde{\lambda}(1-\tilde{\lambda}b)}.$$

where $\tilde{\lambda}^{(2)} = \sum_{j=1}^{s} \alpha_j \lambda_j^2$.

The proof of the theorem relies on the following series of lemmas. Before stating the lemmas we need to introduce additional notations. Let L_i be the number of messages remaining in the i-th station at the service completion of an arbitary message in the station, and T_i be the total number of messages served in a cycle in the i-th station. Write $\tau_i^{(n)}$ for the instance when the n-th service is completed in the i-th station, and $\bar{\tau}_i$ for the starting instance of the reply time from the i-th to the $(i+1)$-th station. Write $d_i = \bar{\tau}_i - \tau_i$.

Lemma 1. The probability generating function of L_i can be written as

$$G_{L_i}(z) = \sum_{j=1}^{s} \alpha_{ji} L_{W_i+B_i}(\lambda_{ji}(1-z)) = E\{\sum_{n=1}^{T_i} z^{L_i(\tau_i^{(n)})}\}/E(T_i) .$$

Proof. We only give the proof of the first identity. The second one is given in Takagi (1986). Note first that L_i is equal to the number of arrivals while a message stays in the i-th station, i.e. $L_i = N_i(W_i + B_i)$. Also, from (2.1) it follows that

$$E\{z^{N_i(W_i+B_i)} | W_i + B_i = x\} = E\{z^{N_i(x)}\} = \sum_{j=1}^{s} \alpha_{ji} \exp\{-\lambda_{ji}(1-z)x\} .$$

The lemma now follows.

Let $F_k(\cdot; i)$ denote the probability generating function of $L_k(\tau_i)$, and define $f_k(i)$ to be its derivative at 1, i.e., $f_k(i) = F_k'(1; i) = \frac{\partial}{\partial z} F_k(z; i)|_{z=1}$. Then we have the following lemma.

Lemma 2. It follows that

$$\begin{aligned} E\{\sum_{n=1}^{T_i} z^{L_i(\tau_i^{(n)})}\} &= \{F_i(z;i) - 1\} \sum_{j=1}^{s} \alpha_{ji} L_{B_i}(\lambda_{ji}(1-z)) \\ &\quad \div \{z - \sum_{j=1}^{s} \alpha_{ji} L_{B_i}(\lambda_{ji}(1-z))\} . \end{aligned}$$

Proof. If one applies the results of the gambler's ruin problem in Takagi (1986, page 42) with $L_i(\tau_i)$ for the initial capital, $L_i(\tau_i^{(n)})$ for the capital remaining after the n-th game, and $N_i(B_i) - 1$ for the net gain on the n-th game, then one gets

$$E\{\sum_{n=1}^{T_i} z^{L_i(\tau_i^{(n)})}\} = G_{N_i(B_i)}(z)\{F_i(z;i) - 1\}/\{z - G_{N_i(B_i)}(z)\} .$$

The lemma now follows from the fact that $G_{N_i(B_i)}(z) = \sum_{j=1}^{s} \alpha_{ji} L_{B_i}\{\lambda_{ji}(1-z)\}$.

Lemma 3. The probability generating function of T_i can be written as

$$G_{T_i}(z) = F_i\{G_{J_i}(z); i\}.$$

Proof. If we let $J_i^{(k)}; k = 1, 2, \ldots$, be i.i.d. J_i, then we can write

$$T_i =_d J_i^{(1)} + \cdots + J_i^{(L_i(\tau_i))}.$$

This implies that

$$E\{z^{T_i} | L_i(\tau_i) = k\} = \{G_{J_i}(z)\}^k,$$

and the lemma follows immediately from this.

The following lemma establishes a relationship between $F_{i,k}(\cdot, \cdot; i)$ and $F_{i,k}(\cdot, \cdot; i+1)$.

Lemma 4. It follows that for $k \neq i; i, k = 1, \ldots, K$

$$\begin{aligned} F_{i,k}(z_1, z_2; i+1) &= \sum_{j=1}^{s} \alpha_{jk} F_{i,k}\{L_{S_i}((1-z_2)\lambda_{jk}), z_2; i\} \\ &\cdot \sum_{l=1}^{s} \sum_{j=1}^{s} \alpha_{li} \alpha_{jk} L_{R_i}((1-z_1)\lambda_{li} + (1-z_2)\lambda_{jk}). \end{aligned}$$

Proof. Note first that $L_i(\tau_{i+1}) = N_i(R_i)$ and $L_k(\tau_{i+1}) = L_k(\tau_i) + N_k(d_i) + N_k(R_i)$ for $k \neq i$. Hence we can write

$$E\{z_1^{L_i(\tau_{i+1})} z_2^{L_k(\tau_{i+1})}\} = E\{z_2^{L_k(\tau_i)}\} E\{z_2^{N_k(d_i)}\} E\{z_1^{N_i(R_i)} z_2^{N_k(R_i)}\}. \tag{3.1}$$

Now since $d_i =_d S_i^{(1)} + \cdots + S_i^{(L_i(\tau_i))}$ where $S_i^{(k)}; k = 1, 2 \ldots$ are i.i.d. S_i and its Laplace transform is given by $L_{d_i}(t) = E\{L_{S_i}(t)^{L_i(\tau_i)}\}$, we have

$$E\{z_2^{N_k(d_i)}\} = \sum_{j=1}^{s} \alpha_{jk} E\{L_{S_i}((1-z_2)\lambda_{jk})^{L_i(\tau_i)}\}. \tag{3.2}$$

Similarly, we have

$$E\{z_1^{N_i(R_i)} z_2^{N_k(R_i)}\} = \sum_{l=1}^{s} \sum_{j=1}^{s} \alpha_{li} \alpha_{jk} L_{R_i}((1-z_1)\lambda_{li} + (1-z_2)\lambda_{jk}). \tag{3.3}$$

The lemma follows from (3.1), (3.2) and (3.3).

Lemma 5. It follows that for $k \neq i; i, k, = 1, \ldots, K$

$$\begin{aligned} f_k(i+1) &= r_i\tilde{\lambda}_k + f_k(i) + b_i\tilde{\lambda}_k f_i(i)/(1 - b_i\tilde{\lambda}_k)\,, \\ f_i(i+1) &= r_i\tilde{\lambda}_i\,. \end{aligned}$$

Proof. For $k \neq i$, by differentiating both sides of the equation in Lemma 4 with respect to z_2, we get

$$\begin{aligned} f_k(i+1) &= \frac{\partial}{\partial z_2} F_{i,k}(z_1, z_2; i+1)|_{z_1=z_2=1} \\ &= \sum_{j=1}^{s} \alpha_{jk}(f_i(i), f_k(i))(-\lambda_{jk}L'_{S_i}(0), 1)^T \sum_{l=1}^{s}\sum_{j=1}^{s} \alpha_{li}\alpha_{jk} \\ &\quad + \sum_{j=1}^{s} \alpha_{jk} \sum_{l=1}^{s}\sum_{j=1}^{s} \alpha_{li}\alpha_{jk}\{-\lambda_{jk}L'_{R_i}(0)\}\,. \end{aligned}$$

Now, since $\sum_{j=1}^{s} \alpha_{jk} = \sum_{l=1}^{s} \alpha_{li} = \sum_{l=1}^{s}\sum_{j=1}^{s} \alpha_{li}\alpha_{jk} = 1$, we have

$$f_k(i+1) = f_i(i)\tilde{\lambda}_k E(S_i) + f_k(i) + r_i\tilde{\lambda}_k\,. \tag{3.4}$$

By noting that from (2.3) $E(S_i) = b_i(1-b_i\tilde{\lambda}_i)^{-1}$, we can establish the first part of the lemma. The second part of the lemma follows by a similar argument.

We are now ready to prove the main theorem.

Proof of the Theorem. First, note that from the first identity of Lemma 1

$$E(W_i) = \tilde{\lambda}_i\{G'_{L_i}(1) - b_1\tilde{\lambda}_i\}\,. \tag{3.5}$$

Now, from Lemma 3 we have $E(T_i) = f_i(i)E(J_i)$, and from (2.4) we can get $E(J_i) = (1 - \tilde{\lambda}_i)^{-1}$. Furthermore, by solving the system of K^2 equations in Lemma 5, we have

$$f_i(i) = \tilde{\lambda}_i(1 - b_i\tilde{\lambda}_i) \sum_{k=1}^{K} r_k/(1 - \sum_{k=1}^{K} b_k\tilde{\lambda}_k)\,.$$

Hence it follows that

$$E(T_i) = \tilde{\lambda}_i \sum_{k=1}^{K} r_k/(1 - \sum_{k=1}^{K} b_k\tilde{\lambda}_k)\,. \tag{3.6}$$

Let a_i be the reciprocal of the right hand side of (3.6). Write $d_i(z)$ for $z - \sum_{j=1}^{s} \alpha_{ji}L_{B_i}\{\lambda_{ji}(1-z)\}$ and $e_i(z)$ for $\{F_i(z;i) - 1\}\sum_{j=1}^{s} \alpha_{ji}L_{B_i}\{\lambda_{ji}(1-z)\}$

Then, by plugging (3.6) and the right hand side of the equation in Lemma 2 into the right hand side of the second equation in Lemma 1, we have

$$G_{L_i}(z) = a_i e_i(z)/d_i(z)\,.$$

Now since $d_i(0) = e_i(0) = 0$, the derivative of G_{L_i} at 1 can be written as

$$G'_{L_i}(1) = a_i\{e''_i(1)d'_i(1) - e'_i(1)d''_i(1)\}/\{2d'_i(1)^2\}\,. \tag{3.7}$$

Note that $d'_i(1) = 1 - b_i\tilde{\lambda}_i$, and that $e'_i(1) = F'_i(1;i) = (1 - b_i\tilde{\lambda}_i)/a_i$. Also, note too that $d''_i(1) = -b_i^{(2)}\sum_{j=1}^{s}\alpha_{ji}\lambda_{ji}^2$, and that $e''_i(1) = F''_i(1;i) + 2b_i\tilde{\lambda}_i F'_i(1;i) = f_{i,i}(i) + 2b_i\tilde{\lambda}_i(1 - b_i\tilde{\lambda}_i)/a_i$. Plugging all these into the formula (3.7) and then into (3.5) establishes the theorem.

REFERENCES

(1) Heffes, H. and Lucantoni, D.M.(1986). A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance, *IEEE Journal on Selected Areas in Communications,* **Sac-4**, 856-868.

(2) Kleinrock, L.(1975). *Queueing Systems, Volume 1: Theory,* John-Wiley & Sons, New York.

(3) Neuts, M.F.(1979). A versatile Markovian point process, *Journal of Applied Probabilty,* **16**, 764-779.

(4) Takagi, H.(1986). *Analysis of Polling Systems,* The MIT Press, Cambridge Massachusetts, and London England.