Bull. Korean Math. Soc. **46** (2009), No. 4, pp. 673–689
DOI 10.4134/BKMS.2009.46.4.673

MAP_1, $MAP_2/G/1$ FINITE QUEUES WITH SERVICE SCHEDULING FUNCTION DEPENDENT UPON QUEUE LENGTHS

Doo Il Choi and Sang Min Lee

Abstract. We analyze MAP_1, $MAP_2/G/1$ finite queues with service scheduling function dependent upon queue lengths. The customers are classified into two types. The arrivals of customers are assumed to be the Markovian Arrival Processes (MAPs). The service order of customers in each buffer is determined by a service scheduling function dependent upon queue lengths. Methods of embedded Markov chain and supplementary variable give us information for queue length of two buffers. Finally, the performance measures such as loss probability and mean waiting time are derived. Some numerical examples also are given with applications in telecommunication networks.

1. Introduction

We analyze MAP_1, $MAP_2/G/1$ finite queues with a service scheduling function dependent upon queue lengths. The customers are classified into two types (type-1 and type-2) according to their service characteristics. The service order of each type is determined by a service scheduling function dependent upon queue lengths [4]. The service time of all customers irrespective of customer type has the same general distribution. A detailed description is given in Section 2.

We analyzed a similar model when the arrivals are Poisson processes [4]. However, it is well known that the Poisson process is not appropriate for modeling bursty arrival streams. The arrivals are thus assumed to be the Markovian Arrival Processes (MAPs), as introduced by Lucantoni et. al. [11]. The MAP is a nonrenewal process, and includes the phase-type renewal process, a Markov-modulated Poisson process (MMPP) and the superpositions of such processes as particular cases [11]. Asmussen and Koole [1] have shown that

Received August 12, 2008; Revised December 29, 2008.
2000 *Mathematics Subject Classification.* 60K25, 68M20, 98B22.
Key words and phrases. $MAP_1, MAP_2/G/1$, service scheduling function, loss, mean delay.
This work was supported by the Korea Research Foundation Grant funded by the Korean Government(KRF-2008-521-C00033).

the MAP is weakly dense in the class of stationary point processes. Therefore, the MAP is a fairly generalized arrival process while still remaining analytically tractable. Specially, it is appropriate to model traffics with bursty and time-correlated properties between interarrivals [5, 9]. Traffics such as voice and video in telecommunication networks have these properties [9].

Our research was motivated by performance analysis of the multiplexer in a multimedia environment, as in the case of the Broadband Integrated Services Digital Network (B-ISDN). The multiplexer such as ATM (Asynchronous Transfer Mode) supports diverse traffic streams with different characteristics. These traffic streams require different Quality of Service (QoS) such as delay and loss. Furthermore, the ATM networks should be utilized as fully as possible. Thus, in order to improve the utilization of ATM networks while meeting the QoS of each traffic streams, distinct treatment for traffic streams with different characteristics is needed. Finally, the scheduling schemes must be applied to support these traffic streams.

The representative scheduling schemes are the schemes with priority. There are static and dynamic priority schemes. As a static priority scheme, the Head of Line (HOL) priority scheme [12, 13] has been applied to satisfy QoS of traffic with stringent delay constraint. However, in HOL priority scheme, the low priority class can suffer by long delay. Recently, to overcome this shortcoming, the dynamic priority schemes including the Queue Length Threshold (QLT) scheduling scheme have been proposed [2, 3, 7, 10]. Purpose of these schemes is for improving the QoS of low priority traffic while meeting delay constraint of high priority.

In this paper, we give an unified analysis of the queueing system with a general service scheduling function. The assumption of MAPs can be applied to many bursty environments. Furthermore, the existing many scheduling schemes can be given as special cases of our model. The performance measures of these schemes are compared throughout the numerical examples. This comparison of performance measures is expected to help the system designers select the optimal scheduling scheme for their given system. The MAP also is compared with Poisson process. We show that the bursty and time-correlated properties have an important effect on the performance of system. Finally, these results emphasize the importance of exact modeling for arrival process.

Following this section, the detailed MAP and model description are given in Section 2. By using the embedded Markov chain and supplementary variable method, we obtain the queue length distribution at departure epochs and an arbitrary time in Section 3. Then, the loss probability and the mean waiting time can be derived. In Section 4, the some numerical examples are presented to illustrate the effectiveness of our proposed queueing system. In particular, we show that the existing many scheduling schemes can be given as a special case of our service scheduling function.

2. Model description and preliminary for analysis

The type-$k(k=1,2)$ customers arrive to the system according to the MAP with representation $(\overline{C}_k, \overline{D}_k)$, where $\overline{C}_k$ and $\overline{D}_k$ are $M_k \times M_k$ matrices. Here M_k is the number of states in the underlying Markov process governing arrivals of type-k customer. There are two separate buffers (buffer I and buffer II) to accommodate type-1 and type-2 customers with different capacities K_1 and K_2 respectively. The customers arriving when the corresponding buffer is full are lost. The service order for customers of each buffer is determined by the service scheduling function $f_k(i,j)$, where $f_k(i,j)$ is the probability that a type-$k(k=1,2)$ customer at service initiation is selected when there are i type-1 customers and j type-2 customers. Clearly, $f_1(i,0)=1, i>0, f_2(0,j)=1, j>0$ and $f_1(i,j)+f_2(i,j)=1$. The service times of customers are independent and identically distributed with distribution function $G(\cdot)$, mean μ and Laplace transform $G^*(s)$. The service of customers in each buffer is based on the first-come first-service. Before proceeding to the analysis, we consider the superposed arrival process of two independent MAPs with representations $(\overline{C}_1, \overline{D}_1)$ and $(\overline{C}_2, \overline{D}_2)$. Let M be $M_1 M_2$. In order to distinguish arrivals of the type-1 and type-2 customer, we introduce the following $M \times M$ matrices:

$$D_1 = \overline{D}_1 \otimes I_2, \qquad D_2 = I_1 \otimes \overline{D}_2,$$
$$C = \overline{C}_1 \oplus \overline{C}_2, \qquad D = D_1 + D_2,$$

where $\otimes$ and $\oplus$ are the Kronecker product and the Kronecker sum [8], and $I_k(k=1,2)$ is the identity matrix of the same order as $\overline{D}_k$.

Let $A_k(t)$ $(k=1,2)$ be the number of arrivals of the type-k customer during the interval $(0,t]$, and $A(t)=A_1(t)+A_2(t)$. We also set $J(t)$ be the state at time t of the underlying Markov process governing the superposed arrival process. Define the joint conditional probabilities as follows:

$$p(n_1,n_2,j,t|i) = Pr\{A_1(t)=n_1, A_2(t)=n_2, J(t)=j | A(0)=0, J(0)=i\},$$
$$n_1,\ n_2 \geq 0, \qquad 1 \leq i,\ j \leq M.$$

By the Chapman-Kolmogorov's forward equation, we have the following differential-difference equations for the matrices

$$P(n_1,n_2,t) \triangleq (p(n_1,n_2,j,t|i))_{1\leq i,j\leq M}$$

of order M:

$$P^{'}(n_1,n_2,t) = P(n_1,n_2,t)C + P(n_1-1,n_2,t)D_1 + P(n_1,n_2-1,t)D_2,$$

where $P(-1,n_2,t)$ and $P(n_1,-1,t)$ are 0 matrices. It is then easily shown that the matrix $P(n_1,n_2,t)$ has the probability generating function

$$\begin{aligned}\overline{P}(z_1,z_2,t) &\triangleq \sum_{n_1=0}^{\infty}\sum_{n_2=0}^{\infty} P(n_1,n_2,t) z_1^{n_1} z_2^{n_2} \\ &= e^{(C+z_1D_1+z_2D_2)t}, \qquad |z_1|\leq 1,\ |z_2|\leq 1.\end{aligned}$$

Let π be the steady-state probability vector of the underlying Markov process $J(t)$. Then π is given by solving the equations $\pi(C+D)=0$, $\pi\mathbf{e}=1$. The mean arrival rate of type-k customer is expressed by $\lambda_k^* = \pi D_k \mathbf{e} (k=1,2)$, where $\mathbf{e}$ is a column vector of 1s.

3. Analysis

3.1. The joint queue length distribution at departure epochs

We consider the joint queue length distribution immediately after departure epochs. Let τ_n $(n \geq 1)$ be the n-th departure epoch of a customer with $\tau_0 = 0$. We introduce the notations

$$N_1(n) = \text{ the queue length of buffer I at time } \tau_n+,$$
$$N_2(n) = \text{ the queue length of buffer II at time } \tau_n+,$$
$$J_n = \text{ the state of the underlying Markov process at time } \tau_n+.$$

The process $\{(N_1(n), N_2(n), J_n), n \geq 0\}$ then forms a Markov chain, and its states are labeled in lexicographic order, that is,

$$(0,0,1)\cdots(0,0,M)(0,1,1)\cdots(0,1,M)\cdots(K_1,K_2,M).$$

We analyze the stationary probability distribution for the joint queue length, defined by

$$x_{k,l,i} = \lim_{n\to\infty} Pr\{N_1(n)=k, N_2(n)=l, J_n=i\},$$
$$0 \leq k \leq K_1,\ 0 \leq l \leq K_2,\ 1 \leq i \leq M,$$
$$\mathbf{x}_{k,l} = (x_{k,l,1}, x_{k,l,2}, \ldots, x_{k,l,M}), \quad 0 \leq k \leq K_1,\ 0 \leq l \leq K_2,$$
$$\mathbf{x}_k = (\mathbf{x}_{k,0}, \mathbf{x}_{k,1}, \ldots, \mathbf{x}_{k,K_2}), \quad 0 \leq k \leq K_1,$$
$$\mathbf{x} = (\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{K_1}).$$

To obtain the joint queue length distribution at departure epochs, we need to know the distribution for number of arrivals during the service time. Thus, we introduce the following matrices:

$$A_{k,l} = \int_0^\infty P(k,l,x)dG(x), \quad k \geq 0, l \geq 0.$$

The (i,j)-element of the matrix $A_{k,l}$ is the conditional joint probability that there are k type-1 arrivals and l type-2 arrivals during the service time and the state of the underlying Markov process is j at the next departure epoch, given that the system is nonempty and the state of the underlying Markov process is i just after a departure.

Furthermore, let

$$A'_{k,l} = (-C^{-1})DA_{k,l}, \quad k \geq 0, l \geq 0.$$

The (i,j)-element of the matrix $A'_{k,l}$ is the conditional joint probability that there are k type-1 arrivals and l type-2 arrivals during the service time and the state of the underlying Markov process is j at the next departure epoch, given

that the system is empty and the state of the underlying Markov process is i just after a departure.
We also introduce the matrices

$$A_{k,\cdot} = \sum_{l=0}^{\infty} \int_0^{\infty} P(k,l,x)dG(x), \quad A_{\cdot,l} = \sum_{k=0}^{\infty} \int_0^{\infty} P(k,l,x)dG(x),$$

$$A_{k,\overline{l}} = \sum_{n=l}^{\infty} A_{k,n}, \quad A_{\overline{k},l} = \sum_{n=k}^{\infty} A_{n,l}, \quad k \geq 0, l \geq 0,$$

$$A^{'}_{k,\overline{K_2}} = \sum_{n=K_2}^{\infty} A^{'}_{k,n}, \quad A^{'}_{\overline{K_1},\overline{K_2}} = \sum_{n_1=K_1}^{\infty} \sum_{n_2=K_2}^{\infty} A^{'}_{n_1,n_2}.$$

The one-step transition probabilities and the matrices are now defined:

$$\begin{aligned} P_{(i,j):(k,l)}(j_1,j_2) = \lim_{n\to\infty} Pr\{&N_1(n+1) = k, N_2(n+1) = l, J_{n+1} = j_2 \\ &\mid N_1(n) = i, N_2(n) = j, J_n = j_1\}, \end{aligned}$$

$$P_{(i,j):(k,l)} = (P_{(i,j):(k,l)}(j_1,j_2))_{1\leq j_1,j_2\leq M}.$$

The one-step transition probabilities are then given by

a. For $i = 0$ and $j = 0$

$$P_{(0,0):(k,l)} = \begin{cases} A^{'}_{k,l}, & \text{if } 0 \leq k < K_1,\ 0 \leq l < K_2, \\ A^{'}_{k,\overline{K_2}}, & \text{if } 0 \leq k < K_1,\ l = K_2, \\ A^{'}_{\overline{K_1},l}, & \text{if } k = K_1,\ 0 \leq l < K_2, \\ A^{'}_{\overline{K_1},\overline{K_2}}, & \text{if } k = K_1,\ l = K_2. \end{cases}$$

b. For $i = 0$ and $j \geq 1$

$$P_{(0,j):(k,l)} = \begin{cases} 0 & \text{if } l < j-1, \\ A_{k,l-j+1}, & \text{if } 0 \leq k < K_1,\ j-1 \leq l < K_2, \\ A_{k,\overline{K_2-j+1}}, & \text{if } 0 \leq k < K_1,\ l = K_2, \\ A_{\overline{K_1},l-j+1}, & \text{if } k = K_1,\ j-1 \leq l < K_2, \\ A_{\overline{K_1},\overline{K_2-j+1}}, & \text{if } k = K_1,\ l = K_2. \end{cases}$$

c. For $i \geq 1$

$$P_{(i,j):(k,l)} = 0, \quad \text{for all } j,l, \qquad \text{if } k < i-1.$$

d. For $1 \leq i < K_1$

$$P_{(i,j):(k,l)} = \begin{cases} 0, & \text{if } l < j-1, \\ f_1(i,j)A_{k-i+1,l-j} + f_2(i,j)A_{k-i,l-j+1}, & \text{if } k < K_1, l < K_2, \\ f_1(i,j)A_{k-i+1,\overline{K_2-j}} + f_2(i,j)A_{k-i,\overline{K_2-j+1}}, & \text{if } k < K_1, l = K_2, \\ f_1(i,j)A_{\overline{K_1-i+1},l-j} + f_2(i,j)A_{\overline{K_1-i},l-j+1}, & \text{if } k = K_1, l < K_2, \\ f_1(i,j)A_{\overline{K_1-i+1},\overline{K_2-j}} + f_2(i,j)A_{\overline{K_1-i},\overline{K_2-j+1}}, & \text{if } k = K_1, l = K_2. \end{cases}$$

e. For $i = K_1$

$$P_{(i,j):(k,l)} = \begin{cases} 0, & \text{if } l < j-1, \\ f_1(K_1,j)A_{0,l-j}, & \text{if } k = K_1 - 1, l < K_2, \\ f_1(K_1,j)A_{0,\overline{K_2-j}}, & \text{if } k = K_1 - 1, l = K_2, \\ f_1(K_1,j)A_{\bar{1},l-j} + f_2(K_1,j)A_{\bar{0},l-j+1}, & \text{if } k = K_1, l < K_2, \\ f_1(K_1,j)A_{\bar{1},\overline{K_2-j}} + f_2(K_1,j)A_{\bar{0},\overline{K_2-j+1}}, & \text{if } k = K_1, l = K_2, \end{cases}$$

where matrices $A_{-1,\cdot}$ and $A_{\cdot,-1}$ are 0 matrices.

Let P be the one-step transition probability matrix of the Markov chain $\{(N_1(n), N_2(n), J_n), n \geq 0\}$. The elements of the matrix P are given by $P_{(i,j):(k,l)}(j_1, j_2)$ in lexicographic order. The vector $\mathbf{x}$ for the joint queue length distribution at departure epochs is then obtained by solving the equations

$$\mathbf{x}\,P = \mathbf{x}, \qquad \mathbf{x}\,\mathbf{e} = 1.$$

3.2. Queue length distribution of each buffer at an arbitrary time

Let $N_1(t)$ and $N_2(t)$ be the queue lengths (excluding customer in service) of buffers I and II at time t respectively. We also introduce the notation

$$\xi(t) = \begin{cases} 0 & \text{if the server is idle at time } t, \\ 1 & \text{if the server is busy at time } t. \end{cases}$$

We first investigate the stationary probability that the system is empty:

$$y_0(j) = \lim_{t\to\infty} Pr\{N_1(t) = 0, N_2(t) = 0, J(t) = j, \xi(t) = 0\}, 1 \leq j \leq M,$$
$$\mathbf{y}_0 = (y_0(1), y_0(2), \ldots, y_0(M)).$$

Since the system is work-conserving, the $\mathbf{y}_0$ doesn't depend on the service discipline. Thus, we may assume the service discipline to be first-come first-service irrespective of customer type. The j-th element $y_0(j)$ of $\mathbf{y}_0$ is then derived by applying the key renewal theorem (e.g. see Theorem 6.3, p. 153, Cinlar [6]):

$$y_0(j) = \sum_{k=1}^{M} \frac{1}{m(0,k)} \int_0^\infty p(0,0,j,t|k)dt,$$

where $m(0,k)$ denotes the mean recurrence time of the state $(0,0,k)$ in the Markov chain $\{(N_1(n), N_2(n), J_n), n \geq 0\}$.
By the fact that the j-th element of the vector $\int_0^\infty P(0,0,t)dt\mathbf{e} = -C^{-1}\mathbf{e}$ is the mean duration of an idle period starting in state j of the underlying Markov chain $\{J_n, n \geq 0\}$, we easily have

$$m(0,k) = Ex_{0,0,k}^{-1},$$

where $E = \mathbf{x}_{0,0}(-C^{-1})\mathbf{e} + \mu$ is the mean interdeparture time of customers [6]. Finally, we obtain the probability $y_0(j)$:

$$y_0(j) = j\text{-th element of } \frac{1}{E}\mathbf{x}_{0,0}(-C^{-1}).$$

Next, we derive the queue length distribution of buffer I at an arbitrary time when the server is busy.
The stationary queue length probability of buffer I is then defined

$$y_n^1(j) = \lim_{t\to\infty} Pr\{N_1(t) = n, J(t) = j, \xi(t) = 1\},\ 1 \le j \le M,$$
$$\mathbf{y}_n^1 = (y_n^1(1), \ldots, y_n^1(M)), \qquad 0 \le n \le K_1.$$

We use the supplementary variable method for analysis. Let $\hat{T}$ and $\tilde{T}$ be the remaining and the elapsed service time for the customer in service respectively. We define the joint probability distribution for the queue length of buffer I and the remaining service time of the customer in service at arbitrary time τ as

$$\alpha(n, j, t)dt = Pr\{N_1(\tau) = n, J(\tau) = j, t < \hat{T} \le t + dt \ \ \xi(\tau) = 1\}.$$

We also define the Laplace transform of $\alpha(n, j, t)$ and the vectors

$$\alpha^*(n, j, s) = \int_0^\infty e^{-st}\alpha(n, j, t)dt,$$
$$\alpha_n^*(s) = (\alpha^*(n, 1, s), \ldots, \alpha^*(n, M, s)), \quad 0 \le n \le K_1.$$

Since the queue length of buffer I at an arbitrary time τ contains the number of arrivals during the elapsed service time from last departure epoch before time τ, we need to know the number of arrivals during the elapsed service time. The conditional probability $\beta(n, j_1, j_2, t)dt$ is defined as

$$\beta(n, j_1, j_2, t)dt = Pr\{\ n \text{ arrivals of type-1 customer during } \tilde{T},$$
$$J(\tau) = j_2,\ t < \hat{T} \le t + dt \mid J(\bar{\tau}) = j_1\}, \quad n \ge 0,$$

where $\bar{\tau}$ is the starting time of the service time which includes the time τ.
We also define the Laplace transform of $\beta(n, j_1, j_2, t)$ and the matrices

$$\beta^*(n, j_1, j_2, s) = \int_0^\infty e^{-st}\beta(n, j_1, j_2, t)dt,$$
$$\beta_n^*(s) = (\beta^*(n, j_1, j_2, s))_{1\le j_1, j_2\le M}.$$

Then, the vectors $\alpha_n^*(s)$ satisfy the following equations:
For $0 \le n < K_1$,

$$\alpha_n^*(s) = \frac{\mu}{E}\left[\mathbf{x}_{0,0}(-C^{-1})D\beta_n^*(s) + \sum_{m=0}^{K_2}\sum_{k=1}^{n+1} f_1(k, m)\mathbf{x}_{k,m}\beta_{n-k+1}^*(s)\right.$$
$$\left. + \sum_{m=1}^{K_2}\sum_{k=0}^{n} f_2(k, m)\mathbf{x}_{k,m}\beta_{n-k}^*(s)\right],$$

and

$$\alpha^*_{K_1}(s) = \frac{\mu}{E}\left[\mathbf{x}_{0,0}(-C^{-1})D\sum_{l=K_1}^{\infty}\beta^*_l(s) + \sum_{m=0}^{K_2}\sum_{k=1}^{K_1} f_1(k,m)\mathbf{x}_{k,m}\sum_{l=K_1-k+1}^{\infty}\beta^*_l(s)\right.$$
$$\left.+\sum_{m=1}^{K_2}\sum_{k=0}^{K_1} f_2(k,m)\mathbf{x}_{k,m}\sum_{l=K_1-k}^{\infty}\beta^*_l(s)\right].$$

By same method as that in Choi, Kim, Choi, and Sung [3], we obtain that

$$\beta^*_n(s) = \frac{1}{\mu}\left[\sum_{k=0}^{n} A_{k,\cdot}R_{n-k}(s) - G^*(s)R_n(s)\right],$$

where $R_n(s) = (sI + C + D_2)^{-1}[(-D_1)(sI + C + D_2)^{-1}]^n$.
Substituting $\beta^*_n(s)$ into above equation $\alpha^*_n(s)$, and then putting $s = 0$, after some algebraic manipulation we obtain the queue length probabilities $\mathbf{y}^1_n = \alpha^*_n(0)$:
For $0 \le n < K_1$,

$$\mathbf{y}^1_n = \frac{1}{E}\left[\mathbf{x}_{0,0}(-C^{-1})D\left\{\sum_{l=0}^{n} A_{l,\cdot}(C+D_2)^{-1}\{D_1(-(C+D_2)^{-1})\}^{n-l}\right.\right.$$
$$\left. -(C+D_2)^{-1}\{D_1(-(C+D_2)^{-1})\}^{n}\right\}$$
$$+\sum_{m=0}^{K_2}\sum_{k=1}^{n+1} f_1(k,m)\mathbf{x}_{k,m}\left\{\sum_{l=0}^{n-k+1} A_{l,\cdot}(C+D_2)^{-1}\{D_1(-(C+D_2)^{-1})\}^{n-k+1-l}\right.$$
$$\left. -(C+D_2)^{-1}\{D_1(-(C+D_2)^{-1})\}^{n-k+1}\right\}$$
$$+\sum_{m=1}^{K_2}\sum_{k=0}^{n} f_2(k,m)\mathbf{x}_{k,m}\left\{\sum_{l=0}^{n-k} A_{l,\cdot}(C+D_2)^{-1}\{D_1(-(C+D_2)^{-1})\}^{n-k-l}\right.$$
$$\left.\left. -(C+D_2)^{-1}\{D_1(-(C+D_2)^{-1})\}^{n-k}\right\}\right],$$

and

$$\mathbf{y}^1_{K_1} = \pi - \sum_{k=0}^{K_1-1}\mathbf{y}^1_k - \mathbf{y}_0.$$

The stationary queue length probability of buffer II is defined as

$$y^2_n(j) = \lim_{t\to\infty} Pr\{N_2(t) = n, J(t) = j, \xi(t) = 1\},\ 1 \le j \le M,$$
$$\mathbf{y}^2_n = (y^2_n(1), \ldots, y^2_n(M)), \qquad 0 \le n \le K_2.$$

By the same method as in the previous case, we obtain the queue length distribution of buffer II.

For $0 \le n < K_2$,

$$\begin{aligned}\mathbf{y}_n^2 =& \frac{1}{E}\Big[\mathbf{x}_{0,0}(-C^{-1})D\Big\{\sum_{l=0}^{n} A_{\cdot,l}(C+D_1)^{-1}\{D_2(-(C+D_1)^{-1})\}^{n-l} \\ &\qquad - (C+D_1)^{-1}\{D_2(-(C+D_1)^{-1})\}^{n}\Big\} \\ &+ \sum_{k=1}^{K_1}\sum_{m=0}^{n} f_1(k,m)\mathbf{x}_{k,m}\Big\{\sum_{l=0}^{n-m} A_{\cdot,l}(C+D_1)^{-1}\{D_2(-(C+D_1)^{-1})\}^{n-m-l} \\ &\qquad - (C+D_1)^{-1}\{D_2(-(C+D_1)^{-1})\}^{n-m}\Big\} \\ &+ \sum_{k=0}^{K_1}\sum_{m=1}^{n+1} f_2(k,m)\mathbf{x}_{k,m}\Big\{\sum_{l=0}^{n-m+1} A_{\cdot,l}(C+D_1)^{-1}\{D_2(-(C+D_1)^{-1})\}^{n-m+1-l} \\ &\qquad - (C+D_1)^{-1}\{D_2(-(C+D_1)^{-1})\}^{n-m+1}\Big\}\Big],\end{aligned}$$

and

$$\mathbf{y}_{K_2}^2 = \pi - \sum_{k=0}^{K_2-1} \mathbf{y}_k^2 - \mathbf{y}_0.$$

Using the queue length distribution of each buffer at an arbitrary time, we obtain the following performance measures:

a. The loss probabilities for type-1 (P_{loss}^1) and type-2 (P_{loss}^2) customers:

$$P_{\text{loss}}^1 = \frac{\mathbf{y}_{K_1}^1 D_1\mathbf{e}}{\lambda_1^*}, \qquad P_{\text{loss}}^2 = \frac{\mathbf{y}_{K_2}^2 D_2\mathbf{e}}{\lambda_2^*}.$$

b. The mean queue lengths of buffer I (L^1) and buffer II (L^2):

$$L^1 = \sum_{i=0}^{K_1} i\mathbf{y}_i^1\mathbf{e}, \qquad L^2 = \sum_{i=0}^{K_2} i\mathbf{y}_i^2\mathbf{e}.$$

By the Little law, we obtain the mean waiting time of a customer in each buffer:

c. The mean waiting times of type-1 (W^1) and type-2 (W^2) customers:

$$W^1 = \frac{L^1}{\lambda_1^*(1-P_{\text{loss}}^1)}, \qquad W^2 = \frac{L^2}{\lambda_2^*(1-P_{\text{loss}}^2)}.$$

4. Numerical examples

In this section, we present some numerical examples. First, by specifying the service scheduling function $f_k(i,j)(k=1,2)$, we show that the existing many scheduling schemes can be given as special cases of our proposed model, and compare the performance measures of the schemes. Next, we investigate the effects of system when the arrivals are MAPs and Poisson processes.

The following schemes are given as special cases of our model:

a. Head of Line priority scheduling scheme (HOL):

$$f_2(i,j) = 1, \quad \text{if } j > 0.$$

b. Shortest Job First scheduling scheme (SJF):

$$f_1(i,j) = \begin{cases} 1, & \text{if } i \le j, \\ 0, & \text{otherwise.} \end{cases}$$

c. Longest Job First scheduling scheme (LJF):

$$f_1(i,j) = \begin{cases} 1, & \text{if } i > j, \\ 0, & \text{otherwise.} \end{cases}$$

d. Bernoulli scheduling scheme (Bernoulli):

$$f_1(i,j) = p, \quad \text{if } i > 0,\ j > 0.$$

e. QLT scheduling scheme with threshold T on buffer I (QLT):

$$f_1(i,j) = \begin{cases} 1, & \text{if } i > T, \\ 0, & \text{otherwise.} \end{cases}$$

f. QLT scheduling scheme with Bernoulli schedule (QLT with Bernoulli):

$$f_1(i,j) = \begin{cases} 1, & \text{if } i > T, \\ p, & \text{otherwise.} \end{cases}$$

We take the following parameters for the numerical examples. First, MAPs with $(\overline{C}_k, \overline{D}_k)$, $k = 1, 2$, are assumed by

$$\overline{C}_1 = \begin{bmatrix} -\delta_{11} - \lambda_{11} & \delta_{11} \\ \delta_{12} & -\delta_{12} - \lambda_{12} \end{bmatrix}, \quad \overline{D}_1 = \begin{bmatrix} \lambda_{11} & 0 \\ 0 & \lambda_{12} \end{bmatrix},$$

$$\overline{C}_2 = \begin{bmatrix} -\delta_{21} - \lambda_{21} & \delta_{21} \\ \delta_{22} & -\delta_{22} - \lambda_{22} \end{bmatrix}, \quad \overline{D}_2 = \begin{bmatrix} \lambda_{21} & 0 \\ 0 & \lambda_{22} \end{bmatrix}.$$

The service time of all customers is assumed to be constant and equal to one. We also assume that the type-2 customers are more delay-sensitive than the type-1 customers. Thus we take the buffer sizes as $K_1 = 10$ and $K_2 = 5$. In all numerical examples, we furthermore set the threshold $T = 5$, the Bernoulli probability $p = 0.5$, and $\delta_{11} = \delta_{12} = 0.1$, $\delta_{21} = \delta_{22} = 0.05$, $\lambda_1^* = \lambda_2^*$ and $\lambda_{12}/\lambda_{11} = \lambda_{22}/\lambda_{21} = 6$.

Figs 1 and 2 display the loss probability and the mean waiting time of the type-1 customer, respectively, as a function of the total effective arrival rate $\lambda_1^* + \lambda_2^*$. From the figures, we can see that QLT, QLT with Bernoulli and LJF scheduling schemes deliver good performance for the loss probability and the mean waiting time.

Figs 3 and 4 display the loss probability and the mean waiting time of the type-2 customer, respectively, as a function of the total effective arrival rate $\lambda_1^* + \lambda_2^*$. From the figures, we can observe that HOL and SJF scheduling schemes give good performance for the loss probability and the mean waiting time. Finally from Figs 1-4, according to the characteristic of supported traffics, an appropriate scheduling scheme must be selected to satisfy Quality of Service (QoS) of the traffic.

Next we compare the performance measures when the arrivals are MAPs and Poisson processes. The cases of HOL, LJF, and QLT with Bernoulli are compared. As a function of the total effective arrival rate $\lambda_1^* + \lambda_2^*$, the loss probability and the mean waiting time for HOL are given in Figs 5 and 6, those for LJF in Figs 7 and 8, and those for QLT with Bernoulli in Figs 9 and 10. From the figures, we can see that the loss probability and the mean waiting time when the arrival follows MAP are larger than those when the arrival follows Poisson process. These results are attributed to the bursty and time-correlated properties. Finally, these characteristics of source traffic have an important effect on the performance of system.

In conclusion, we compared the performance measures of diverse service scheduling schemes in this paper. We expect that these results will be helpful to system designers in selecting an appropriate service scheduling scheme and thereby optimizing their system. Furthermore we also compared MAPs and Poisson arrivals, and confirmed that the burstiness of source traffic substantially affects the performance of the system. Therefore, it is very important to precisely model source traffic.

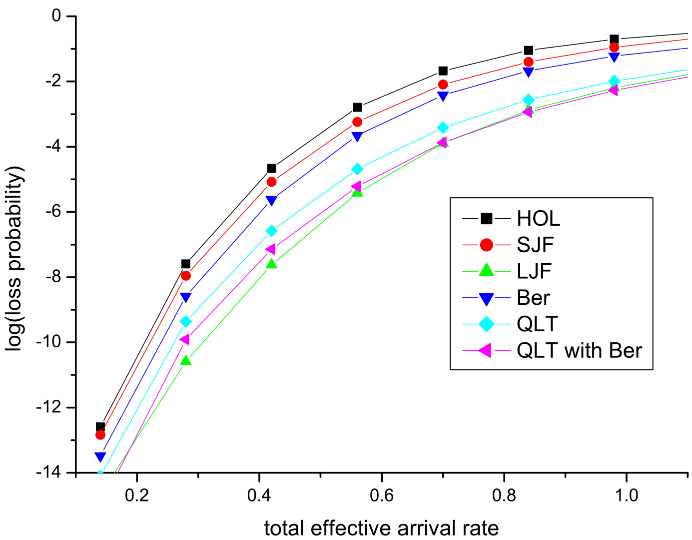

FIGURE 1. total effective arrival rate vs log(loss probability) for type-1 customer

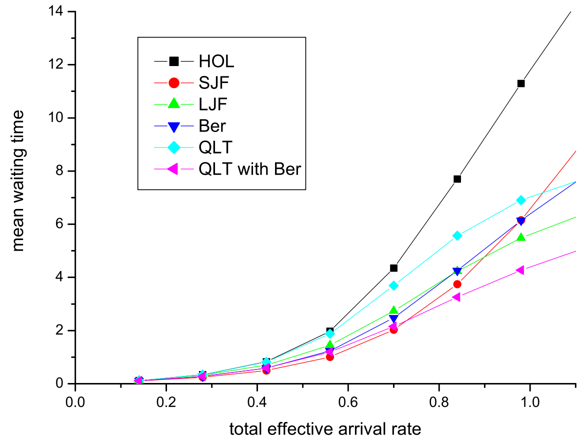

FIGURE 2. total effective arrival rate vs mean waiting time for type-1 customer

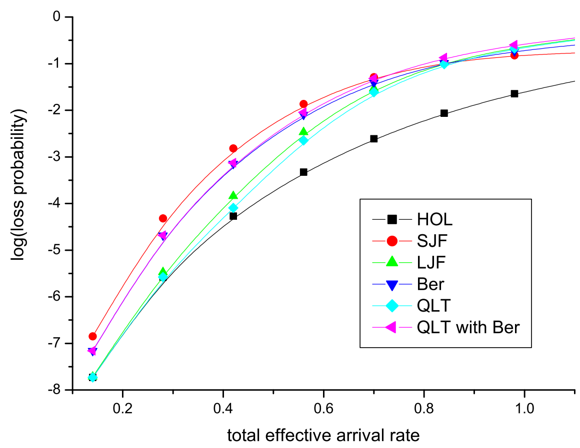

FIGURE 3. total effective arrival rate vs log(loss probability) for type-2 customer

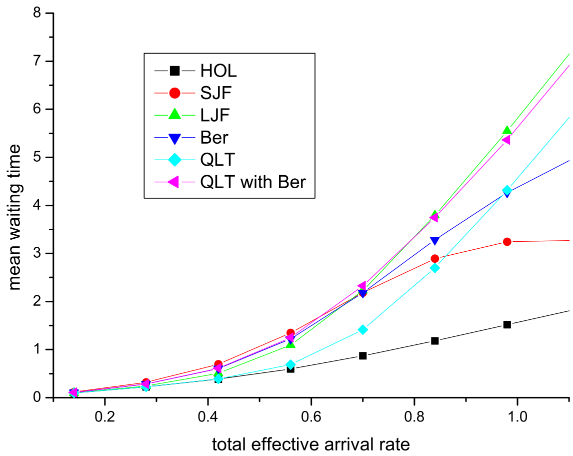

FIGURE 4. total effective arrival rate vs mean waiting time for type-2 customer

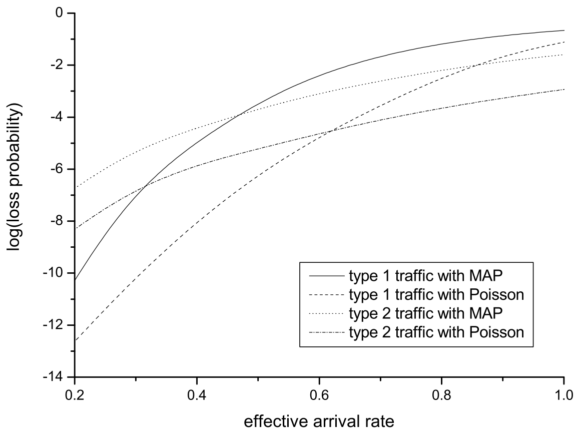

FIGURE 5. effective arrival rate vs log(loss probability) for HOL

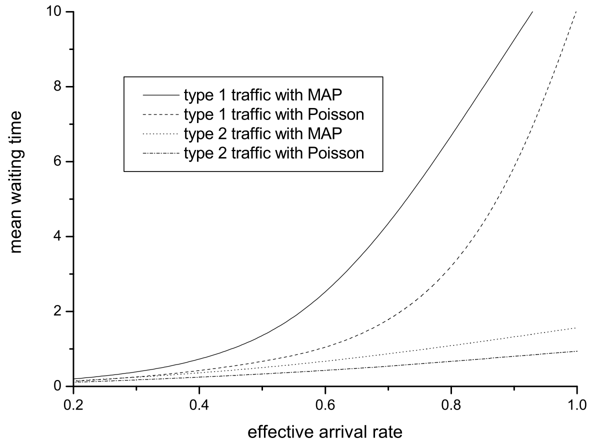

FIGURE 6. effective arrival rate vs mean waiting time for HOL

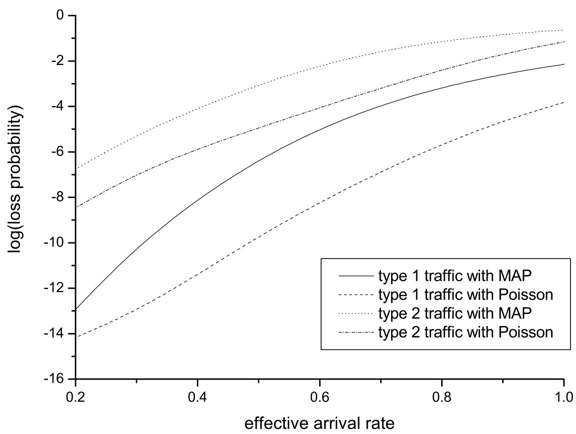

FIGURE 7. effective arrival rate vs log(loss probability) for SJF

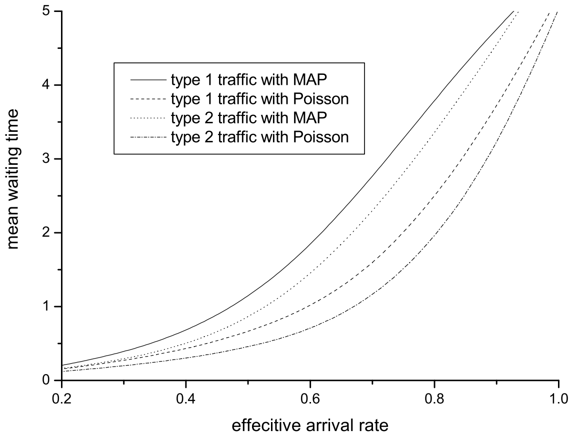

FIGURE 8. effective arrival rate vs mean waiting time for SJF

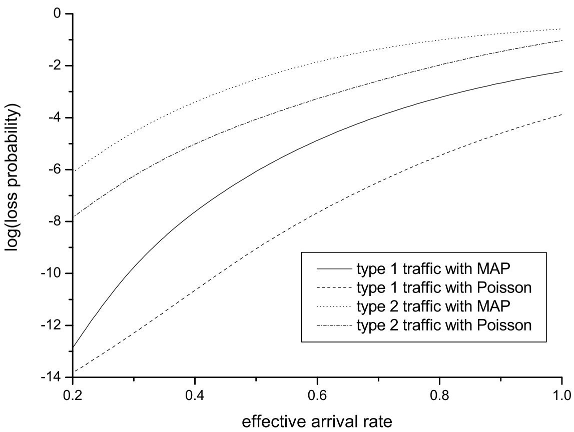

FIGURE 9. effective arrival rate vs log(loss probability) for QLT with Bernoulli

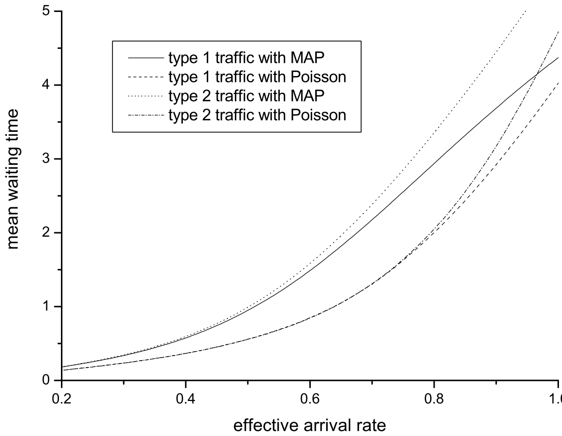

FIGURE 10. effective arrival rate vs mean waiting time for QLT with Bernoulli

References

[1] S. Asmussen and G. Koole, *Marked point processes as limits of Markovian arrival streams*, J. Appl. Probab. **30** (1993), no. 2, 365–372.

[2] D. I. Choi, B. D. Choi, and D. K. Sung, *Performance analysis of priority leaky bucket scheme with queue-length-threshold scheduling policy*, IEE Proc. Comm. **145** (1998), no. 6, 395–401.

[3] B. D. Choi, Y. C. Kim, D. I. Choi, and D. K. Sung, *An analysis of M, MMPP/G/1 queue with QLT scheduling policy and Bernoulli schedule*, IEICE Tran. on Comm. **E81-B** (1998), no. 1, 13–22.

[4] D. I. Choi, T. S. Kim, and S. M. Lee, *Analysis of a queueing system with a general service scheduling function, with applications to telecommunication network traffic control*, European J. Oper. Res. **178** (2007), no. 2, 463–471.

[5] D. I. Choi and Y. Lee, *Performance analysis of a dynamic priority queue for traffic control of bursty traffics in ATM networks*, IEE Proc. Commun. **148** (2001), no. 6, 395–401.

[6] E. Cinlar, *Markov renewal theory*, Advances in Appl. Probability **1** (1969), 123–187.

[7] S. S. Fratini, *Analysis of a dynamic priority queue*, Comm. Statist. Stochastic Models **6** (1990), no. 3, 415–444.

[8] A. Graham, *Kronecker Products and Matrix Calculus*: *with applications*, Ellis Horwood Series in Mathematics and its Applications. Ellis Horwood Ltd., Chichester; Halsted Press [John Wiley & Sons, Inc.], New York, 1981.

[9] H. Heffes and D. M. Lucantoni, *A Markov-modulated characterization of packetized voice and data traffic and related statistical multiplexer performance*, IEEE J. Select. Areas Commun. **6** (1986), 856–868.

[10] C. Knessl, D. I. Choi, and C. Tier, *A dynamic priority queue model for simultaneous service of two traffic types*, SIAM J. Applied Mathematics **63** (2002), no. 2, 398–422.

[11] D. M. Lucantoni, K. S. Meier-Hellstern, and M. F. Neuts, *A single-server queue with server vacations and a class of nonrenewal arrival processes*, Adv. in Appl. Probab. **22** (1990), no. 3, 676–705.

[12] H. Nassar and H. Al. Mahdi, *Queueing analysis of an ATM multimedia multiplexer with nonpreemptive priority*, IEE Proc. Commun. **150** (2003), no. 3, 189–196.

[13] A. Sugahara, T. Takine, Y. Takahashi, and T. Hasegawa, *Analysis of a nonpreemptive priority queue with SPP arrivals of high class*, Performance Evaluation **21** (1995), no. 3, 215–238.

Doo Il Choi
Department of applied Mathematics
Halla University
Kangwon 220-712, Korea
E-mail address: dichoi@halla.ac.kr

Sang Min Lee
Converged Service Development Department
Platform Laboratory, KT
Daejon 305-811, Korea