On-line concurrent estimation of priority queueing systems with feedback controlled and non-renewal input streams

Fumio Ishizaki
Department of Information and Telecommunication Engineering
Nanzan University, 27 Seirei, Seto 489-0863, Japan
`ishizaki@it.nanzan-u.ac.jp`

Abstract

We consider a discrete-time single-server priority queueing system where the arrival process is comprised of feedback controlled streams and non-renewal streams. Such a priority queueing system frequently appears in various applications. For the priority queueing systems, we develop an on-line concurrent performance estimation technique, which uses a *proportional relation* between the stationary distributions of the queueing systems with different buffer capacity. Our technique is useful to various controls for dynamic buffer allocation in multimedia networks.

1 Introduction

In many applications, from an observation of a queueing system with a buffer capacity, predicting performance of the queueing systems with different buffer capacity is often required. For instance, in packet/cell networks, various controls for dynamic buffer allocation at router/switch need to concurrently estimate the performance measures in queueing systems with different buffer capacity from an observation of a "real" queueing system.

The sample path constructability techniques including Augmented System Analysis (ASA) [4, 5, 6] and Time Warping Algorithm (TWA) [3] are the most widely known method for concurrent performance estimation. From an observation of a sample path associated with the system with a parameter value θ, the sample path constructability techniques try to construct a corresponding sample path which is associated with the system with different parameter value θ_m and is stochastically equivalent to the original sample path. As a result, the sample path constructability techniques can estimate the performance measures in the system with the parameter value θ_m from the constructed sample path. However, the sample path constructability techniques are not applicable to queueing system where the arrival process is comprised of streams with feedback control based on the queue length and non-renewal streams [9]. The reason for the inapplicability of the sample path constructability techniques is that the existence of the feedback control and the non-renewal streams in the queueing systems makes it difficult to construct the sample paths associated with the queueing systems with different buffer capacity from an observation of a queueing system. Such queueing systems with feedback controlled streams and non-renewal streams frequently appear in various applications. For example, in packet/cell networks, feedback control based on the queue length is usually done and the arrival process of packet/cell is a bursty and time-correlated process. Thus, for concurrent performance estimation, methods which can be applied to such queueing systems are strongly needed. Recently, to overcome the inapplicability, Ishizaki et al. [9] have developed a method for concurrent performance estimation, which uses a *proportional relation* [7, 8, 9, 10, 13] holding between the stationary distributions of queueing systems with different buffer capacity.

In this paper, we consider a feedback controlled queueing system with priority, which is an extension of the queueing system without priority [9], and we develop an on-line concurrent performance estimation method for the priority queueing system by extending the method developed in [9]. The method presented in this paper concurrently estimates various performance measures of the priority queueing systems with different buffer capacity.

Priority mechanism is considered useful in many applications. For example, to meet the different QoS (quality-of-service) requirements of the diverse traffic that will be carried at multimedia networks, some switches are now being designed to allow the connections to be partitioned into priority classes [2]. It is thus expected that our method for priority queueing system will be helpful to various controls for buffer allocation in multimedia networks.

The remainder of the paper is organized as follows. In Section 2, a discrete-time single-server priority queueing system considered in this paper is described. In Section 3, we establish a proportional relation between the stationary distributions of the queueing systems with different buffer capacity. The proportional relation enables us to express various performance measures in the queueing system with buffer capacity K_1 in terms of the stationary distribution of the queueing system with buffer capacity K_0, where $0 < K_1 \leq K_0 \leq \infty$. Such expressions of the loss probability, the mean queue length and the mean sojourn time are shown in Section 4. In Section 5, for on-line concurrent performance estimation, we provide their estimators which can be evaluated from an observation of a single sample path of the queueing system with buffer capacity K_0. Some simulation results are presented in Section 6. Conclusions are drawn in Section 7.

2 Model

This section describes a discrete-time priority queueing system to be studied in this paper. In the queueing system, time is divided into equal intervals referred to as *slots*, and the slot length is equal to a unit time. The queueing system is comprised of L buffers and a single-server. There are L priority classes of customers and each class has its own buffer for exclusive use. The capacity for class l customers is equal to $K^{(l)}$ ($0 < K^{(l)} \le +\infty$). Class 0 customers have the highest priority and the other classes are prioritized in the descending order. Class l customers are served only when there is no customer with the higher class, i.e., all the higher class buffers are empty. The priority rule is preemptive. Note here that under the preemptive priority rule, any performance associated with a class is not affected by the existence of the lower priority classes, and the dynamics of a buffer is independent of those of the buffers for the lower classes.

Each priority class has two arrival traffic streams : a controlled stream and an uncontrolled stream. Customers from each traffic stream arrive at the system in batch. The arrival of a batch occurs at the beginning of a slot immediately after the slot boundary (i.e., early arrival model [14]). An arriving class l customer ($l = 0, \ldots, L-1$) is accommodated in the lth buffer when the buffer space is available and is discarded when the lth buffer is full. More precisely, when a total of m customers belonging to class l from both traffic streams arrive to find k available space in the lth buffer, k out of the m arriving customers are randomly selected to be accommodated in the buffer, and the remaining $m-k$ arriving customers are discarded. The customers accommodated into the buffer are served by the single server. The service time of customer is deterministic and is equal to a unit time. The service of customer starts at the beginning of a slot and ends at the end of the slot (i.e., on the slot boundary).

The lth uncontrolled stream ($l = 0, \ldots, L-1$) is a nonrenewal time-correlated arrival process. The batch size of the class l customers arriving from the lth uncontrolled stream in a slot is governed by a stationary and time-homogeneous Markov chain. This Markov chain is referred to as the lth underlying Markov chain hereafter. The number of states in the underlying Markov chain may be finite or infinite; thus, the uncontrolled stream may model traffic with temporal correlations over a wide range of time scales. On the other hand, the lth controlled stream ($l = 0, \ldots, L-1$) is subject to feedback control based on the queue length of the lth buffer. More specifically, the batch size of the class l customers arriving from the lth controlled stream in a slot is probabilistically determined by the queue length of the lth buffer at the beginning of the previous slot.

We now describe the queueing model in more detail. We begin with the introduction of notations. Let $A_n^{(l)}$ ($n = 1, 2, \ldots; l = 0, \ldots, L-1$) denote a random variable representing the size of the class l batch arriving from the uncontrolled stream in the nth slot. Also, let $B_n^{(l)}$ ($n = 1, 2, \ldots; l = 0, \ldots, L-1$) denote a random variable representing the size of the class l batch arriving from the controlled stream in the nth slot. Let $X_n^{(l)}$ ($n = 0, 1, \ldots; l = 0, \ldots, L-1$) denote a random variable on $\{0, \ldots, K^{(l)}\}$ representing the queue length (including a customer in service if any) of the lth buffer in the nth slot. The dynamics of the stochastic sequence $\{X_n^{(l)}\}$ is then represented by the following recursion:

$$X_{n+1}^{(l)} = \min\left((X_n^{(l)} - \delta_n^{(l)})^+ + A_{n+1}^{(l)} + B_{n+1}^{(l)}, K^{(l)}\right),$$

where $\delta_n^{(l)}$ is defined as

$$\delta_n^{(l)} = \mathbf{1}_{\{\sum_{k=0}^{l-1} X_n^{(k)} = 0\}},$$

$(\cdot)^+ = \max\{\cdot, 0\}$, $\mathbf{1}_E$ denotes the indicator function for event E and for notational convenience, we define $\sum_{k=0}^{-1} X_n^{(k)} = 0$. Let $Z_n^{(l)}$ ($n = 1, 2, \ldots; l = 0, \ldots, L-1$) denote a random variable representing the number of the lth class lost customers in the nth slot. $Z_n^{(l)}$ is given by

$$Z_n^{(l)} = \left((X_{n-1}^{(l)} - \delta_{n-1}^{(l)})^+ + A_n^{(l)} + B_n^{(l)} - K^{(l)}\right)^+.$$

For the uncontrolled arrival stream $\{A_n^{(l)}\}$, we introduce some notations. The lth underlying Markov chain is denoted by $\{S_n^{(l)}\}_{n=0}^{\infty}$ and it has $M^{(l)} + 1$ states labeled from 0 to $M^{(l)}$, where $M^{(l)}$ is possibly infinite. We assume that all the underlying Markov chains are independent with each other. The size of the class l batch arriving in a slot is probabilistically determined based on the state of the lth underlying Markov chain in the slot. In other words, given that the value of $S_n^{(l)}$, $A_n^{(l)}$ is conditionally independent of all other random variables. We assume that the state of the underlying Markov chain cannot be observed. The one-step state transition matrix of the lth underlying Markov chain is denoted by $\boldsymbol{U}^{(l)} = \{U_{i,j}^{(l)}\}$ ($i, j = 0, \ldots, M^{(l)}; l = 0, 1$), where $U_{i,j}^{(l)} = \Pr\{S_{n+1}^{(l)} = j | S_n^{(l)} = i\}$. We assume that $\boldsymbol{U}^{(l)}$ is irreducible and positive recurrent. Let $a_{j,k}^{(l)}$ denote the probability that the class l batch size of the customers arriving from the uncontrolled steam in a slot is equal to k, given that the lth underlying Markov chain is in state j. Namely, $a_{j,k}^{(l)}$ is defined as

$$a_{j,k}^{(l)} = \Pr\{A_n^{(l)} = k \mid S_n^{(l)} = j\},$$

for $j = 0, \dots, M^{(l)}$, $k = 0, 1, 2, \dots$, and $l = 0, \dots, L-1$. Also, let $\boldsymbol{A}_k^{(l)}$ ($k = 0, 1, \dots; l = 0, \dots, L-1$) denote $(M^{(l)}+1) \times (M^{(l)}+1)$ matrices whose (i,j)th element $[\boldsymbol{A}_k^{(l)}]_{i,j}$ ($i, j = 0, \dots, M^{(l)}; l = 0, \dots, L-1$) is given by

$$[\boldsymbol{A}_k^{(l)}]_{i,j} = \Pr\{A_{n+1}^{(l)} = k, S_{n+1}^{(l)} = j \mid S_n^{(l)} = i\} = a_{j,k}^{(l)} U_{i,j}^{(l)}.$$

Note that $\boldsymbol{A}_k^{(l)}$ represents the one-step state transition matrix of the class l underlying Markov chain when the class l batch size of k arrive at the queueing system from the uncontrolled stream.

Next, for the controlled arrival stream $\{B_n^{(l)}\}$, we introduce some notations. The batch size arriving from the controlled stream in a slot is probabilistically determined by the queue length in the previous slot. In other words, given that the value of $X_n^{(l)}$, $B_{n+1}^{(l)}$ is conditionally independent of all other random variables. Let $b_{j,k}^{(l)}$'s ($j = 0, \dots, M^{(l)}; k = 0, 1, \dots; l = 0, \dots, L-1$) denote the conditional probabilities which determine the class l batch size in a slot based on the queue length of the lth buffer in the previous slot. $b_{j,k}^{(l)}$'s are given by

$$b_{j,k}^{(l)} = \Pr\{B_{n+1}^{(l)} = k \mid X_n^{(l)} = j\}.$$

We close this section with a key assumption to ensure the proportional relation exists. The assumption imposes a limitation on the matrices $\boldsymbol{A}_0^{(l)}$ ($l = 0, \dots, L-1$).

Assumption 1 *For* $l = 0, \dots, L-1$, *there exists a* $1 \times (M^{(l)}+1)$ *probability row vector* $\boldsymbol{a}^{(l)}$ *satisfying*

$$\boldsymbol{A}_0^{(l)} = \boldsymbol{A}_0^{(l)} \boldsymbol{e}^{(l)} \boldsymbol{a}^{(l)},$$

where $\boldsymbol{e}^{(l)}$ *is an* $(M^{(l)}+1) \times 1$ *column vector whose elements are all equal to one.*

Note that Assumption 1 is equivalent to the following statement: For $l = 0, \dots, L-1$, any $i_0^{(l)}, i_1^{(l)} \in \{0, \dots, M^{(l)}\}$ and $j^{(l)}$,

$$\Pr\{S_{n+1}^{(l)} = j^{(l)} \mid A_{n+1}^{(l)} = 0, S_n^{(l)} = i_0^{(l)}\} = \Pr\{S_{n+1}^{(l)} = j^{(l)} \mid A_{n+1}^{(l)} = 0, S_n^{(l)} = i_1^{(l)}\}.$$

Assumption 1 implies that if no customer from the lth ($l = 0, \dots, L-1$) uncontrolled stream arrives in a slot, the lth uncontrolled stream regenerates at the slot. As a result, the queueing process $\{X_n^{(l)}\}$ of the lth buffer regenerates at slots where the queue length of the lth buffer changes from k to $k-1$ for each fixed k ($0 < k \leq K^{(l)}$).

A simple example of the uncontrolled stream satisfying Assumption 1 is the superposition of heterogeneous on-off sources and heterogeneous Bernoulli sources with the following feature. On-periods of each on-off source follow an arbitrary distribution, while off-periods follow a geometric distribution. During on-period, each on-off source generates exactly one customer in each slot, and it does not generate customers during off-period. Each Bernoulli source generates a customer with a fixed probability in each slot.

3 Proportional relation

We hereafter consider two queueing systems both of which are identical to one described in the previous section but their lth buffer capacity are K_0 and K_1, respectively, where $0 < K_1 \leq K_0 \leq +\infty$. In what follows, we refer to the system with capacity K_i as K_i-system for $i = 0, 1$ and, to emphasize the system capacity, we put the superscript (K_i) on the quantities associated with the K_i-system. In the next section, for concurrent performance estimation of the queueing systems described in the previous section, we will develop a method to concurrently estimate the performance measures of K_1-system from an observation of K_0-system. The method uses a proportional relation holding between K_0-system and K_1-system , i.e., the stationary queue length distribution of K_1-system is expressed as that of K_0-system multiplied by a constant. In this section, we show that the proportional relation holds and the constant is expressed in terms of the distribution associated with only K_0-system.

Due to the limitation of space, in this paper, we omit the analysis and proof (for the analysis and proof, see [8]), and we present the result for the queueing systems with two priority classes (i.e., $L = 2$) only. However, the result can be easily extended to that of queuing systems with any L classes. Note that to estimate the performance measures for the 0th class (i.e., the highest class), the result for the feedback controlled queueing system without priority [9] can be directly applied, because the preemptive priority mechanism considered in this paper does not affect any performance measures for the highest class. Therefore, we hereafter focus on the performance measures only for the 1st class in the queueing system with two priority classes.

Note that the setting which we have made so far makes $\{(X_n^{(1,K_i)}, S_n^{(1)}, X_n^{(0)}, S_n^{(0)})\}$ ($i = 0, 1$) a Markov chain. We hereafter denote this Markov chain by $\mathcal{M}^{(K_i)}$ ($i = 0, 1$). Now we introduce several assumptions for the Markov chain $\mathcal{M}^{(K_i)}$ ($i = 0, 1$). As will be shown, the next two assumptions guarantee that the performance measures of K_1-system can be estimated by utilizing the proportional relation and observing K_0-system.

Assumption 2 *The Markov chain* $\mathcal{M}_i$ ($i = 0, 1$) *is stationary and ergodic.*

Assumption 3 *There exists an integer* j ($0 \leq j \leq K_1 - 1$) *such that* $\Pr\{X_n^{(1,K_i)} = j\} > 0$ *for* $i = 0, 1$.

Assumption 2 guarantees the existence and uniqueness of the stationary distributions of the Markov chain $\mathcal{M}^{(K_i)}$. In our model, we can consider the case where the state space is divided into some disjoint subsets and the stationary and ergodic queueing process (satisfying Assumption 2) is a.s. in one of such subsets. Assumption 3 ensures that the queueing processes in K_0- and K_1-systems belong to a common subset.

The following theorem establishes a proportional relation between the stationary distribution of the Markov chain $\mathcal{M}^{(K_0)}$ associated with K_0-system and that of the Markov $\mathcal{M}^{(K_1)}$ associated with K_1-system. The proof is provided in [8].

Theorem 1 *Under Assumptions 1, 2, and 3, there exists a positive constant c such that*

$$\Pr\{X_n^{(0)} = i_0, S_n^{(0)} = j_0, X_n^{(1,K_1)} = i_1, S_n^{(1)} = j_1\} = c\Pr\{X_n^{(0)} = i_0, S_n^{(0)} = j_0, X_n^{(1,K_0)} = i_1, S_n^{(1)} = j_1\} \quad (1)$$

for any i_0, j_0, j_1 and $i_1 = 0, 1, \ldots, K_1 - 1$.

We make the following assumption to obtain a simple expression of the constant c.

Assumption 4 $\Pr\{X_n^{(0)} = 0, A_{n+1}^{(1)} = 0\} > 0$.

Assumption 4 is made only to exclude the trivial case where there exists an integer j $(0 \le j \le K_1 - 1)$ such that $X_n^{(K_i)} = j$ for all $n = 0, 1, \ldots$ and $i = 0, 1$. In fact, we can easily remove Assumption 4 such as, if $\Pr\{X_n^{(0)} = 0, A_{n+1}^{(1)} = 0\} = 0$, then $\Pr\{X_n^{(1,K_0)} \le K_1 \mid X_n^{(0)} = 0, A_{n+1}^{(1)} = 0\}$ is replaced with 1 (i.e., $c = 1$) in Theorem 2 below.

The following theorem provides an expression of the constant c, which includes the stationary distribution of the Markov chain $\mathcal{M}^{(K_0)}$ associated with K_0-system but does not include that of the Markov chain $\mathcal{M}^{(K_1)}$ associated with K_1-system. The proof is provided in [8].

Theorem 2 *Under Assumption 4, the constant c is given by*

$$c = \frac{1}{\Pr\{X_n^{(1,K_0)} \le K_1 \mid X_n^{(0)} = 0, A_{n+1}^{(1)} = 0\}}. \quad (2)$$

4 Expression of performance measures

In this section, we derive expressions of some steady-state performance measures in K_1-system as the functions of the steady-state distributions in K_0-system. Those expressions allow us to concurrently estimate the performance measures in the queueing systems with different buffer capacity. In this paper, we present only the expression of the performance measures without their derivations (for the derivations, see [8]).

Before presenting the expressions of the performance measures, we consider the mean number of customers arriving at the 1st buffer in a slot. Let $\lambda^{(K_i)}$ $(i = 0, 1)$ denote the mean number of customers arriving at the 1st buffer in K_i-system in a slot. $\lambda^{(K_i)}$ is then defined as $\lambda^{(K_i)} = \mathrm{E}[A_n^{(1)} + B_n^{(1,K_i)}]$. Note that since the batch size $B_n^{(1,K_i)}$ of customers arriving from the controlled stream in a slot depends on the queue length, $\lambda^{(K_1)}$ is not equal to $\lambda^{(K_0)}$ in general. From the definition, it immediately follows that

$$\lambda^{(K_i)} = \mathrm{E}[A_n^{(1)}] + \sum_{k=1}^{\infty}\sum_{m=0}^{K_i} k b_{m,k}^{(1)} \Pr\{X_n^{(1,K_i)} = k\}.$$

We now derive an expression of $\lambda^{(K_1)}$ as the function of the stationary distribution of K_0-system. Applying Theorem 1, after a straightforward calculation, we obtain

$$\begin{aligned}\lambda^{(K_1)} = {} & c\sum_{i=0}^{K_1-1}\sum_{l=1}^{\infty} l b_{i,l}^{(1)} \Pr\{X_n^{(1,K_0)} = i\} \\ & +(1 - c\Pr\{X_n^{(1,K_0)} \le K_1 - 1\})\sum_{l=1}^{\infty} l b_{K_1,l}^{(1)} \\ & +\mathrm{E}[A_n^{(1)}], \end{aligned} \quad (3)$$

where c is given by (2).

First, we present the expression of the mean queue length $L^{(K_1)} = \mathrm{E}[X_n^{(1,K_1)}]$ of the 1st buffer in K_1-system. Applying Theorem 1, after a straightforward calculation, we obtain

$$L^{(K_1)} = K_1 - c\sum_{i=0}^{K_1-1}(K_1 - i)\Pr\{X_n^{(1,K_0)} = i\}. \quad (4)$$

Next, we show the expression of the mean sojourn time $W^{(K_1)}$ of customer entering the 1st buffer in K_1-system. Using Little's formula and applying Theorem 1, we obtain

$$W^{(K_1)} = \frac{K_1 - c\sum_{i=0}^{K_1-1}(K_1 - i)\Pr\{X_n^{(1,K_0)} = i\}}{\Pr\{X_n^{(0)} = 0\} - c\Pr\{X_n^{(0)} = 0, X_n^{(K_0)} = 0\}}. \quad (5)$$

Finally, we consider the loss probability $P_{\text{loss}}^{(K_1)}$ of customers arriving at the 1st buffer in K_1-system. $P_{\text{loss}}^{(K_1)}$ is defined by $P_{\text{loss}}^{(K_1)} = \mathrm{E}[Z_n^{(1,K_1)}]/\mathrm{E}[A_n^{(1,K_1)}]$. Using Rate Conservation Law [11, 12, 1] and applying Theorem 1, we obtain

$$\begin{aligned}P_{\text{loss}}^{(K_1)} &= 1 - \frac{1}{\lambda^{(K_1)}}\Big[\Pr\{X_n^{(0)} = 0\} \\ &\qquad -c\Pr\{X_n^{(0)} = 0, X_n^{(1,K_0)} = 0\}\Big]. \end{aligned} \quad (6)$$

5 Estimator

This section presents on-line concurrent estimators for the performance measures in K_1-system. The estimators derived in this section can predict the performance measures in K_1-system from an observation of K_0-system. This allows us to concurrently estimate the performance measures in the queueing system with different buffer capacities.

The estimators shown below determine performance measures of K_1-system by observing a single sample path of K_0-system for N slots, where N is a positive integer. Let $\hat{P}^{(K_1)}_{\rm loss}(N)$, $\hat{L}^{(K_1)}(N)$ and $\hat{W}^{(K_1)}(N)$ denote the estimators for the loss probability, mean queue length, and mean delay of the perturbed system, respectively. From the expressions in (2), (3), (4), (5) and (6), we immediately obtain the following estimators.

$$\hat{P}^{(K_1)}_{\rm loss}(N) = 1 - \frac{1}{\hat{\lambda}^{(K_1)}(N)}\left[\hat{P}_{0,*}(N) - \hat{c}(N)\hat{P}_{0,0}(N)\right], \tag{7}$$

$$\hat{L}^{(K_1)}(N) = K_1 - \hat{c}(N)\hat{f}_5(N),$$

$$\hat{W}^{(K_1)}(N) = \frac{\hat{L}^{(K_1)}(N)}{\hat{P}_{0,*}(N) - \hat{c}(N)\hat{P}_{0,0}(N)},$$

where

$$\hat{c}(N) = \frac{\sum_{n=0}^{N-1} \mathbf{1}_{\{X^{(0)}_n = 0\}}\mathbf{1}_{\{A^{(1)}_{n+1}=0\}}}{\sum_{n=0}^{N-1} \mathbf{1}_{\{X^{(0)}_n = 0\}}\mathbf{1}_{\{A^{(1)}_{n+1}=0\}}\mathbf{1}_{\{X^{(1,K_0)}_n \le K_1\}}},$$

$$\begin{aligned}\hat{\lambda}^{(K_1)}(N) &= \hat{c}(N)\hat{f}_1(N) + (1-\hat{c}(N)\hat{f}_2(N))\hat{f}_3(N) \\ &\quad + \hat{f}_4(N),\end{aligned}$$

$$\hat{f}_1(N) = \frac{1}{N}\sum_{n=0}^{N-1} B^{(1,K_0)}_{n+1}\mathbf{1}_{\{X^{(1,K_0)}_n \le K_1 - 1\}},$$

$$\hat{f}_2(N) = \frac{1}{N}\sum_{n=0}^{N-1} \mathbf{1}_{\{X^{(1,K_0)}_n \le K_1 - 1\}},$$

$$\hat{f}_3(N) = \frac{\sum_{n=0}^{N-1} B^{(1,K_0)}_{n+1}\mathbf{1}_{\{X^{(1,K_0)}_n = K_1\}}}{\sum_{n=0}^{N-1}\mathbf{1}_{\{X^{(1,K_0)}_n = K_1\}}},$$

$$\hat{f}_4(N) = \frac{1}{N}\sum_{n=0}^{N-1} A^{(1)}_{n+1},$$

$$\hat{f}_5(N) = \frac{1}{N}\sum_{n=0}^{N-1}(K_1 - X^{(1,K_0)}_n)\mathbf{1}_{\{X^{(1,K_0)}_n \le K_1 - 1\}},$$

$$\hat{P}_{0,*}(N) = \frac{1}{N}\sum_{n=0}^{N-1}\mathbf{1}_{\{X^{(0)}_n = 0\}},$$

$$\hat{P}_{0,0}(N) = \frac{1}{N}\sum_{n=0}^{N-1}\mathbf{1}_{\{X^{(0)}_n = 0\}}\mathbf{1}_{\{X^{(1,K_0)}_n = 0\}}.$$

Note that when $N \to \infty$, $\hat{c}(N)$ and $\hat{f}_3(N)$ converge to c and $\sum_{l=1}^{\infty} l b_{K_1,l}$, respectively, with probability one under Assumption 3. We thus see that $\hat{P}^{(K_1)}_{\rm loss}(N)$, $\hat{L}^{(K_1)}(N)$ and $\hat{W}^{(K_1)}(N)$ are consistent estimators for $P^{(K_1)}_{\rm loss}$, $L^{(K_1)}$ and $W^{(K_1)}$, respectively.

6 Simulation results

In this section, we present some simulation results and compare estimates obtained through our method with those obtained through the standard simulation procedure. In the simulation results, we assume that the class l controlled stream $(l = 0, 1)$ composes of $N^{(l)}$ homogeneous controlled sources. Each controlled source generates one customer with probability $\gamma^{(l)}(\eta^{(l)})^n$ when the queue length of the lth buffer in the previous slot is equal to n. $b^{(l)}_{j,k}$ $(l = 0, 1)$ is then given by

$$b^{(l)}_{j,k} = \begin{pmatrix} N^{(l)} \\ k \end{pmatrix}\left(\gamma^{(l)}(\eta^{(l)})^j\right)^k\left(1-\gamma^{(l)}(\eta^{(l)})^j\right)^{N^{(l)}-k},$$

for $0 \le k \le N^{(l)}$ and $b^{(l)}_{j,k} = 0$ for $k > N^{(l)}$. We assume that the class l uncontrolled stream composes of $J^{(l)}$ homogeneous on-off sources with geometrically distributed on-off periods. On-periods (resp. off-periods) of the class l uncontrolled source are i.i.d. and geometrically distributed with mean $(1-\alpha^{(l)})^{-1}$ (resp. $(1-\beta^{(l)})^{-1}$). Each uncontrolled source generates exactly one customer in each slot during on-periods, and it does not generate customers during off-periods.

In the simulation results, we set the parameters as follows. For the 0th class streams, we set $N^{(0)} = 1$, $\gamma^{(0)} = 0.200$, $\eta^{(0)} = 0.994$. $J^{(0)} = 1$, $\alpha^{(0)} = 0.680$ and $\beta^{(0)} = 0.990$. For the 1st class streams, we set $N^{(1)} = 3$, $\gamma^{(1)} = 0.200$, and $\eta^{(1)} = 0.994$, $J^{(1)} = 5$, $\alpha^{(1)} = 0.680$ and $\beta^{(1)} = 0.990$. The buffer capacity $K^{(0)}$ of the 0th buffer is set to 50 in both K_0- and K_1-systems, the buffer capacity K_0 of the 1st buffer in K_0-system is set to 45, and that in K_1-system is set to 35.

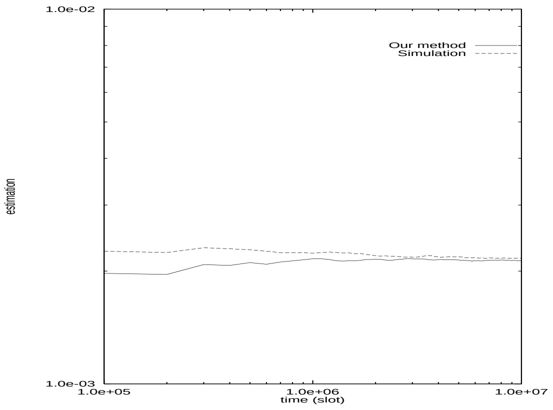

Figure 1: $P^{(K_1)}_{\rm loss} - P^{(K_0)}_{\rm loss}$

We compare the estimation of $P^{(K_1)}_{\rm loss} - P^{(K_0)}_{\rm loss}$ obtained through the method developed in this paper (denoted by "Our method" in figures) with that obtained through the standard simulation procedure (denoted by "Simulation" in figures). In the method developed in this paper, K_0-system is simulated, and a simulated sample path of K_0-system is observed. In the simulated sample path of K_0-system, four stochastic processes and two random variables, namely the sequence

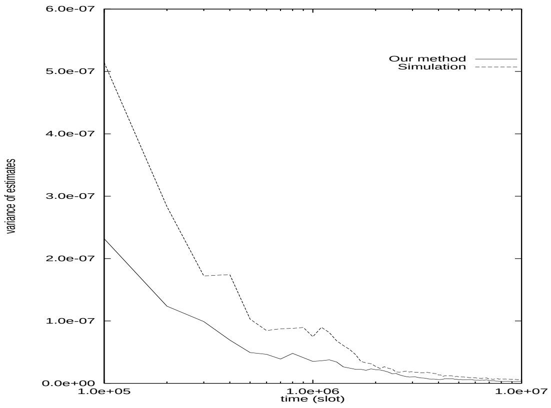

Figure 2: Variance of estimates

of batch sizes of the customers arriving from the lth uncontrolled stream $\{A_n^{(l,K_0)}\}$ $(l = 0, 1)$, the sequence of the batch sizes of the customers arriving from the lth controlled stream $\{B_n^{(l)}\}$ $(l = 0, 1)$ and the initial queue lengths $X_0^{(0)}$ and $X_0^{(1,K_0)}$ are observed. $P_{\text{loss}}^{(K_1)}$ is estimated by (7) from the observation of K_0-system. On the other hand, in the standard simulation procedure, not only K_0-system but also K_1-system is simulated to estimate $P_{\text{loss}}^{(K_1)} - P_{\text{loss}}^{(K_0)}$.

Fig. 1 shows the estimations of $P_{\text{loss}}^{(K_1)} - P_{\text{loss}}^{(K_0)}$ as a function of the observation periods (N). Each estimation is obtained by collecting 30 simulation runs and estimates and taking the average of the 30 estimates. Fig. 2 displays the variance of the 30 estimates. Fig. 2 shows that our method yields good estimates with low variance compared to the standard simulation procedure.

7 Conclusions

This paper extends the proportional relation method developed in [9] for the priority queueing system. Further, without any modification, the result obtained in this paper can be directly applied to the priority queueing system where the service time of customer is geometric [13].

Our method does not assume the knowledge about the arrival processes (e.g, $U_{i,j}^{(l)}$, $a_{j,k}^{(l)}$ and $b_{j,k}^{(l)}$) and the computational complexity of the estimators is irrespective of the numbers $M^{(l)}$ of the states in the underlying Markov chains. This on-line feature is important for practical applications, because our method can trace the changes of the arrival processes. Due to the online feature and applicability to priority queueing systems with feedback controlled and non-renewal input streams, our method is useful to various controls for dynamic buffer allocation, especially, in multimedia networks.

References

[1] F. Baccelli and P. Brémaud, *Elements of queueing theory*, Springer-Verlag, New York, 1994.

[2] A. W. Berger and W. Whitt, Extending the effective bandwidth concept to networks with priority classes, *IEEE Commun. Mag.*, pp.78–83, Aug., 1998.

[3] C. G. Cassandras and C. G. Panayiotou, Concurrent sample path analysis of discrete event systems, *Journal of Discrete Event Dynamic Systems*, Vol.9, pp.171–195, 1999.

[4] C. G. Cassandras and S. G. Strickland, On-line sensitivity analysis of Markov chains, *IEEE Trans. Autom. Contr.*, Vol.34, pp.76–86, 1989.

[5] C. G. Cassandras and S. G. Strickland, Observable augmented systems for sensitivity analysis of Markov and semi-Markov processes, *IEEE Trans. Autom. Contr.*, Vol.34, pp.1026–1037, 1989.

[6] C. G. Cassandras and S. G. Strickland, Sample path properties of timed discrete event systems, *Proc. of the IEEE*, Vol.77 pp.59–71, 1989.

[7] F. Ishizaki, Exact relation between loss probability and queue length in an ATM multiplexer with correlated arrivals and periodic vacations, *Proc. of the 14th International Conference on Information Networking (ICOIN-14)*, pp.1A-3.1–1A-3.8, 2000.

[8] F. Ishizaki, On-line concurrent estimation of priority queueing systems using the proportional relation, preparation, 2000.

[9] F. Ishizaki, G. C. Lin and T. Suda, On-line sensitivity analysis of feedback controlled queueing systems with respect to buffer capacity, *Proc. of IEEE CDC '99*, pp.40–50, 1999.

[10] F. Ishizaki and T. Takine, Loss probability in a finite discrete-time queue in terms of the steady state distribution of an infinite queue, *Queueing Systems*, Vol.31, pp.317-326, 1999.

[11] M. Miyazawa, Comparison of the loss probability of the $GI^X/GI/1/k$ queue with a common traffic intensity, *Journal of Oper. Res. Soc. of Japan*, Vol.32, pp.505–516, 1989.

[12] M. Miyazawa and Y. Takahashi, Rate conservation principle for discrete-time queues, *Queueing Systems*, Vol.12, pp.215–230, 1992.

[13] N. Miyoshi and F. Ishizaki, A unified approach to the proportional relation for discrete-time single-server queues, submitted, 2000.

[14] H. Takagi, *Queueing Analysis, Vol.3*, North-Holland, Amsterdam, 1993.