PAPER *Special Issue on Internet Technology*

Internet Traffic Modeling: Markovian Approach to Self-Similar Traffic and Prediction of Loss Probability for Finite Queues

Shoji KASAHARA[†], ***Regular Member***

SUMMARY It has been reported that IP packet traffic exhibits the self-similar nature and causes the degradation of network performance. Therefore it is crucial for the appropriate buffer design of routers and switches to predict the queueing behavior with self-similar input. It is well known that the fitting methods based on the second-order statistics of counts for the arrival process are not sufficient for predicting the performance of the queueing system with self-similar input. However recent studies have revealed that the loss probability of finite queuing system can be well approximated by the Markovian input models. This paper studies the time-scale impact on the loss probability of $MMPP/D/1/K$ system where the $MMPP$ is generated so as to match the variance of the self-similar process over specified time-scales. We investigate the loss probability in terms of system size, Hurst parameters and time-scales. We also compare the loss probability of resulting $MMPP/D/1/K$ with simulation. Numerical results show that the loss probability of $MMPP/D/1/K$ are not significantly affected by time-scale and that the loss probability is well approximated with resulting $MMPP/D/1/K$.

key words: *internet traffic modeling, self-similarity, fitting, $MMPP/D/1/K$, loss probability*

1. Introduction

Recently a number of high-quality, high-resolution measurements of Internet traffic have been carried out and analyzed. They have shown that traffic from those networks appears to be self-similar with long-range dependence (LRD) [3], [6]. Self-similar traffic is characterized by that the correlation never vanishes in a large time-scale. Its traffic looks the same regardless of time-scales over a long range interval. This fractal behavior makes traffic very bursty. These properties of self-similar traffic are quite different from those of traditional traffic models such as Poisson process and Markovian arrival process (*MAP*) [9]. This observation has initiated studies of new models such as chaotic maps, fractional Brownian motion (FBM) and fractional autoregressive integrated moving average (FARIMA) [3]. They can describe the self-similar behavior in a relatively simple manner. However, queueing theoretical techniques developed in the past are hardly applicable to these models.

Manuscript received January 19, 2001.

Manuscript revised March 12, 2001.

[†]The author is with the Graduate School of Information Science, Nara Institute of Science and Technology, Ikoma-shi, 630-0101 Japan.

On the other hand, a number of models based on traditional traffic models have been proposed. One approach is to emulate self-similarity over a certain range of time-scales with finite state Markovian models. In [10], we have proposed a fitting method for self-similar traffic in terms of *MMPP*. Our fitting method is based on [1] where traffic is modeled by the superposition of several two state *MMPP*s. In [1], the authors proposed the fitting method which is mainly focused on the covariance structure of the second-order self-similar process. More precisely, the parameters of *MMPP* are determined so as to match the autocorrelation function which is approximately evaluated. In our method, however, the parameters are determined so as to match the variance of measured traffic which is exactly evaluated. In the following, we call the resulting *MMPP* a pseudo self-similar process.

It is well known that the queueing performance deteriorates with self-similar traffic whose Hurst parameter is between 0.5 and 1, i.e., LRD traffic [3]. Hence it is important for the appropriate buffer design of routers and switches to predict the queueing behavior under self-similar traffic with LRD*. In general, the fitting methods based on the second-order statistics of counts for the arrival process are not sufficient for predicting the queueing performance [1]. In our previous study [10], we studied the queueing system with infinite buffer and pseudo self-similar input. Numerical results showed that our fitting method works well in the sense of imitating statistical characteristics of self-similar traffic and that it does not work well for predicting the queueing behavior with resulting $MMPP/D/1$ even when the Hurst parameter is moderate.

Recently, [4] discussed the impact of the LRD on the buffer occupancy and indicated that LRD does not affect the buffer occupancy when the busy periods of the system are not large. In [8], the authors considered

*Note that sometimes the actual router architecture is modeled as a queuing system with multiple servers and a finite single-buffer. In general, the queueing system with multiple servers whose service time distribution is general is not analyzable. In this paper, we focus on the single-server queueing system with finite buffer and self-similar input and discuss the possibility of prediction for loss probability with the analytical model.

the critical time scale (CTS) and showed that the buffer behavior at the time-scale beyond the CTS is not significantly affected. Therefore, there is a possibility of accurate prediction of the queueing behavior using the fitting method based on the second-order statistics of counts for the arrival process if the appropriate time-scale is taken into consideration.

In this paper, we consider the time-scale impact on the queue length distribution using an *MMPP* generated so as to match the variance over specified time-scales. We use our fitting method for imitating self-similar traffic with LRD and consider the resulting $MMPP/D/1/K$ system. We investigate the loss probability in terms of system capacity, Hurst parameter and time-scales.

This paper is organized as follows. In Sect. 2, we overview the characteristics of the self-similar process. In Sect. 3, we present the variance fitting method for the exact self-similar process. In Sect. 4, we summarize the analytical results of $MAP/G/1/K$, which is the generalization of $MMPP/D/1/K$. In Sect. 5, we show several numerical examples for the loss probability of $MMPP/D/1/K$ and illustrate the effects of time-scale, Hurst parameter and the system size. Finally, some concluding remarks are given in Sect. 6.

2. Self-Similar Process

In this section, we overview the concept of self-similarity of the stochastic process. First, we summarize Cox's definitions of self-similarity [2] and then, we show the equivalent definitions to those of Cox.

We consider a second-order stationary process $X = \{X_t : t = 1, 2, ...\}$ with the variance σ^2 and the autocorrelation function $r(k)$, where $r(k)$ is given as

$$r(k) = \frac{\mathrm{Cov}(X_t, X_{t+k})}{\mathrm{Var}(X_t)}.$$

In the context of packet traffic, X_t corresponds to the number of packets that arrive during the tth time slot. We also consider a new sequence of $X_t^{(m)}$ which is obtained by averaging the original sequence in non-overlapping blocks. That is,

$$X_t^{(m)} = \frac{1}{m}\sum_{i=1}^{m} X_{(t-1)m+i}, \quad t = 1, 2, \ldots .$$

The new sequence is also a second-order stationary process with the autocorrelation function $r^{(m)}(k)$.

Let δ^2 denote the central second difference operator defined by that for any function $f(x)$,

$$\delta^2(f(x)) = \{f(x+1) - f(x)\} - \{f(x) - f(x-1)\}.$$

Then, definitions of self-similar process are given by the following.

Definition 2.1: X is called *exactly* second-order self-similar with the Hurst parameter $H = 1 - \beta/2$ if

$$r(k) = \frac{1}{2}\delta^2(k^{2-\beta}). \tag{1}$$

Definition 2.2: X is called *asymptotically* second-order self-similar with the Hurst parameter $H = 1-\beta/2$ if

$$r^{(m)}(k) \to \frac{1}{2}\delta^2(k^{2-\beta}), \quad \text{as } m \to \infty. \tag{2}$$

Note that (1) implies that for all $m = 1, 2, \ldots,$

$$r^{(m)}(k) = r(k). \tag{3}$$

We are interested in the range $0.5 < H < 1$ because the process has the long-range dependence. In the case that $H = 0.5$, X is a second-order pure noise with $\mathrm{Var}(X^{(m)}) = \mathrm{Var}(X)/m$.

Let $L(x)$ denote the slowly varying function at infinity, i.e. for any $n > 0$,

$$\lim_{x\to\infty} \frac{L(nx)}{L(x)} = 1.$$

We can define the self-similar process with the variance of the averaged process equivalent to 2.1 and 2.2. The readers are referred to [10], [13] for details.

Definition 2.3: X is called exactly second-order self-similar with the Hurst parameter $H = 1 - \beta/2$ if

$$\mathrm{Var}(X^{(m)}) = \sigma^2 m^{-\beta}. \tag{4}$$

Definition 2.4: X is called *asymptotically* second-order self-similar with the Hurst parameter $H = 1-\beta/2$ if

$$\mathrm{Var}(X^{(m)}) \sim L(m)m^{-\beta}, \quad \text{as } m \to \infty.$$

In our fitting method, we consider the self-similarity under Definition 2.3, that is, we develop the fitting method using (4).

3. Variance Fitting Method

In this section we describe the process of determining the parameters of *MMPP*s so as to mimic the self-similar process [10]. That is, their values are obtained so as to match the variance over several time-scales.

We use a continuous-time *MMPP* for modeling self-similar traffic. We construct an *MMPP* with apparently self-similar behavior over several time-scales by superposing several *MMPP*s. First, consider two-state *MMPP*s with different time-scales. That is, the mean sojourn time of each process is in accordance with the different time-scale. Let us superpose them to make a new *MMPP*. When we see this process in a large time-scale, it looks like an ordinary two-state *MMPP*. If we look in a smaller time-scale, we find that each state is composed of a smaller *MMPP*. This can be repeated only finite times.

The resulting *MMPP* is not self-similar from the definitions in the previous section since it looks constant when the time-scale is larger than the time constant in itself. However, it can emulate self-similarity over several time-scales. Thus, it is practically sufficient to use the process which has self-similarity over only several time-scales to model real traffic.

We assume that the number of states of every underlying *MMPP* is two. So the *MMPP* obtained by the above method is also described by the superposition of several interrupted Poisson processes (*IPP*) and one Poisson process. We assume that the *MMPP* under consideration consists of $d(>1)$ *IPP*s and a Poisson process. We also assume that two modulating parameters of each *IPP* are equal. Then for $1 \leq i \leq d$, we can describe ith *IPP* as follows

$$\boldsymbol{Q}_i = \begin{bmatrix} -\sigma_i & \sigma_i \\ \sigma_i & -\sigma_i \end{bmatrix}, \ \boldsymbol{\Lambda}_i = \begin{bmatrix} \lambda_i & 0 \\ 0 & 0 \end{bmatrix}.$$

Hence the superposition can be described as follows

$$\boldsymbol{Q}=\boldsymbol{Q}_1 \bigoplus \boldsymbol{Q}_2 \bigoplus \cdots \bigoplus \boldsymbol{Q}_d,$$
$$\boldsymbol{\Lambda}=\boldsymbol{\Lambda}_1 \bigoplus \boldsymbol{\Lambda}_2 \bigoplus \cdots \bigoplus \boldsymbol{\Lambda}_d \bigoplus \lambda_p,$$

where $\bigoplus$ means the Kronecker's sum and λ_p is the arrival rate of the Poisson process to be superposed. The whole arrival rate of the process λ is given by

$$\lambda = \lambda_p + \sum_{i=1}^{d} \frac{\lambda_i}{2}. \tag{5}$$

Parameters which we have to determine are σ_i, $\lambda_i(1 \leq i \leq d)$, and λ_p.

First, as preliminary we define the notations used in the procedure and describe some assumptions. Let $N_{t|i}$ be the number of arrivals in the ith *IPP* during the tth time slot and $N_{t|p}$ be the number of arrivals in the Poisson process, and let $N_{t|i}^{(m)}$ and $N_{t|p}^{(m)}$ be respectively the averaged processes of $N_{t|i}$ and $N_{t|p}$. We assume that

$$\mathrm{Var}\left(X_t^{(m)}\right) = \mathrm{Var}\left(\sum_{i=1}^{d} N_{t|i}^{(m)} + N_{t|p}^{(m)}\right).$$

We obtain the variance of the i-th *IPP* as [10]

$$\mathrm{Var}\left(N_{t|i}^{(m)}\right) = \frac{\lambda_i}{2m} + \left\{\frac{1}{4m\sigma_i} - \frac{1}{8m^2\sigma_i^2}(1-e^{-2m\sigma_i})\right\}\lambda_i^2.$$

The corresponding variance of the Poisson process is λ_p/m. Because the variance of a process which is a superposition of independent subprocesses equals the sum of individual variances, the variance of the whole process is given by

$$\mathrm{Var}\left(X_t^{(m)}\right) = \frac{\lambda_p}{m} + \sum_{i=1}^{d} \mathrm{Var}\left(N_{t|i}^{(m)}\right) = \frac{\lambda}{m} + \frac{1}{4}\sum_{i=1}^{d} \eta_i \lambda_i^2, \tag{6}$$

where

$$\eta_i = \frac{1}{m\sigma_i} - \frac{1}{2m^2\sigma_i^2}(1-e^{-2m\sigma_i}).$$

Using (4) and (6), we match the variance at d different points m_i $(1 \leq i \leq d)$. Suppose the range of time-scales over which we want the process to express self-similarity of the original process is $m_{\min} \leq m \leq m_{\max}$, then m_i is defined by

$$m_i = m_{\min} a^{i-1} \quad (1 \leq i \leq d),$$

where

$$a = \left(\frac{m_{\max}}{m_{\min}}\right)^{\frac{1}{d-1}}, \qquad d > 1. \tag{7}$$

From the property of η_i, we have

$$\frac{\lambda}{m} < \mathrm{Var}(X_t^{(m)}) < \frac{\lambda}{m} + \lambda^2. \tag{8}$$

We must choose m_i such that (8) is satisfied at any m_i. This condition comes from that we use a simple *IPP* as a sub-process. Practically, this condition never matters when m is large, but sometimes $\mathrm{Var}(X_t^{(m)})$ is too small when m is small. Therefore, we should be careful to choose m_1, which is large enough to make $\mathrm{Var}(X_t^{(m_1)})$ larger than λ/m_1.

Next, we assume the following relation between σ_i and m_i

$$m_i\sigma_i = \mathrm{const} \quad (1 \leq i \leq d).$$

That is, σ_i can be described as

$$\sigma_i = \frac{m_1}{m_i}\sigma_1 \quad (1 \leq i \leq d). \tag{9}$$

This assumption comes from the intuitive understanding that a self-similar process looks the same in any time-scale. By this assumption, we can reduce the number of the parameters to be determined. That is, if we determine σ_1, we can obtain the values of σ_i $(2 \leq i \leq d)$ by using (9). Furthermore, we can obtain λ_p from the following equation if we determine λ_i.

$$\lambda = \lambda_p + \sum_{i=1}^{d} \frac{\lambda_i}{2}.$$

Now the parameters we need to find are only σ_1 and λ_i.

In the following, we describe the procedure of determining these parameters. We show the parameters preliminarily required for our fitting procedure in Table 1.

Procedure of Parameter Fitting

Table 1 Preliminarily required parameters.

Parameter	Meaning
λ	Arrival rate of the whole process
$m_{\min}$, $m_{\max}$	Minimum and maximum of time-scales over which self-similarity is taken into consideration
σ^2	Variance
H	Hurst parameter
d	Number of *IPP*s

Step 0. Find the range of σ_1 heuristically and fix σ_1.

Step 1. Determine λ_i as the function of σ_i. From (4) and (6), we have

$$\sigma^2 \begin{bmatrix} m_1^{-\beta} \\ m_2^{-\beta} \\ \vdots \\ m_d^{-\beta} \end{bmatrix} = \lambda \begin{bmatrix} m_1^{-1} \\ m_2^{-1} \\ \vdots \\ m_d^{-1} \end{bmatrix} + \boldsymbol{B} \begin{bmatrix} \lambda_1^2 \\ \lambda_2^2 \\ \vdots \\ \lambda_d^2 \end{bmatrix}, \tag{10}$$

where $\boldsymbol{B}$ is the $d \times d$ matrix whose (i,j) element is

$$B_{ij} = \frac{1}{4m_i\sigma_j} - \frac{1}{8m_i^2\sigma_j^2}(1 - e^{-2m_i\sigma_j}). \tag{11}$$

Solving this, we determine λ_i as the function of σ_i.

Step 2. Determine the value of σ_1 from the range found in Step 0. Here we consider the absolute value of difference between the log-scales variance curve of the process given by (6) and that of the self-similar process given by (4). Note that this value is the function of m. We take the integral of the difference over $m_{\min} \le m \le m_{\max}$. Since the integral is the function of σ_1, we determine the value of σ_1 so as to minimize that integral by appropriate search technique.

Step 3. Determine the values of λ_i from (10).

In step 1, it is necessary that $\boldsymbol{B}$ is non-singular. It is difficult to prove the non-singularity of $\boldsymbol{B}$ for any positive integer of d, however, we can show that if a in (7) is sufficiently large, $\boldsymbol{B}$ is non-singular for any σ_1. We have discussed the non-singularity of the matrix $\boldsymbol{B}$ in [10].

When we minimize the integral in step 2, we must be careful to keep the values of λ_i and λ_p larger than zero. We consider the minimum at the log-scale because we can treat smaller time-scales more carefully.

4. *MAP/G/1/K*

In this section, we briefly summarize the analytical results of *MAP/G/1/K*, a single server queue with finite capacity and Markovian arrival process [11]. Note that *MMPP/D/1/K* is the special case of *MAP/G/1/K*.

In *MAP/G/1/K*, the customer arrives at the system according to the *MAP* represented by $(\boldsymbol{C}, \boldsymbol{D})$, where $\boldsymbol{C}$ and $\boldsymbol{D}$ are $M \times M$ matrices. The service time S is generally and identically distributed with the distribution function $S(x)$.

The irreducible matrix $\boldsymbol{C}+\boldsymbol{D}$ is the infinitesimal generator of the underlying Markov process restricted to the states $\{1, \cdots, M\}$. Let $\boldsymbol{\pi}$ denote the stationary vector of $\boldsymbol{C}+\boldsymbol{D}$, i.e.

$$\boldsymbol{\pi}(\boldsymbol{C}+\boldsymbol{D}) = \boldsymbol{0}, \quad \boldsymbol{\pi e} = 1, \tag{12}$$

where $\boldsymbol{e}$ denotes the column vector of ones.

Let $\boldsymbol{A}_k$ $(k \ge 0)$ denote an $M \times M$ matrix whose (i,j)th element represents the conditional probability that k customers arrive at the system during a service time of a customer and the underlying Markov chain is in phase j at the end of the service given that the underlying Markov chain is in phase i at the beginning of the service. Then $\boldsymbol{A}_k$ satisfies the following equation

$$\sum_{k=0}^{\infty} \boldsymbol{A}_k z^k = \int_0^{\infty} e^{(\boldsymbol{C}+z\boldsymbol{D})x} dS(x).$$

We also define $\boldsymbol{B}_k$ $(k \ge 1)$ as

$$\boldsymbol{B}_k = \sum_{n=k}^{\infty} \boldsymbol{A}_k.$$

Let $\boldsymbol{x}_k$ $(0 \le k \le K-1)$ denote a $1 \times M$ vector whose ith element represents the stationary joint probability that the number of customers in the system at departures is k and the phase of the arrival process is i. Then the transition probability matrix $\boldsymbol{P}$ is given by

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{EA}_0 & \boldsymbol{EA}_1 & \cdots & \boldsymbol{EA}_{K-2} & \boldsymbol{EB}_{K-1} \\ \boldsymbol{A}_0 & \boldsymbol{A}_1 & \cdots & \boldsymbol{A}_{K-2} & \boldsymbol{B}_{K-1} \\ \boldsymbol{O} & \boldsymbol{A}_0 & \cdots & \boldsymbol{A}_{K-3} & \boldsymbol{B}_{K-2} \\ \boldsymbol{O} & \boldsymbol{O} & \cdots & \boldsymbol{A}_{K-4} & \boldsymbol{B}_{K-3} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{O} & \boldsymbol{O} & \cdots & \boldsymbol{A}_1 & \boldsymbol{B}_2 \\ \boldsymbol{O} & \boldsymbol{O} & \cdots & \boldsymbol{A}_0 & \boldsymbol{B}_1 \end{bmatrix},$$

where $\boldsymbol{E} = (-\boldsymbol{C})^{-1}\boldsymbol{D}$. $\boldsymbol{x}_k$ satisfies the following equations

$$\boldsymbol{x}_k = \boldsymbol{x}_0 \boldsymbol{EA}_k + \sum_{n=1}^{k+1} \boldsymbol{x}_n \boldsymbol{A}_{k+1-n}, \quad 0 \le k \le K-2,$$

$$\boldsymbol{x}_{K-1} = \boldsymbol{x}_0 \boldsymbol{EB}_{K-1} + \sum_{n=1}^{K-1} \boldsymbol{x}_n \boldsymbol{B}_{K-n}.$$

We can calculate $\boldsymbol{x}_k$'s in the iterative manner [11].

Let $\boldsymbol{y}_k$ $(0 \le k \le K)$ denote a $1 \times M$ vector whose ith element is the stationary joint probability that the number of customers in the system is k and the phase of the arrival process is i at an arbitrary time. We define the vector probability generating function $\boldsymbol{y}(z)$ as

$$\boldsymbol{y}(z) = \sum_{k=0}^{K} \boldsymbol{y}_k z^k.$$

In general, the following relation between the probability generating function for arbitrary points and that for departure points is hold under any service disciplines [12]

$$\boldsymbol{y}(z)(\boldsymbol{C}+z\boldsymbol{D})=\lambda(z-1)\boldsymbol{x}(z), \tag{13}$$

where λ is the mean arrival rate of the *MAP* and given by $\lambda = \boldsymbol{\pi}\boldsymbol{D}\boldsymbol{e}$. Suppose that all arriving customers can accommodate the system and that the customer who finds the system full immediately leave the system. Then $\boldsymbol{x}(z)$ in (13) is given by

$$\begin{aligned}\boldsymbol{x}(z) &= \frac{1-\boldsymbol{y}_0\boldsymbol{e}}{\rho}\sum_{k=0}^{K-1}\boldsymbol{x}_k z^k \\ &\quad + \frac{\rho-(1-\boldsymbol{y}_0\boldsymbol{e})}{\rho}\frac{\boldsymbol{y}_K\boldsymbol{D}}{\boldsymbol{y}_K\boldsymbol{D}\boldsymbol{e}}z^K,\end{aligned} \tag{14}$$

where $\rho = \lambda E[S]$. Comparing the coefficients of z^0 in (14), we obtain

$$\boldsymbol{y}_0 = \lambda\frac{1-\boldsymbol{y}_0\boldsymbol{e}}{\rho}\boldsymbol{x}_0(-\boldsymbol{C})^{-1}.$$

Multiplying both sides of the above equation by $\boldsymbol{e}$ yields

$$\boldsymbol{y}_0\boldsymbol{e} = \frac{\boldsymbol{x}_0(-\boldsymbol{C})^{-1}}{E[S]+\boldsymbol{x}_0(-\boldsymbol{C})^{-1}}.$$

Similarly, comparing the coefficients of z^k ($k = 1,\cdots,K-1$) in (14) yields

$$\boldsymbol{y}_k\boldsymbol{C}+\boldsymbol{y}_{k-1}\boldsymbol{D} = (1-\boldsymbol{y}_0\boldsymbol{e})(\boldsymbol{x}_{k-1}-\boldsymbol{x}_k)/E[S].$$

From this, $\boldsymbol{y}_k$ ($k=1,\cdots,K-1$) can be calculated by

$$\begin{aligned}\boldsymbol{y}_k &= \boldsymbol{y}_{k-1}\boldsymbol{D}(-\boldsymbol{C})^{-1} \\ &\quad +(1-\boldsymbol{y}_0\boldsymbol{e})(\boldsymbol{x}_{k-1}-\boldsymbol{x}_k)(-\boldsymbol{C})^{-1}/E[S].\end{aligned}$$

Finally, we obtain $\boldsymbol{y}_K$ from$\boldsymbol{y}_K = \boldsymbol{\pi}-\sum_{k=0}^{K-1}\boldsymbol{y}_k$. Hence the blocking probability P_b is given by

$$P_b = \boldsymbol{y}_K\boldsymbol{e}. \tag{15}$$

The loss probability P_l is defined as the ratio of lost customers to the arriving customers. Then P_l satisfies the following relationship between the server utilization and the offered load

$$1-\boldsymbol{y}_0\boldsymbol{e} = \rho(1-P_l).$$

Therefore P_l is given by

$$P_l = 1-\frac{(1-\boldsymbol{y}_0\boldsymbol{e})}{\rho}. \tag{16}$$

5. Numerical Results

In this section, we show some numerical examples of the loss probability for *MMPP*$/D/1/K$ under several conditions. We calculate P_l according to the way of Sect. 4 and [11]. Here we set the followings:

$$\boldsymbol{C} = \boldsymbol{Q}-\boldsymbol{\Lambda}, \quad \boldsymbol{D} = \boldsymbol{\Lambda},$$

where $\boldsymbol{Q}$ is the infinitesimal generator of *MMPP* and $\boldsymbol{\Lambda}$ is the arrival rate matrix. $\boldsymbol{Q}$ and $\boldsymbol{\Lambda}$ are calculated by our fitting method described in Sect. 3. Table 2 shows the values of preliminary parameters used in Figs. 2 to 11. We calculate P_l varying the mean service time, i.e., traffic intensity (offered load) ρ. We also investigate the behavior of P_l in terms of system size K. In all figures, TS denotes the time-scale, K the system size, and ρ the traffic intensity.

Table 2 Parameters.

Parameter	Value
λ	1.0
σ^2	0.6
$m_{\min}$	10^2
$m_{\max}$	10^4, 10^5, 10^6, 10^7
H	0.6, 0.7, 0.8, 0.9
d	4

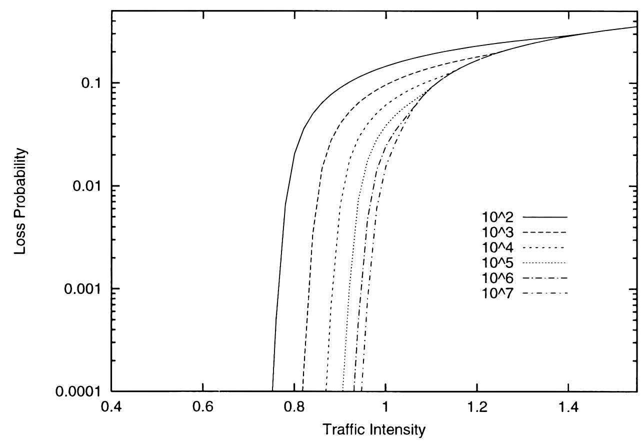

Fig. 1 Loss probability. ($H = 0.8$, $K = 100$)

5.1 Loss Behavior of *MMPP*$/D/1/K$

It is known that there exists a CTS in which the correlation structure in the input process has an impact on the loss probability. We first investigate the time-scale impact on the loss probability. Figure 1 shows the loss probabilities with $H = 0.8$ and $K = 100$ in cases of $TS = 10^2$ to 10^7. The horizontal axis means the traffic intensity and the vertical axis represents the loss probability. The loss probabilities are calculated with the same parameter values as Table 2 except $m_{\min} = 3$ and $d = 2$. From Fig. 1, we observe that the loss probability becomes small as the time-scale increases. This implies that the CTS of resulting *MMPP*$/D/1/K$ is quite small and hence we can expect the less variation of loss probability with time-scales large enough. In the following, we investigate the loss probability under the long time-scale for details.

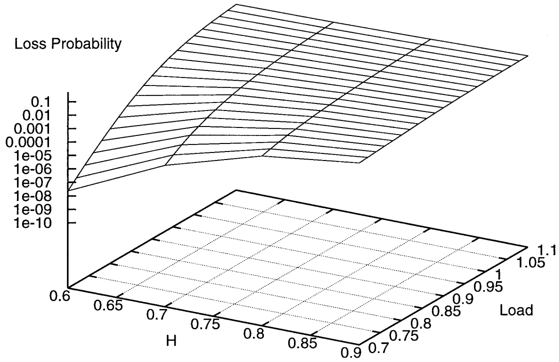

Fig. 2 Loss probability. ($K = 30$, $TS = 10^5$)

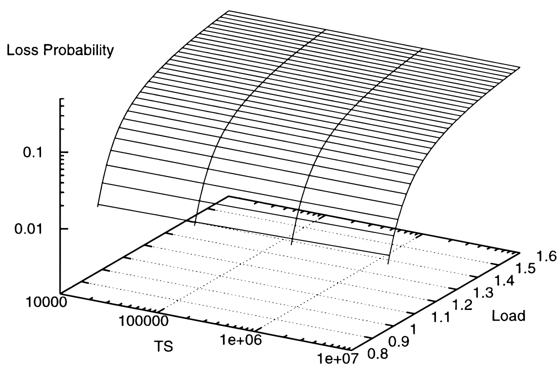

Fig. 3 Loss probability. ($H = 0.6$, $K = 50$)

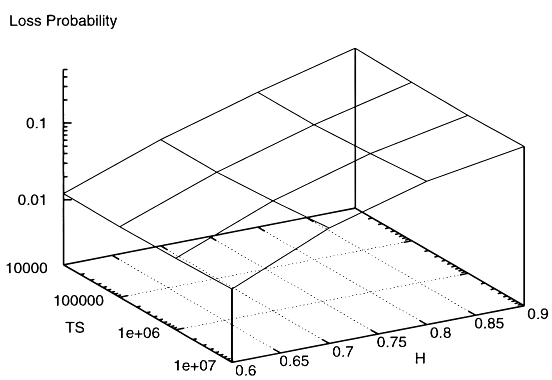

Fig. 4 Loss probability. ($\rho = 1.0$, $K = 100$)

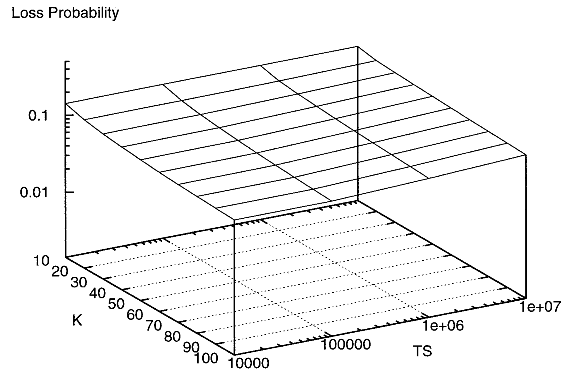

Fig. 5 Loss probability. ($\rho = 1.0$, $H = 0.8$)

Figure 2 represents the loss probability with $TS = 10^5$ and $K = 30$. The loss probability is calculated with varying H and ρ. We observe that the loss probability becomes large as H and ρ are getting large. This tendency agrees with our intuition.

Figure 3 shows the loss probability with $K = 50$ and $H = 0.6$. The loss probability is calculated with varying TS and ρ. From Fig. 3, we observe that the loss probability becomes large as ρ becomes large. However there is little difference of loss probabilities in terms of the time-scale.

Figure 4 illustrates the loss probability with $\rho = 1.0$ and $K = 100$. The loss probability is calculated with varying TS and H. From the figure, we observe that the loss probability increases gradually when H becomes large. We also observe that the system size affects the loss probability. However, there are no great differences when the time-scale changes.

Figure 5 represents the loss probability with $\rho = 1.0$ and $H = 0.8$, we observe that the loss probability decreases gradually when K becomes large. However it does not change greatly by the time-scale.

From Figs. 3 to 5, we observe that the time-scale does not have strong impact on the loss probability. This implies that the time-scale greater than or equal to 10^4 is long enough for capturing the stationary loss behavior of queueing system with pseudo self-similar input. Therefore, in the next subsection, we compare the loss probability obtained from resulting $MMPP/D/1/K$ with simulation and investigate the accuracy of $MMPP/D/1/K$.

5.2 Comparison of $MMPP/D/1/K$ with Simulation

For our simulation, simulated self-similar traffic trace is needed. We generate fractional Brownian traffic (FBT) based on the FBM with the random midpoint displacement (RMD) algorithm [5] and use it as self-similar traffic. The readers are referred to [5] for details. Using RMD algorithm, we generated sample sequences with nominal $H = 0.7,\ 08$ and 0.9, respectively, keeping $\lambda = 1.0$ and $\sigma^2 = 0.6$ as in Table 2.

Figures 6 to 11 show the comparison of analytical results with simulation. In all figures, horizontal axis means the traffic intensity and the vertical axis represents the loss probability. We plot results of $MMPP/D/1/K$ under $TS = 10^4,\ 5,\ 6,$ and 7 with lines and simulation results with cross points. Figures 6, 8 and 10 are for $K = 10$ while Figs. 7, 9 and 11 for $K = 100$.

When $K = 10$ (Figs. 6, 8 and 10), analytical results give upper bounds of simulation and the discrepancy between analytical and simulation results becomes small as the traffic intensity increases. In particular, analytical results exhibit the good agreement with sim-

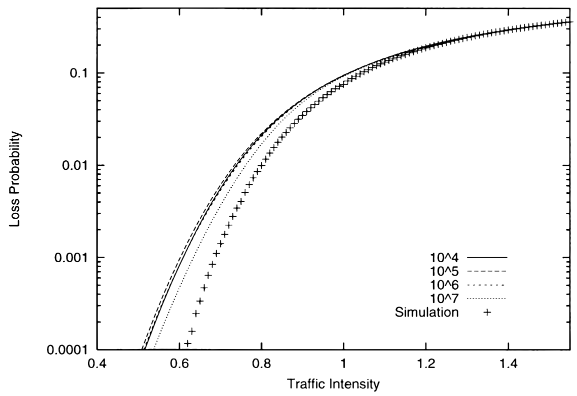

Fig. 6 Loss probability. ($H=0.7$, $K=10$)

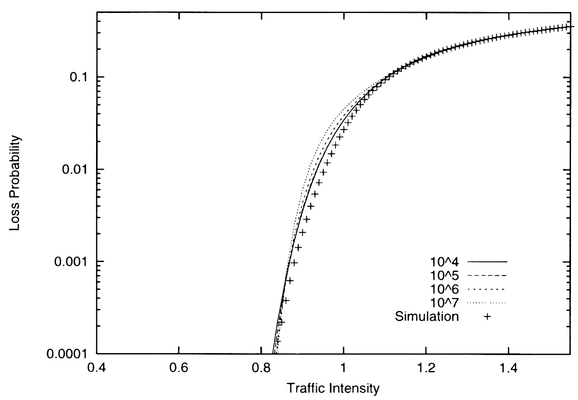

Fig. 7 Loss probability. ($H=0.7$, $K=100$)

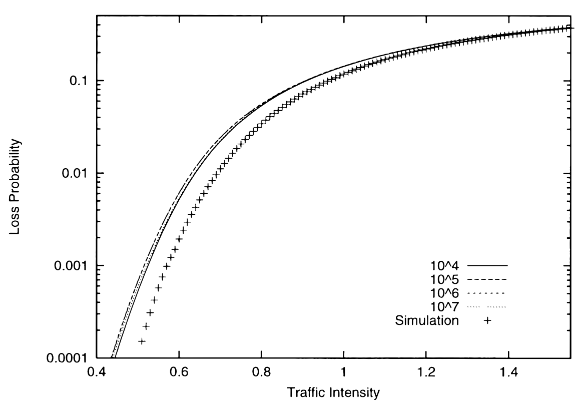

Fig. 8 Loss probability. ($H=0.8$, $K=10$)

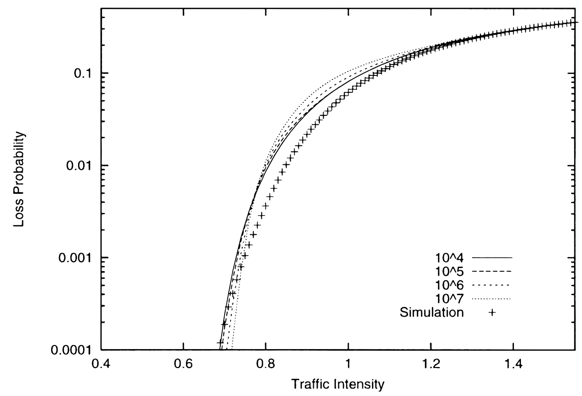

Fig. 9 Loss probability. ($H=0.8$, $K=100$)

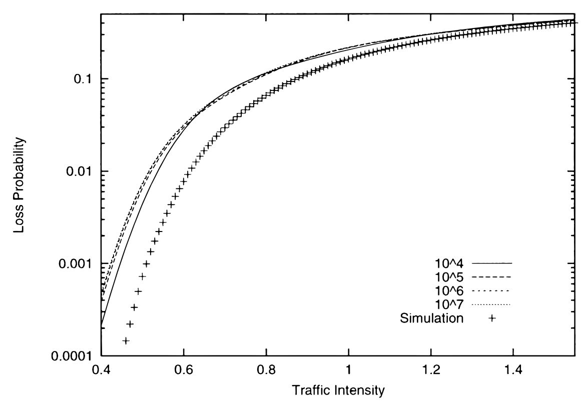

Fig. 10 Loss probability. ($H=0.9$, $K=10$)

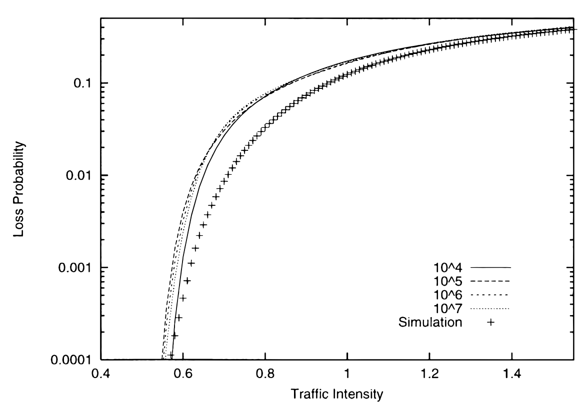

Fig. 11 Loss probability. ($H=0.9$, $K=100$)

ulation when the traffic intensity becomes greater than 1.0.

From Figs. 7 and 9, we observe that the agreement of analytical results with simulation in the case of $K=100$ becomes better than that under $K=10$. Note that the variance fitting method fails to predict the tail distribution of queue length of $MMPP/D/1$ even when the Hurst parameter is moderate [10]. In a practical sense, the finite queueing system is more desirable than the system with infinite buffer and hence this result is quite attractive for the buffer design.

Unfortunately, from Fig. 11, the discrepancy between analytical results and simulation does not improve so much. This is mainly due to the large Hurst parameter. However, analytical results succeed in giving upper bounds which seems to be relatively tight. Therefore the approach of pseudo self-similar input seems to be still useful even when the Hurst parameter is quite large.

6. Conclusion

In this paper, we investigated the loss probability of $MMPP/D/1/K$ where $MMPP$ is generated so as to mimic the variance-time curve of the self-similar process over several time-scales. From the numerical examples, we observed that the loss probability is sensitive to the offered load, system size and Hurst parameter while not sensitive to the time-scale.

Our numerical results represent a significant meaning when we consider modeling and analysis of the queueing system with self-similar input. As we stated in Introduction, it is known that the fitting methods based on the second-order statistics of counts for the arrival process are not sufficient for predicting the queueing performance correctly. However, this is the case of queueing system with infinite buffer and not the case of that with finite buffer.

Our numerical results reveal that the time-scale does not have a strong impact on the loss probability of resulting $MMPP/D/1/K$ if we consider the time-scale long enough. In addition, from the comparison of analytical results with simulation, the queueing system with finite buffer and pseudo self-similar input gives the good approximation for the loss probability even when the Hurst parameter is large. Therefore, in a practical sense, the variance fitting method seems to be enough to predict the loss behavior of the finite queuing system with self-similar input if we consider the appropriate time-scale.

References

[1] A.T. Anderson and B.F. Nielsen, "A Markovian approach for modeling packet traffic with long-range dependence," IEEE J. Sel. Areas Commun., vol.16, no.5, pp.719–732, June 1998.

[2] D.R. Cox, "Long-range dependence: A review," in Statistics: An appraisal, eds. H.A. David and H.T. David, pp.55–74, Ames, IA: The Iowa State University Press, 1984.

[3] A. Erramilli, O. Narayan, and W. Willinger, "Fractal queueing models," in Frontiers in Queueing, ed. J.H. Dshalalow, pp.245–269, CRC Press, 1997.

[4] D.P. Heyman and T.V. Lakshman, "What are the implications of long-range dependence for VBR-video traffic engineering?," IEEE/ACM Trans. Networking, vol.4, no.3, pp.301–317, June 1996.

[5] W.-C. Lau, A. Erramilli, J.L. Wang, and W. Willinger, "Self-similar traffic generation: The random midpoint displacement algorithm and its properties," Proc. ICC'95, pp.466–472, Seattle, WA, 1995.

[6] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of ethernet traffic (extended version)," IEEE/ACM Trans. Networking, vol.2, no.1, pp.1–15, 1994.

[7] S. Robert and J.Y. Le Boudec, "New models for pseudo self-similar traffic," Perform. Eval., vol.30, pp.57–68, 1997.

[8] B.K. Ryu and A. Elwalid, "The importance of long-range dependence of VBR video traffic in ATM traffic engineering: Myths and realities," ACM SIGCOMM'96, pp.3–14, 1996.

[9] V. Paxson and S. Floyd, "Wide area traffic: The failure of Poisson modeling," IEEE/ACM Trans. Networking, vol.3, no.3, pp.226–244, 1995.

[10] T. Yoshihara, S. Kasahara, and Y. Takahashi, "Practical time-scale fitting of self-similar traffic with Markov-modulated Poisson process," Telecommunication Systems, vol.17, nos.1–2, pp.185–211, 2001.

[11] T. Takine, "Introduction to numerical solution techniques of finite-state Markov chains and its applications," Preprint, 1995.

[12] T. Takine and Y. Takahashi, "On the relationship between queue lengths at a random instant and at a departure in the stationary queue with BMAP arrivals," Stochastic Models, vol.14, no.3, pp.601–610, 1998.

[13] B. Tsybakov and N.D. Georganas, "On self-similar traffic in ATM queues: Definitions, overflow probability bound, and cell delay distribution," IEEE/ACM Trans. Networking, vol.5, no.3, pp.397–409, 1997.

Shoji Kasahara received the B.Eng., M.Eng. and Dr.Eng. degrees from Kyoto University, Kyoto, Japan, in 1989, 1991, and 1996, respectively. He was with the Educational Center for Information Processing, Kyoto University from 1993 to 1997 as an Assistant Professor. In 1996, he was a visiting scholar of University of North Carolina at Chapel Hill, NC, USA. In 1997, he moved to the Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, Ikoma, Nara, Japan, where he has been an Associate Professor since August 1998. His research interests include queueing theory and performance analysis of computer and communication systems. Dr. Kasahara is a member of the IEEE, the Information Processing Society of Japan and the Operations Research Society of Japan.