Journal of the Operations Research
Society of Japan
Vol. 36, No. 1, March 1993

ON DISCRETE-TIME SINGLE-SERVER QUEUES WITH MARKOV MODULATED BATCH BERNOULLI INPUT AND FINITE CAPACITY

Toshiaki Tsuchiya
Tokyo Institute of Technology

Yoshitaka Takahashi
NTT Laboratories

(Received March 30, 1992; Revised December 10,1992)

Abstract This paper investigates a discrete-time system with Markov modulated batch Bernoulli process (MMBP) inputs, general service times, and a finite capacity waiting room (queue). The FIFO and space priority service rules are considered. The main motivation for studying this type of queue is its potential applicability to time-slotted communication systems in broadband ISDN (B-ISDN). Applying the supplementary variable technique and the matrix analysis approach, we obtain the steady-state distributions of the number of customers in the system as well as loss probabilities under individual service rules.

1. Introduction

There are a lot of practical systems that operate on a discrete-time basis. Classical examples are synchronous communication channels, e.g. slotted ALOHA [13]. More recent examples are packet switching systems and ATM (Asynchronous Transfer Mode) systems in broadband ISDN's (Integrated Service Digital Networks), see e.g. [6, 7]. In such systems, all events (arrivals and departures) are allowed only at regularly spaced points in time. Performance issues in these systems then necessitate discrete-time queues.

A common feature in recent communication systems is that we frequently encounter a single-server queueing situation with priority or without priority where the input process is not renewal but correlated. For example, it is well known that the input process of packetized voice traffic at a statistical multiplexer (at a single-server) does not form a Poisson process (nor renewal) but bursty and correlated [11]. In order to capture the effect of correlated input process, a discrete-time version of Markov modulated Poisson process (MMPP [11]) called *Markov modulated batch Bernoulli process* (MMBP) has been proposed and the statistical characterization measures of two-state MMBP, so called *switched batch Bernoulli process* (SBBP), has been investigated in Hashida et al. [10]. This feature should motivate the study of discrete-time queues with MMBP or SBBP inputs.

Several previous works on the discrete-time single-server queues with MMBP or SBBP inputs are worth mentioning. Hashida et al. [10] analyzed the $SBBP/G/1$ queue and found the probability generating function (pgf) of the number of customers in the system. Daigle et al. [3] have analyzed the $MMBP/D/1$ queue via the algorithmic approach. Subsequently, Hashida et al. [9], Khamisky et al. [12] and Stavrakakis [17] have analyzed the $SBBP/D/1$ queue with the head-of-the-line priority service rule via the embedded Markov chain approach. In these results, however, the capacity of the waiting room (queue) has been assumed to be infinite. These approaches [3, 9, 10, 12, 17] cannot be directly applied to finite capacity queues.

Recently, under renewal assumptions, there has been much interest in analyzing the discrete-time single-server queues with finite capacity. Gravey et al. [7] has analyzed a

$Geo/D/1/K$ queue, where Geo denotes the Bernoulli input and K denotes the queue capacity (the maximum number of customers allowed in the system). Tran-Gia et al. [20] and Murata et al. [15] have respectively analyzed the batch input $GI^X/D/1/K$ queue and the mixed input $Geo + GI/D/1/K$ queue via the Fast Fourier Transform (FFT) technique. However, there seems to be very few literature on the correlated input discrete-time queues with finite capacity. The primary purpose of this paper is to provide an approach to evaluate the loss (or sometimes called as overflow) probabilities in a correlated input discrete-time queue with finite capacity.

In the followings, the ordinary priority in the queueing literature, e.g. head-of-the-line [9, 12], will be called as *time priority*, since customer classes are prioritized subject to the delay times. For the previous works on discrete-time queues with time priority, see Takahashi et al. [19] and references therein. When the customer classes are prioritized subject to the allowed loss probabilites rather than the delay times, however, it seems that the time priority rules are not so useful. The *space priority* service rule, namely push-out [18] or buffer reservation scheme [6, 14] which will be described precisely later, has been appearing in the ATM systems and has been analyzed via the continuous-time queueing models [6, 14, 18]. Under space priority rule, we can realize different loss probabilities for different classes of customers.

In this paper, we consider a discrete-time $MMBP/G/1/K$ finite capacity queue without and with space priority. The queueing model has a potential applicability to the recently-developed time-slotted systems. In Section 2, we describe the MMBP model by straightforwardly extending the one of Hashida et al. [10]. In Section 3, we subsequently treat the $MMBP/G/1/K$ under FIFO service rule. The approach taken here is based on the supplementary variable technique as in [10] and the matrix analysis method, which enables us to obtain the steady-state distribution of the number of customers in the system. Section 4 is devoted to the analysis of a two-class $MMBP/G/1/K$ space priority queue. For either the push-out or buffer reservation scheme, we respectively derive the individual class loss probabilities, by using the results in Section 3 together with the qualitative results such as the conservation law (for loss probabilities) and the conditional GASTA (Geometric Arrivals See Time Averages) property [21]. We also present numerical results showing a performance difference between the push-out and buffer reservation schemes in Section 5.

It should be noted that the input process treated here covers the previously analyzed discrete-time queue input models in literature [3, 7, 10, 9, 12, 13, 17, 19].

2. Description of the system

We assume the time axis is divided into a sequence of equal-length intervals $[0,1), [1,2), \cdots, [n-1, n), \cdots$. We call interval $[n-1, n)$ as the *n-th slot*. The state of the system changes only around the slot boundaries. Customers arrive in a batch at the beginning of a slot. When a service is completed at a slot, the served customer leaves the system at the end of the slot, and then the next customer is served from the beginning of the next slot. Hence, the service times can be counted by the number of slots. The residual service times are assumed to decrease by 1 at the end of a slot.

In order to represent correlated and bursty type traffic, we introduce a *Markov modulated batch Bernoulli process* (MMBP). It is a slightly extended one of a switched batch Bernoulli process (SBBP) introduced by Hashida et al. [10]. The MMBP is a bivariate process $\{Y_n, X_n\}$. The process $\{Y_n\}$ is an M-state discrete-time Markov chain representing the state of the arrival process during the n-th slot, and X_n is the number (batch size) of arrivals to the system at the beginning of the n-th slot. We shall refer to $\{Y_n\}$ as the *modulating Markov chain* (MMC) and its state the *phase*. The MMC is assumed to be ergodic and its transition probability matrix is denoted by

$$W \equiv \begin{pmatrix} \overline{w}_1 & w_{12} & \cdots & w_{1M} \\ w_{21} & \overline{w}_2 & \cdots & w_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ w_{M1} & w_{M2} & \cdots & \overline{w}_M \end{pmatrix} \tag{2.1}$$

where $w_{ij} \equiv P\{Y_n = j \mid Y_{n-1} = i\}$ and $\overline{w}_i \equiv P\{Y_n = i \mid Y_{n-1} = i\}$.

The batch size X_n depends on the phase Y_n of the MMC. We represent the distribution of it as follows.

$$a_i(k) \equiv P\{X_n = k \mid Y_n = i\} \text{ and } \overline{a}_i(k) \equiv P\{X_n \geq k \mid Y_n = i\} = 1 - \sum_{l=0}^{k-1} a_i(l). \tag{2.2}$$

To avoid a trivial case, we assume that there exists some i such that $a_i(0) < 1$. The mean batch size is then given as

$$E(X_n) = \sum_{i=1}^{M} E(X_n \mid Y_n = i) P\{Y_n = i\} = \sum_{i=1}^{M} \alpha_i \omega_i, \tag{2.3}$$

where α_i is the conditional mean batch size $E(X_n \mid Y_n = i)$, and $\omega = (\omega_1, \cdots, \omega_M)$ is the steady-state probability vector of the MMC.

The service times of individual customers are assumed to be independent, identically distributed random variables, which we denote by S. We assume that S has a mean $\beta = \sum_{l=1}^{\infty} l b_l$ where $b_l = P\{S = l\}$.

The capacity of the system is limited to K places for customers waiting and in service. When the arriving batch is larger in size than the number of unoccupied places, it fills the available positions and the remainder of the batch are rejected (a partially rejected model, see e.g. Baba [1]).

To describe the state of the system, let N_n be the number of customers in the system and R_n the residual service time of the customer in service during the n-th slot. If $N_n = 0$, then R_n is set to be equal to 0. The triple-variate process $\{N_n, R_n, Y_n\}$ then forms a Markov chain. We restrict our attention to the steady-state behavior, and express generic random variables by omitting subscript n.

3. The $MMBP/G/1/K$ queue without priority

In this section, we treat an $MMBP/G/1/K$ queue with a single class of customers. We use the term *no priority* to distinguish this system from those we consider in Section 4. We analyze it through the supplementary variable technique as in Hashida et al. [10].

3.1 Basic difference equations

Let $p_i(k, l)$ be the steady-state probability $P\{N = k, R = l, Y = i\}$, that is, the number of customers in the system is k, the residual service time is l and the phase is i. When $N = 0$, we write $p_i(0)$ instead of $p_i(k, l)$.

Let us consider a transition to state (k, l, i) during $(n + 1)$-st slot from possible states during n-th slot. Then we have

$$p_i(0) = \{p_i(0) + p_i(1,1)\} a_i(0) \overline{w}_i + \sum_{\substack{j=1 \\ j \neq i}}^{M} [\{p_j(0) + p_j(1,1)\} a_j(0) w_{ji}], \tag{3.1}$$

$$(3.2)\quad p_i(k,l) = b_l\Bigg[\{p_i(0)a_i(k)+\sum_{m=1}^{k+1}p_i(m,1)a_i(k+1-m)\}\overline{w}_i + \sum_{\substack{j=1\\ j\neq i}}^{M}\{p_j(0)a_j(k)+\sum_{m=1}^{k+1}p_j(m,1)a_j(k+1-m)\}w_{ji}\Bigg] + \sum_{m=1}^{k}\Big\{p_i(m,l+1)a_i(k-m)\overline{w}_i + \sum_{\substack{j=1\\ j\neq i}}^{M}p_j(m,l+1)a_j(k-m)w_{ji}\Big\}$$
$$(k=1,\cdots,K-1).$$

When $k = K$, arriving customers may be rejected. So,

$$(3.3)\quad p_i(K,l) = b_l\Bigg[\{p_i(0)\overline{a}_i(K)+\sum_{m=1}^{K}p_i(m,1)\overline{a}_i(K+1-m)\}\overline{w}_i + \sum_{\substack{j=1\\ j\neq i}}^{M}\{p_j(0)\overline{a}_j(K)+\sum_{m=1}^{K}p_j(m,1)\overline{a}_j(K+1-m)\}w_{ji}\Bigg] + \sum_{m=1}^{K}\Big\{p_i(m,l+1)\overline{a}_i(K-m)\overline{w}_i + \sum_{\substack{j=1\\ j\neq i}}^{M}p_j(m,l+1)\overline{a}_j(K-m)w_{ji}\Big\}.$$

3.2 The queue-length distribution at arbitrary epochs

Now, let us consider the distribution of the number of customers in the system. For brevity, we call this the *queue-length* distribution in this paper. In order to rewrite the equations (3.1), (3.2) and (3.3) in matrix forms, we introduce some vectors and matrices.

For $k = 1,2,\cdots,K$ and $l = 1,2,\cdots$, let $\mathbf{p}(k,l) \equiv \Big(p_1(k,l),\cdots,p_M(k,l)\Big)$ and $\mathbf{p}(k) \equiv \sum_{l=1}^{\infty}\mathbf{p}(k,l)$. The i-th element $p_i(k)$ of $\mathbf{p}(k)$ is then a joint probability that there are k customers in the system and the phase is i. We also introduce KM-dimensional vectors

$$(3.4)\qquad \mathbf{r}(l) \equiv \Big(\mathbf{p}(1,l),\mathbf{p}(2,l),\cdots,\mathbf{p}(k,l)\Big),\qquad l=1,2,\cdots$$

and

$$(3.5)\qquad \tilde{\mathbf{r}}(1) \equiv \Big(\mathbf{p}(0)+\mathbf{p}(1,1),\mathbf{p}(2,1),\cdots,\mathbf{p}(K,1)\Big).$$

Note that $\tilde{\mathbf{r}}(1)$ is identical with $\mathbf{r}(1)$ except for the first subvector. We write diagonal matrices $\mathbf{A}_k \equiv \mathrm{diag}[a_1(K),a_2(K),\cdots,a_M(k)]$ and $\overline{\mathbf{A}}_k \equiv \mathrm{diag}[\overline{a}_1(k),\overline{a}_2(k),\cdots,\overline{a}_M(K)]$.

Equations (3.2) and (3.3) then can be expressed as

$$(3.6)\qquad \mathbf{r}(l) = b_l\tilde{\mathbf{r}}(1)\tilde{\mathbf{C}}\mathbf{V} + \mathbf{r}(l+1)\mathbf{A}\mathbf{V},\qquad l=1,2,\cdots$$

where

$$(3.7)\qquad \mathbf{A} \equiv \begin{pmatrix} \mathbf{A}_0 & \mathbf{A}_1 & \cdots & \mathbf{A}_{K-2} & \overline{\mathbf{A}}_{K-1} \\ & \mathbf{A}_0 & \cdots & \mathbf{A}_{K-3} & \overline{\mathbf{A}}_{K-2} \\ & & \ddots & \vdots & \vdots \\ & 0 & & \mathbf{A}_0 & \overline{\mathbf{A}}_1 \\ & & & & \mathbf{I} \end{pmatrix}, V \equiv \begin{pmatrix} \mathbf{W} & & & 0 \\ & \mathbf{W} & & \\ & & \ddots & \\ & 0 & & \mathbf{W} & \\ & & & & \mathbf{W} \end{pmatrix},$$

and

$$\tilde{\mathbf{C}} \equiv \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_{K-1} & \overline{\mathbf{A}}_K \\ \mathbf{A}_0 & \mathbf{A}_1 & \cdots & \mathbf{A}_{K-2} & \overline{\mathbf{A}}_{K-1} \\ & \ddots & \ddots & \vdots & \vdots \\ & 0 & \ddots & \mathbf{A}_1 & \overline{\mathbf{A}}_2 \\ & & & \mathbf{A}_0 & \overline{\mathbf{A}}_1 \end{pmatrix}. \tag{3.8}$$

Note that matrices $\mathbf{A}$ and $\mathbf{V}$ are stochastic (and so is $\mathbf{AV}$), and that $\tilde{\mathbf{C}}$ is substochastic. Expressing (3.1) in a matrix form, we have

$$\mathbf{p}(0) = \mathbf{p}(0)\mathbf{A}_0\mathbf{W} + \mathbf{p}(1,1)\mathbf{A}_0\mathbf{W}. \tag{3.9}$$

From Theorem 2.1 and Corollary 3 of Seneta [16], the inverse of $\mathbf{I} - \mathbf{A}_0\mathbf{W}$ exists and is non-negative. So, (3.9) becomes

$$\mathbf{p}(0) + \mathbf{p}(1,1) = \mathbf{p}(1,1)\{\mathbf{I} - \mathbf{A}_0\mathbf{W}\}^{-1}. \tag{3.10}$$

Substituting (3.10) into (3.5) gives that

$$\tilde{\mathbf{r}}(1) = \mathbf{r}(1)\mathbf{I}_0, \text{ where } \mathbf{I}_0 \equiv \begin{pmatrix} \{\mathbf{I} - \mathbf{A}_0\mathbf{W}\}^{-1} & & & 0 \\ & \mathbf{I} & & \\ 0 & & \ddots & \\ & & & \mathbf{I} \end{pmatrix}. \tag{3.11}$$

If we write a stochastic matrix $\mathbf{C} \equiv \mathbf{I}_0\tilde{\mathbf{C}}$, (3.6) becomes

$$\mathbf{r}(l) = b_l\mathbf{r}(1)\mathbf{CV} + \mathbf{r}(l+1)\mathbf{AV}, \qquad l = 1, 2, \cdots \tag{3.12}$$

We are interested in the queue-length distribution, namely, M-dimensional vector $\mathbf{p}(0)$ and KM-dimensional vector $\pi \equiv (\mathbf{p}(1), \mathbf{p}(2), \cdots, \mathbf{p}(K)) = \sum_{l=1}^{\infty} \mathbf{r}(l)$. We first derive $\mathbf{r}(1)$ and $\mathbf{p}(0)$ and then get π from $\mathbf{r}(1)$. An iterated use of (3.12) leads us to

$$\mathbf{r}(l) = \mathbf{r}(1)\mathbf{CV}\sum_{n=l}^{\infty} b_n(\mathbf{AV})^{n-l}. \tag{3.13}$$

For convenience, we write $\mathbf{D}(l) = \mathbf{CV}\sum_{n=l}^{\infty} b_n(\mathbf{AV})^{n-l}$. If we let $l = 1$, (3.13) becomes a vector equation $\mathbf{r}(1) = \mathbf{r}(1)\mathbf{D}(1)$. Thus, $\mathbf{r}(1)$ is proportional to the invariant probability vector $\mathbf{x} = (\mathbf{x}_1, \cdots, \mathbf{x}_K)$ of $\mathbf{D}(1)$.

Let us consider the structure of $\mathbf{D}(1)$. Matrix $\mathbf{A}$ is stochastic and block upper triangular and so is $(\mathbf{AV})^n$. Since $\{b_l\}$ is a distribution, the series $\sum_{n=1}^{\infty} b_n(\mathbf{AV})^{n-1}$ converges and the limit matrix is stochastic and block upper triangular, too. On the other hand, stochastic matrix $\mathbf{C}$ has a similar structure to $\tilde{\mathbf{C}}$ in (3.8) and so do $\mathbf{CV}$. As a result of further consideration, we know that stochastic matrix $\mathbf{D}(1)$ has the form

$$\mathbf{D}(1) = \begin{pmatrix} \mathbf{E}_1 & \mathbf{E}_2 & \cdots & \mathbf{E}_{K-1} & \mathbf{F}_K \\ \mathbf{D}_0 & \mathbf{D}_1 & \cdots & \mathbf{D}_{K-2} & \mathbf{F}_{K-1} \\ & \ddots & \ddots & \vdots & \vdots \\ & 0 & \ddots & \mathbf{D}_1 & \mathbf{F}_2 \\ & & & \mathbf{D}_0 & \mathbf{F}_1 \end{pmatrix}. \tag{3.14}$$

Moreover, the Markov chain represented by $\mathbf{D}(1)$ is seen to be ergodic from the assumption of ergodicity of $\{Y_n\}$. Thus, if the inverse of

$$\mathbf{D}_0 = \mathbf{A}_0\mathbf{W}\sum_{n=1}^{\infty} b_n(\mathbf{A}_0\mathbf{W})^{(n-1)} \tag{3.15}$$

exists, we can determine $\mathbf{x}$ uniquely from the vector equations

$$\mathbf{x}_{k+1} = \{\mathbf{x}_k - \mathbf{x}_1\mathbf{E}_k - \sum_{n=2}^{k}\mathbf{x}_n\mathbf{D}_{k+1-n}\}\mathbf{D}_0^{-1} \qquad k = 1,2,\cdots,K-1, \tag{3.16}$$

$$\mathbf{x}_K = \sum_{n=1}^{K}\mathbf{x}_n\mathbf{F}_{K+1-n}. \tag{3.17}$$

Letting $k = 1,2,\cdots,K-1$, we have iteratively $\mathbf{x}_2,\mathbf{x}_3,\cdots,\mathbf{x}_K$ in terms of $\mathbf{x}_1$ from (3.16) with the consequence that (3.17) becomes the vector equation for $\mathbf{x}_1$. So, we get $\mathbf{x}$ completely by solving (3.17) for $\mathbf{x}_1$.

From $\mathbf{x}$, we know $\mathbf{r}(1)$ up to a multiple constant, say, c. Note that we also have $\mathbf{p}(0)$ with unkown c, since $\mathbf{p}(0)$ can be expressed in terms of the first subvector $\mathbf{p}(1,1)$ of $\mathbf{r}(1)$ from (3.9) or (3.10).

The multiple constant c can be obtained from the Little's Law:

$$\beta\mathbf{r}(1)\mathbf{e} = 1 - \mathbf{p}(0)\mathbf{e}, \ \mathbf{e} = (1,1,\cdots,1)^T. \tag{3.18}$$

Since the service time of a customer is at least 1, $\mathbf{r}(1)\mathbf{e} = c$ is the rate of departing customers from the system, and is the same as the rate of entering customers to the system. Hence, the left-hand side of (3.18), $\beta\mathbf{r}(1)\mathbf{e}$, equals to the probability that the server is busy.

Next, let us consider about π. Summing up (3.13) over l, we get

$$\pi = \mathbf{r}(1)\mathbf{CV}\sum_{n=1}^{\infty}\bar{b}_n(\mathbf{AV})^{n-1}. \tag{3.19}$$

We also write $\mathbf{D} = \mathbf{CV}\sum_{n=1}^{\infty}\bar{b}_n(\mathbf{AV})^{n-1}$. Since $\{b_l\}$ has mean $\beta < \infty$, the series in (3.19) converges and $\mathbf{D}$ is well defined. Matrix $\mathbf{D}$ has a similar structure to $\mathbf{D}(1)$. Since the vector $\mathbf{r}(1)$ is already computed, we immediately have π. This concludes the computation of the queue-length distribution.

Once we get π and $\mathbf{p}(0)$, we can evaluate performance measures. For example, the loss probability B is calculated as follows. Applying the Little's law at the server, we have

$$E(X_n)\{1 - B\}\beta = 1 - \mathbf{p}(0)\mathbf{e}. \tag{3.20}$$

The right-hand side represents the mean number of customers in the server, while the left-hand side represents the throughput multiplied by the mean service time. Thus, we have B from (2.3) and (3.20) as

$$B = 1 - \frac{1 - \mathbf{p}(0)\mathbf{e}}{(\omega_1\alpha_1 + \cdots + \omega_M\alpha_M)\beta}. \tag{3.21}$$

4. The $MMBP/G/1/K$ queue with space priority

We consider two classes of customers, such as voices and data in communication networks, having different requirements for quality of service. We assume class-1 customers are loss-insensitive but delay-sensitive, while class-2 customers are loss-sensitive but delay-insensitive. In order to ensure different requirements for individual classes, we treat two schemes of space priority called push-out and buffer reservation. In both schemes, the buffer is divided into two parts, sized K_1- including the server and $K_2 = K - K_1$. Buffer management policies change according to the number of customers in the system, as we describe in Section 4.1 (push-out) and in Section 4.2 (buffer reservation).

In the sequel, an arriving batch consists of both class-1 and class-2 customers and is sorted in advance so that class-1 customers arrive before class-2 customers. The number of class-1 and class-2 customers are independent, and service time distributions for both classes of customers are the same. We use superscript j to denote quantities corresponding to class-j. For example, $a_i^{(j)}$ is the number of class-j customers in arriving batch when the phase is i.

4.1 System with push-out scheme

We first consider the push-out scheme [6, 14, 18], which is one of selective discarding mechanisms. While there exist some unoccupied places, the system behaves as if there is only one class. A class-2 customer who finds all the places occupied may enter the system, if there is at least one class-1 customer waiting at position $k > K_1$. When it occurs, the last class-1 customer in the system is pushed out (rejected) instead of the arriving class-2 customer, and the arriving class-2 customer joins at the end of the waiting line. So, an arriving class-1 customer who finds K_1 or more customers in the system may join the waiting line, but he might be pushed out. In order for a class-1 customer not to be pushed out while he is at position $k > K_1$, the number of class-2 customers behind him (namely, class-2 customers arrive with him and later) must not exceed $K - k$. Once a class-1 customer proceeds to the position $k \leq K_1$, he is never pushed out and is eventually served.

The queue-length distribution in the push-out system is the same as in the no priority system in Section 3. Because when the system is full, each arriving customer causes a loss of either the arriving customer or a pushed out class-1 customer, and the service distributions for the two classes are the same. It follows that the total loss probability is invariant between the no priority and the push-out systems.

If we let $B^{(1)}$ and $B^{(2)}$ be the loss probabilities for corresponding classes and B the total loss probability, we have

$$\alpha^{(1)}B^{(1)} + \alpha^{(2)}B^{(2)} = (\alpha^{(1)} + \alpha^{(2)})B = \alpha B, \tag{4.1}$$

where $\alpha^{(j)}$ is the mean number of class-j customers in a batch and α is the mean batch size (see Sumita and Ozawa [18]). Since we know the total loss probability B from the result of Section 3, it is then sufficient for us to compute either $B^{(1)}$ or $B^{(2)}$.

Behavior of a class-1 customer

As we stated before, a class-1 customer at position $k > K_1$ might not be served. The probability that a class-1 customer is eventually served depends on his position, arrivals of class-2 customers behind him, and the service times of the customers before him. For the purpose we tag a class-1 customer in the system and pay attention to his behavior.

Now, let $N_{pos}^{(1)}$ be the position of the tagged class-1 customer in the system and $N_{behind}^{(2)}$ the number of class-2 customers behind him. We also denote by R and Y the residual service time and the phase as before. We are interested in the conditional probability

$$f_i(k,x,l) \equiv P\{\text{the tagged class-1 customer is eventually served } \mid N^{(1)}_{pos} = K_1 + k, N^{(2)}_{behind} = x, R = l, Y = i\}. \tag{4.2}$$

From these conditional probabilities we have a vector form

$$\mathbf{f}(k,x,l) \equiv (f_1(k,x,l),\cdots,f_M(k,x,l))^T. \tag{4.3}$$

Assume that $N^{(1)}_{pos} = K_1 + k, N^{(2)}_{behind} = x$ and $R \geq 2$ during a slot. In order for the tagged class-1 customer not to be pushed out at the beginning of next slot, the number of class-2 customers arriving must not exceed $K_2 - k - x$.

Thus, conditioning by the number of class-2 customers in an arriving batch, we have

$$\mathbf{f}(k,x,l) = \sum_{m=0}^{K_2-k-x} \mathbf{A}^{(2)}_m \mathbf{W}\mathbf{f}(k,x+m,l-1) \qquad (1 \leq k \leq K_2, x \leq K_2 = k, l \geq 2), \tag{4.4}$$

and

$$\mathbf{f}(k,x,1) = \sum_{m=0}^{K_2-k+1-x} \mathbf{A}^{(2)}_m \mathbf{W}\sum_{l=1}^{\infty} b_l \mathbf{f}(k-1,x+m,l) \qquad (2 \leq k \leq K_2, x \leq K_2 - k). \tag{4.5}$$

If $N^{(1)}_{pos} = 1$ and $R = 1$ during the n-th slot, the tagged class-1 customer can step to position K_1 at the beginning of the $n+1$-th slot. Then, he is never pushed out and is eventually served. So, we have

$$\mathbf{f}(1,x,1) = (1,\cdots,1)^T \qquad (x \leq K_2 - 1). \tag{4.6}$$

The loss-probabilities

Let us deal with the class-1 loss probability by tagging a class-1 customer in an arriving batch. We consider the probability that the tagged class-1 customer is eventually served. In order to describe environments for the tagged class-1 customer, we introduce some random variables. Let Y_T be the arrival phase at when the batch including the tagged class-1 customer arrives. We denote by $X^{(1)}_T$ the number of class-1 customers including the tagged class-1 customer and by $X^{(2)}_T$ the number of class-2 customers who arrive together with the tagged class-1 customer. The arriving batch size including the tagged class-1 customer is then $X^{(1)}_T + X^{(2)}_T$. We also denote the tagged class-1 customer's position in arriving batch by T_p. Then, it follows that

$$P\{Y_T = i\} = \frac{\omega_i \alpha^{(1)}_i}{\omega_1\alpha^{(1)}_1 + \omega_2\alpha^{(1)}_2 + \cdots + \omega_M\alpha^{(1)}_M}. \tag{4.7}$$

Following the argument in Takahashi et al. [19], we have

$$P\{X^{(1)}_T = k \mid Y_T = i\} = \frac{k a^{(1)}_i(k)}{\alpha^{(1)}_i}, \qquad P\{X^{(2)}_T = k \mid Y_T = i\} a^{(2)}_i(k), \tag{4.8}$$

and

$$P\{T_p = l \mid Y_T = i\} = \sum_{k=1}^{\infty} P\{X^{(1)}_T = k \mid Y_T = i\}\frac{1}{k} = \frac{\overline{a}^{(1)}_i(l)}{\alpha^{(1)}_i}. \tag{4.9}$$

It is also helpful for us to introduce the queue-length distribution at class-1 customers' arrival epochs. Let $Q^{(1)}$ be the number of customers in the system which the batch including the tagged class-1 customer 'sees' at arrival epoch and $q_i^{(1)}(k)$ be the conditional probability $P\{Q^{(1)} = k \mid Y_T = i\}$. Here, we assume that an arriving batch 'sees' the state of the system after departure (if any) and before its arrival.

We further let Q_n be the number of customers in the system at time n, after departure and before arrival, if any. Here, 'at time n' means after departure and before arrivals. Then, it is easily seen that $\{Q_n = 0\}$ occurs if and only if $\{N_n = 0\}$ or $\{N_n = 1, R_n = 1\}$ occurs. For $k = 1, \cdots, K$, we have similar relations between Q_n and (N_n, Y_n), too. Moreover, from the conditional-GASTA property ([21], or see Appendix), we can equate $P\{Q^{(1)} = k \mid Y_T = i\}$ with $P\{Q = k \mid Y = i\}$. Thus, we have

$$
\begin{aligned}
(4.10) \qquad q_i^{(1)}(0) &= P\{Q^{(1)} = 0 \mid Y_T = i\} = P\{Q = 0 \mid Y = i\} \\
&= P\{N = 0 \mid Y = i\} + P\{N = 1, R = 1 \mid Y = i\} \\
&= \{p_i(0) + p_i(1,1)\}/\omega_i,
\end{aligned}
$$

$$
\begin{aligned}
(4.11) \quad q_i^{(1)}(k) &= P\{Q^{(1)} = k \mid Y_T = i\} \qquad (1 \le k \le K-1) \\
&= P\{N = k, R \ge 2 \mid Y = i\} + P\{N = k+1, R = 1 \mid Y = i\} \\
&= \{\sum_{l=2}^{\infty} p_i(k,l) + p_i(k+1,1)\}/\omega_i = \{p_i(k) - p_i(k,1) + p_i(k+1,1)\}/\omega_i,
\end{aligned}
$$

and

$$
\begin{aligned}
(4.12) \qquad q_i^{(1)}(K) &= P\{Q^{(1)} = K \mid Y_T = i\} = P\{N = K, R \ge 2 \mid Y = i\} \\
&= \sum_{l=2}^{\infty} p_i(K,l)/\omega_i \\
&= \{p_i(K) - p_i(K,1)\}/\omega_i,
\end{aligned}
$$

where Q is the generic random variable of Q_n, and $N, R, Y, p_i(k)$ and $p_i(k,l)$ are quantities described in Section 2 and in Section 3.

Using the above results, we have

(4.13)

$$
\begin{aligned}
P_s(i) &\equiv P\{\text{the tagged class-1 customer is eventually served } \mid Y_T = i\} \\
&= \sum_{k=1}^{K_1} P\{Q^{(1)} + T_p = k \mid Y_T = i\} \\
&\quad + \sum_{k=1}^{K_2} P\{Q^{(1)} + T_p = K_1 + k, \text{tagged customer is eventually served } \mid Y_T = i\} \\
&= \sum_{k=1}^{K_1} \sum_{l=0}^{k-1} q_i^{(1)}(l) \frac{\overline{a}_i^{(1)}(k-l)}{\alpha_i^{(1)}} + \frac{1}{\omega_i \alpha_i^{(1)}} \sum_{r=1}^{\infty} \Bigg[\sum_{k=1}^{K_2} \Bigg[p_i(0)\overline{a}_i^{(1)}(K_1+k)b_r \\
&\quad + \sum_{m=1}^{K_1+k} p_i(m,1)\overline{a}_i^{(1)}(K_1+k-m+1)b_r + \sum_{m=1}^{K_1+k-1} p_i(m,r+1)\overline{a}_i^{(1)}(K_1+k-m) \Bigg] \\
&\quad \times \sum_{l=0}^{K_2-k} a_i^{(2)}(k)\{\overline{w}_i f_i(k,l,r) + \sum_{j \ne i} w_{ij} f_j(k,l,r)\} \Bigg].
\end{aligned}
$$

From (4.7) and (4.13), we finally obtain the loss probability for class-1 as

$$B^{(1)} = 1 - [P_s(1)P\{Y_T = 1\} + \cdots + P_s(M)P\{Y_T = M\}]. \tag{4.14}$$

The loss probability $B^{(2)}$ is obtained through (4.1).

4.2 System with buffer reservation scheme

Let us introduce the buffer reservation scheme [6, 14]. The buffer management policy is rather simple than that of push-out scheme. The system behaves as if there is a single class while the number of customers in the system is less than K_1. When K_1 or more customers are in the system, only class-2 customers are allowed to enter the system, and all the arriving class-1 customers are immediately rejected. Once a customer enters the system, he is never rejected regardless of class.

Recall that we started with (3.12) to get the steady-state distribution in Section 3 for the no priority system. With the buffer reservation scheme, we similarly have the following relationship:

$$\mathbf{r}(l) = b_l \mathbf{r}(1)\mathbf{C}^{(1)}\mathbf{C}^{(2)}\mathbf{V} + \mathbf{r}(l+1)\mathbf{A}^{(1)}\mathbf{A}^{(2)}\mathbf{V}, \tag{4.15}$$

where

$$\mathbf{A}^{(1)} \equiv \begin{pmatrix} \mathbf{A}_0^{(1)} & \mathbf{A}_1^{(1)} & \cdots & \mathbf{A}_{K_1-2}^{(1)} & \overline{\mathbf{A}}_{K_1-1}^{(1)} & \\ & \mathbf{A}_0^{(1)} & \cdots & \mathbf{A}_{K_1-3}^{(1)} & \overline{\mathbf{A}}_{K_1-2}^{(1)} & 0 \\ & & \ddots & \vdots & \vdots & \\ & 0 & & \mathbf{A}_0^{(1)} & \overline{\mathbf{A}}_1^{(1)} & \\ & & & & \mathbf{I} & \\ & & 0 & & & \mathbf{I} \end{pmatrix}, \tag{4.16}$$

$$\mathbf{A}^{(2)} \equiv \begin{pmatrix} \mathbf{A}_0^{(2)} & \mathbf{A}_1^{(2)} & \mathbf{A}_2^{(2)} & \cdots & \mathbf{A}_{K-2}^{(2)} & \overline{\mathbf{A}}_{K-1}^{(2)} \\ & \mathbf{A}_0^{(2)} & \mathbf{A}_1^{(2)} & \cdots & \mathbf{A}_{K-3}^{(2)} & \overline{\mathbf{A}}_{K-2}^{(2)} \\ & & \mathbf{A}_0^{(2)} & \cdots & \mathbf{A}_{K-4}^{(2)} & \overline{\mathbf{A}}_{K-3}^{(2)} \\ & & & \ddots & \vdots & \vdots \\ & 0 & & & \mathbf{A}_0^{(2)} & \overline{\mathbf{A}}_1^{(2)} \\ & & & & & \mathbf{I} \end{pmatrix}, \tag{4.17}$$

$$\mathbf{C}^{(1)} \equiv \tilde{\mathbf{I}}_0\mathbf{A}^{(1)}, \qquad \tilde{\mathbf{I}}_0 \equiv \begin{pmatrix} \{\mathbf{I} - \mathbf{A}_0^{(1)}\mathbf{A}_0^{(2)}\mathbf{W}\}^{-1} & & 0 \\ & \mathbf{I} & \\ 0 & & \ddots & \\ & & & \mathbf{I} \end{pmatrix}, \tag{4.18}$$

and

$$\mathbf{C}^{(2)} \equiv \begin{pmatrix} \mathbf{A}_1^{(2)} & \mathbf{A}_2^{(2)} & \mathbf{A}_3^{(2)} & \cdots & \mathbf{A}_{K-1}^{(2)} & \overline{\mathbf{A}}_{K}^{(2)} \\ \mathbf{A}_0^{(2)} & \mathbf{A}_1^{(2)} & \mathbf{A}_2^{(2)} & \cdots & \mathbf{A}_{K-2}^{(2)} & \overline{\mathbf{A}}_{K-1}^{(2)} \\ & \mathbf{A}_0^{(2)} & \mathbf{A}_1^{(2)} & \cdots & \mathbf{A}_{K-3}^{(2)} & \overline{\mathbf{A}}_{K-2}^{(2)} \\ & & \ddots & \ddots & \vdots & \vdots \\ & 0 & & \ddots & \mathbf{A}_1^{(2)} & \overline{\mathbf{A}}_2^{(2)} \\ & & & & \mathbf{A}_0^{(2)} & \overline{\mathbf{A}}_1^{(2)} \end{pmatrix}. \tag{4.19}$$

The vector $\mathbf{r}(l)$ and matrix $\mathbf{V}$ are given by (3.4) and (3.7) respectively. Starting with (4.15) instead of (3.12), we have the queue-length distribution through similar arguments to those in Section 3.

The loss probabilities

Applying the Little's law at the server, we get

$$\left[\alpha^{(1)}\{1-B^{(1)}\}+\alpha^{(2)}\{1-B^{(2)}\}\right]\beta = 1-\mathbf{p}(0)\mathbf{e}. \tag{4.20}$$

It remains to obtain one of the loss probabilities $B^{(1)}$ or $B^{(2)}$.

Let $X^{(1)}$ be the number of arriving class-1 customers and $N_{Loss}^{(1)}$ be the number of class-1 customers rejected at the beginning of a slot. When $X^{(1)}=0$, $N_{Loss}^{(1)}$ equals to 0.

We denote by Q the generic random variable of Q_n introduced in Section 4.1, and by $q_i(k)$ the conditional probability $P\{Q=k \mid Y=i\}$. Then, it follows that

$$\begin{aligned} q_i(0) &= \{p_i(0)+p_i(1,1)\}/\omega_i \\ q_i(k) &= \{p_i(k)-p_i(k,1)+p_i(k+1,1)\}/\omega_i \qquad (1\le k\le K-1) \\ q_i(K) &= \{p_i(K)-p_i(K,1)\}/\omega_i \end{aligned} \tag{4.21}$$

Since the capacity of the system is limited to K_1 for class-1 customers, arriving class-1 customers will be rejected when $Q+X^{(1)}$ exceeds K_1. Thus, we have

$$\begin{aligned} P\{N_{Loss}^{(1)}=0 \mid Y=i\} &= \sum_{k=0}^{K_1-1} P\{Q=k, X^{(1)}\le K_1-k \mid Y=i\} \\ &\quad + \sum_{k=0}^{K_2} P\{Q=K_1+k, X^{(1)}=0 \mid Y=i\} \\ &= \sum_{k=0}^{K_1-1} q_i(k)\sum_{l=0}^{K_2} a_i^{(1)}(l) + \sum_{k=0}^{K_2} q_i(K_1+k)a_i^{(1)}(0), \end{aligned} \tag{4.22}$$

and for $l=1,2,\cdots$

$$\begin{aligned} P\{N_{Loss}^{(1)}=l \mid Y=i\} &= \sum_{k=0}^{K_1-1} P\{Q=k, X^{(1)}=K_1-k+l \mid Y=i\} \\ &\quad + \sum_{k=0}^{K_2} P\{Q=K_1+k, X^{(1)}=l \mid Y=i\} \\ &= \sum_{k=0}^{K_1-1} q_i(k)a_i^{(1)}(K_1-k+l) + \sum_{k=0}^{K_2} q_i(K_1+k)a_i^{(1)}(l). \end{aligned} \tag{4.23}$$

Then, we have

$$\begin{aligned} E(N_{Loss}^{(1)} \mid Y=i) &= \sum_{l=1}^{\infty} lP\{N_{Loss}^{(1)}=l \mid Y=i\} \\ &= \sum_{k=0}^{K_1-1} q_i(k)\sum_{l=1}^{\infty} la_i^{(1)}(K_1-k+l) + \sum_{k=0}^{K_2} q_i(K_1+k)\sum_{l=1}^{\infty} la_i^{(1)}(l). \end{aligned} \tag{4.24}$$

Finally, we obtain class-1 loss probability $B^{(1)}$ as

$$(4.25)\qquad B^{(1)}=\frac{E(N^{(1)}_{Loss})}{E(X^{(1)})}=\frac{\omega_1 E(N(1)_{Loss}\mid Y=1)+\cdots+\omega_M E(N^{(1)}_{Loss}\mid Y=M)}{\omega_1\alpha^{(1)}_1+\cdots+\omega_M\alpha^{(1)}_M}.$$

The loss probability for class-2 $B^{(2)}$ is immediately derived from (4.20) and (4.25).

5. Numerical results

In this section, we assume that $M=2$. Namely, the input precess is SBBP. The arriving batch distribution of class-j customers in phase i is assumed to be Poisson with parameter $\alpha^{(j)}_i$. Also, we assume a deterministic service ($b_1=1,\beta=1$).

The ratio of class-2 customers to all customers is taken as 20% ($\alpha^{(2)}_i/\alpha_i=0.2$). To represent bursty traffic, we assume phase 2 is bursty ($\alpha_2>\alpha_1$) and burstiness (α_2/α_1) is taken as 3. We also assume that the mean sojourn times in individual phase are $10^5(=1/\overline{w}_1)$ and $2\times10^4(=1/\overline{w}_2)$ so that the fraction of bursty state becomes 1/6.

Figure 1 shows the classwise loss probabilities for different positions of threshold ($K_1=K-K_2$) under the fixed total buffer size $K=64$ with push-out or with buffer reservation scheme. In both systems, the differences between the two loss probabilities increase as the value of K_1 decreases. With push-out scheme, class-1 loss probability is almost constant due to the conservation law for the total probability, but the decreace of class-2 loss probability is relatively small. Contrarily, with buffer reservation scheme, the decrement of class-2 loss probability is large, but it is due to sacrificing the class-1 loss probability. Comparing two space priority schemes under the common condition of loss probabilities, push-out is superior to buffer reservation in total buffer size (K), while buffer reservation is superior to push-out in the value of K_2.

Next, we investigated the number of necessary buffer size for an offered load $E(X)/\beta=E(X)$ under the requirements about classwise loss probabilities such that

- The loss probability for class-1 customers ($B^{(1)}$) must be less than 10^{-6}.
- The loss probability for class-2 customers ($B^{(2)}$) must be less than 10^{-10}.

For comparison, we examined three buffering schemes: no priority, push-out and buffer reservation (figure 2). When the no-priority scheme is selected, it is assumed that the total loss probability must be less than 10^{-10}.

As we can see from figure 2, the superiority of using space priority is evident: for an offered load, the required buffer length can be reduced, or inversely, the admissible mean load can be increased for a given buffer length. The superiority of push-out scheme to buffer reservation scheme decreases according as the buffer length increases.

6. Concluding remarks

We have analyzed the M-state MMBP input finite queues without and with space priority. Starting to solve the M-linear equations, we have recursively obtained the total number of customers in the individual system. We have also derived the loss probabilities for the FIFO (no priority) and two-class priority systems.

It is left for future work to study the interdeparture and overflow process from the queues. It is also worthwhile to study a more general queue with multiple priority classes from a practical point of view.

Acknowledgment

The authors are grateful to Professor Yukio Takahashi of Tokyo Institute of Technology for his helpful suggestions and constructive comments.

Appendix–Conditional GASTA

Here we introduce the conditional-GASTA (Geometric Arrivals See Time Averages) property, a discrete-time analogue of conditional PASTA [4].

We are interested in some discrete-time based systems to which arrivals depend on some stochastic processes, defined on a probability space $(\Omega, \mathcal{F}, P)$. In order to denote these systems, we consider the following random variables.

$X_n (n = 1, 2, \cdots)$: state of a system in $[n-1, n)$

$Y_n (n = 1, 2, \cdots)$: arrival phase in $[n-1, n)$

$B_n (n = 1, 2, \cdots)$: number of arriving customer (0 or 1) at time n

$1\{E\}$: indicator function of the event $\{E\}$

For every n, X_n and Y_n take values on appropriate metric spaces S_X and S_Y, respectively.

We denote by s an arbitrary but fixed value in S_Y. We assume s satisfies the following conditions:

(i) The limit

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} 1\{Y_k = s\} \tag{A.1}$$

exists and positive with probability one.

(ii) There exists a Bernoulli process $\{B_n^s\}$ with parameter p_s, s.t. $B_n = B_n^s$ whenever $Y_n = s$; that is,

$$1\{Y_n = s\} B_n = 1\{Y_n = s\} B_n^s. \tag{A.2}$$

Let A be an arbitrary subset of S_X such that $\{X_n \in A\} \in \mathcal{F}$ for every $n (n = 1, 2, \cdots)$. We define two key quantities:

$$V_n^s \equiv \sum_{k=1}^{n} 1\{X_k \in A\} 1\{Y_k = s\} / \sum_{k=1}^{n} 1\{Y_k = s\}, \tag{A.3}$$

the fraction of time $X \equiv \{X_n\}$ spends in A, when restricting ourselves to those intervals in $[0, n)$ during which $Y \equiv \{Y_n\}$ is in state s, and

$$W_n^s \equiv \sum_{k=1}^{n} 1\{X_k \in A\} 1\{Y_k = s\} B_k / \sum_{k=1}^{n} 1\{Y_k = s\} B_k, \tag{A.4}$$

the fraction of arrivals (or the numbers of arriving batches for batch input model) who find X in A, when restricting ourselves to those arrivals who find Y in state s during $[0, n)$.

The result and its proof

Here is a discrete-time analogue of *Adapted Lack of Anticipation Assumption* (ALAA) in [4].

ALAA: *For every* $n (n = 1, 2, \cdots)$, B_n^s *and* $\{(X_k, Y_k) \mid k = 1, 2, \cdots, n\}$ *are independent.*

We are now in a position to state the conditional-GASTA property. The proof is analogous to the proof of Theorem 1 in Wolff [22].

Theorem. *Under ALAA, there exists a random variable V^s such that $V_n^s \to V^s$ w.p.1 as $n \to \infty$ if and only if there exists a random variable W^s such that $W_n^s \to W^s$ w.p.1 as $n \to \infty$, and in this case $V^s = W^s$ w.p.1.*

Proof. We denote the process

$$\mathcal{R}_n^s \equiv 1\{X_n \in A\}1\{Y_n = s\}(B_n^s - p^s), \qquad n = 1, 2, \cdots \tag{A.5}$$

$$\mathcal{U}_n^s \equiv \sum_{k=1}^{n} \mathcal{R}_k^s \equiv \mathcal{W}_n^s - np_s\mathcal{V}_n^s, \qquad n = 1, 2, \cdots \tag{A.6}$$

where

$$\mathcal{V}_n^s \equiv \sum_{k=1}^{n} 1\{X_k \in A\}1\{Y_k = s\}/n, \tag{A.7}$$

and

$$\mathcal{W}_n^s \equiv \sum_{k=1}^{n} 1\{X_k \in A\}1\{Y_k = s\}B_k = \sum_{k=1}^{n} 1\{X_k \in A\}1\{Y_k = s\}B_k^s. \tag{A.8}$$

It is obvious that $E\{(\mathcal{R}_n^s)^2\} \leq 1$ and that $\sum_{n=1}^{\infty} E\{(\mathcal{R}_n^s)^2\}/n^2$ converges. Furthermore, it is seen by our ALAA that $E(\mathcal{R}_n^2 \mid \mathcal{U}_{n-1}) = 0$ and this leads that $\{\mathcal{U}_n^2\}$ forms a discrete-time martingale.

Applying the convergence theorem for martingales (Theorem 3, p.243 of Feller [5]), we have

$$\frac{1}{n}\mathcal{U}_n^s \to 0 \text{ w.p.1 as } n \to \infty. \tag{A.9}$$

From (A.9), we thus have

(A.10)
$$\lim_{n\to\infty} \left\{ \sum_{k=1}^{n} 1\{X_k \in A\}1\{Y_k = s\}B_k^s / \sum_{k=1}^{n} B_k^s \right\} =^{\text{w.p.1}} \lim_{n\to\infty} \left\{ \sum_{k=1}^{n} 1\{X_k \in A\}1\{Y_k = s\}/n \right\}.$$

Substituting $A = \Omega$ into (A.10) gives

$$\lim_{n\to\infty} \left\{ \sum_{k=1}^{n} 1\{Y_k = s\}B_k^s / \sum_{k=1}^{n} B_k^s \right\} =^{\text{w.p.1}} \lim_{n\to\infty} \left\{ \sum_{k=1}^{n} 1\{Y_k = s\}/n \right\}. \tag{A.11}$$

Recalling (A.3) and (A.4), we obtain the theorem statement from the ratio of (A.10) to (A.11). □

Remark. In the case of $Y_n \equiv s$, our theorem reduces of Theorem 1 in Halfin [8].

References

[1] Baba, Y.: The $M^X/G/1$ queue with finite waiting room, *Journal of the Operations Research Society Japan*, Vol. 27 (1984) 260–273.

[2] Blondia, C.: The $N/G/1$ finite capacity queue, *Stochastic Models*, Vol. 5 (1989) 273–294.

[3] Daigle, J.N. , Lee, Y. and Magalhães, M.N.: Discrete Time Queues with Phase Dependent Arrivals, *Proceedings IEEE INFOCOM '90*, (1990) 728–732.

[4] van Doorn, E.A. and Regterschot, G.J.K.: Conditional PASTA, *Operations Research Letters*, Vol. 7 (1988) 229–232.

[5] Feller, W.: *An Introduction to Probability Theory and Applications*, 2nd ed., John Wiley & Sons Inc. (1971).

[6] García, J. and Casals, O.: Space Priority mechanisms with bursty traffic, *Proc. Int'l Conf. on the Performance of Distributed Systems and Integrated Communication Networks* (Hasegawa, T., Takagi, H. and Takahashi, Y. eds), North-Holland (1991) 363–382.

[7] Gravey, A. , Louvion, J.R. and Boyer, P.: On the $Geo/D/1$ and $Geo/D/1/n$ Queues, *Performance Evaluation*, Vol. 11 (1990) 117–125.

[8] Halfin, S.: Batch Delays Versus Customer Delays, *The Bell System Technical Journal*, Vol. 62 (1983) 2011–2015.

[9] Hashida, O. and Takahashi, Y.: A discrete-time priority queue with Switched Batch Bernoulli Process inputs and constant service time, *Proc. 13-th Int'l Teletraffic Congress* (Jensen, A. and Iversen, V.B. eds), North-Holland (1991) 521–526.

[10] Hashida, O., Takahashi, Y. and Shimogawa, S.: Switched Batch Bernoulli Process (SBBP) and the Discrete-Time $SBBP/G/1$ Queue with Application to Statistical Multiplexer Performance, *IEEE J. Selected Areas in Communications*, Vol. 9 (1991) 394-401.

[11] Heffes, H. and Lucantoni, D.M.: A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance, *IEEE J. Selected Areas in Communications*, Vol. 4 (1986) 856–868.

[12] Khamisky, A. and Sidi, M.: Discrete-time priority queueing systems with two-state Markov modulated arrival processes, *Proc. IEEE INFOCOM '91*, (1991) 1456–1463.

[13] Kobayashi, H.: Discrete-time queueing systems, *Probability Theory and Computer Science* (Louchard, G. and Latouche, G. eds), Academic Press (1983) 53–85.

[14] Kröner, H., Hébuterne,G. , Boyer, P. and Gravey, A.: Priority Managements in ATM Switching Nodes, *IEEE J. Selected Areas in Communications*, Vol. 9 (1991) 418–427.

[15] Murata, M., Oie, Y. and Suda, T.: Analysis of a discrete-time single-server queue with bursty inputs for traffic control in ATM networks, *IEEE J. Selected Areas in Communications*, Vol. 8 (1990) 447–458.

[16] Seneta, E.: *Non-negative Matrices and Markov Chains*, Springer-Verlag, 2nd edition (1981)

[17] Stavrakakis, I.: Statistical multiplexing under non-i.i.d. packet arrival process and different priority policies, *Performance Evaluation*, Vol. 12 (1991) 181–189.

[18] Sumita, S. and Ozawa, T.: Achievability of performance objectives in ATM switching nodes, *Proc. Int'l Sem on the Performance of Distributed and Parallel Systems*, (Hasegawa, T., Takagi, H. and Takahashi, Y. eds), North-Holland (1989) 45–56.

[19] Takahashi, Y. and Hashida, O.: Delay analysis of discrete-time priority queue with structured inputs, *Queueing Systems*, Vol. 8 (1991) 149–164.

[20] Tran-Gia, P. and Ahamadi, H.: Analysis of a discrete-time $G^{[X]}/D/1-S$ queueing system with applications in packet switching systems, *Proc. IEEE INFOCOM'88*, (1988) 861–870.

[21] Tsuchiya, T. , Takahashi, Y. and Takahashi, Y.: Conditional GASTA (Geometric Arrivals See Time Averages), *The 1991 Autumn Conference of ORSJ*, 1-D-6 (1991).

[22] Wolff, R.W.: Poisson Arrivals See Time Averages, *Operations Research*, Vol. 20 (1982) 223–231.

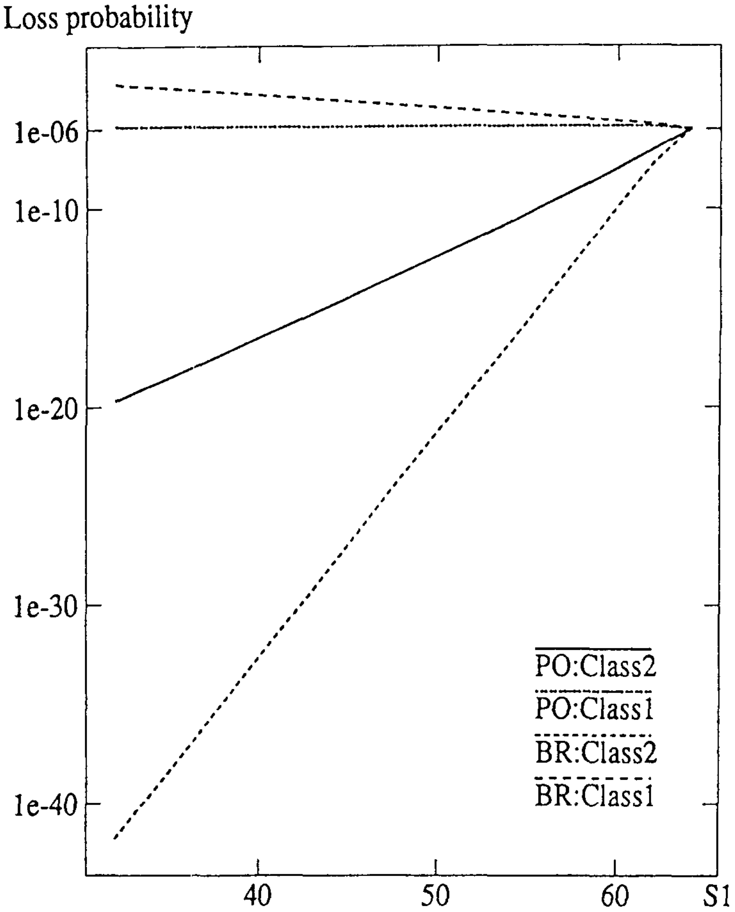

Figure 1: Loss probabilities for different values of $S_1(K = 64)$

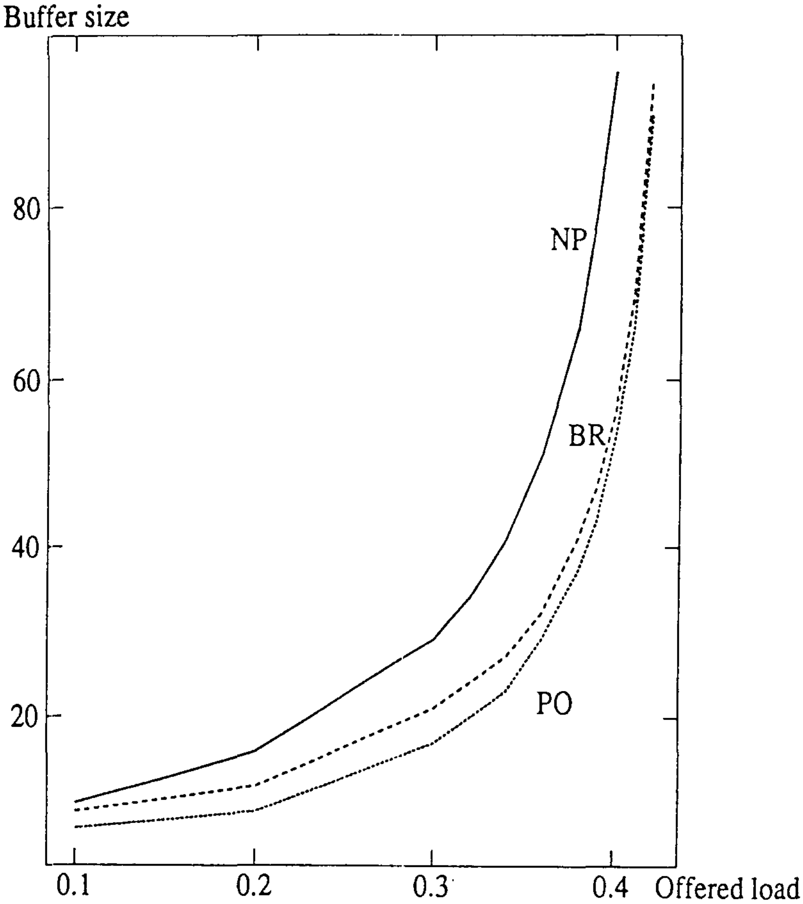

Figure 2: Average load versus buffer size

Toshiaki Tsuchiya & Yoshitaka Takahashi:
NTT Telecommunication Networks Laboratories
9-11 Midori-cho, 3-Chome, Musashino-shi
Tokyo 180, JAPAN