

PERFORMANCE ANALYSIS OF THE LEAKY BUCKET SCHEME WITH QUEUE LENGTH DEPENDENT ARRIVAL RATES

DOO IL CHOI AND SANG MIN LEE

ABSTRACT. In this paper, we analyze a leaky bucket (LB) scheme with queue length dependent arrival rates. In other words, if the queue length exceeds an appropriate threshold value on buffer, the arrivals need to be controlled. In ATM networks, if the congestion occurs, the input traffics must be controlled (reduced) for congestion resolution. By the bursty and correlated properties of traffics, the arrivals are assumed to follow a Markov-modulated Poisson process (MMPP). We derive the loss probability and the waiting time distribution for arbitrary cell. The analysis is done by using the embedded Markov chain and supplementary variable method. We also present some numerical examples to show the effects of our proposed LB scheme.

1. Introduction

In telecommunication networks concluding B-ISDN, it needs to control the input traffic so as to prevent the network from reaching an unacceptable congestion level. The leaky bucket (LB) scheme proposed by Turner [11] has been known as one of the most promising methods for preventive congestion control and policing functions. For the LB scheme without a data buffer, Butto et al. [2] analyzed the system with a on-off type bursty source as $G/D/1/N$ queue. Each burst has a duration Z (random variable) and a bit rate b bit/sec. During the burst, the cells are assumed to arrive periodically. In order to reduce the cell loss probability, the data buffer is installed in the LB scheme. All the following models have a data buffer. Many papers about the LB scheme have been investigated by the unslotted system (i.e., by continuous-time

Received March 13, 2006.

2000 Mathematics Subject Classification: 60K25, 68M20.

Key words and phrases: queuing analysis, leaky bucket, MMPP, telecommunication networks.

queueing system). With a Poisson arrival process as input process, Sidi et al. [9] analyzed the LB scheme with both the finite and infinite buffer capacity. They obtained the distributions of queue length, the waiting time and the inter-departure time. Kim et al. [6] also analyzed the finite buffer LB scheme with a Markov Modulated Poisson Process (MMPP) and derived the cell loss probability and the waiting time distribution. They showed the effects of the system parameters by varying the ratio of the arrival rates and the sojourn times of each arrival state for the MMPP. There are the analysis of LB scheme by the slotted system (i.e., by the discrete-time queueing system). Ahmadi et al. [1] and Sohraby et al. [10] analyzed the LB scheme with the cell transmission at only slot boundary. In Ahmadi et al. [1], the cell arrivals in a slot are characterized by a batch process, and the arrivals in successive slots are independent and identically distributed. The solution method is based on the matrix analytic approach. Sohraby et al. [10] extended the batch arrival process to a finite state discrete Markovian Arrival Process with infinite buffer and obtained the queue length distribution. For a Poisson and a discrete-time MMPP arrival process with finite and infinite buffer, Wu et al. [12] analyzed the queue length distribution and obtained the ratio of the variance of the interdeparture times to the variance of the interarrival times.

On the other side, in order to support Quality of Service (QoS) of traffics there has been the analysis of the diverse LB schemes. The representatives are the LB scheme with a dynamic token generation interval or priority [3, 4, 5, 7, 8, 13]. In papers [3, 7, 8] the token generation interval alternates according to the buffer occupancy, maintaining the same weighted average token generation interval reserved at call setup by an admission controller. Lee and Un [7] analyzed the LB scheme with on-off data source in which the token generation interval during on-period is somewhat smaller than that of off-period. They analyzed the performance of the LB scheme by using the fluid-flow method. Recently, Choi et al. [3, 4, 8] analyzed the LB scheme with the dynamic token generation intervals, in which the token generation interval is changed according to buffer occupancy. They assumed the arrival process to be the MMPP in continuous-time ([8]) and Markov Modulated Bernoulli Process (MMBP) in discrete-time ([3, 4]). Also, Choi et al [5] analyzed the LB scheme with priority. They classified traffics into two types according to their priority. The high priority type is transmitted by queue-length-threshold (QLT) scheduling policy depending on buffer

occupancy of low priority cells. The special case of QLT scheduling policy is just the Head of Line (HOL) priority scheme [13].

In B-ISDN, by the more many cell generation and transmission, the network may be reached at congestion level. The information on network maybe lost, and finally the network resources become useless. Thus, the input traffic must be controlled at appropriate level of buffer occupancy. We analyze the LB scheme with queue length dependent arrival rates. In other words, we place a threshold L on the buffer (that is, it is to indicate congestion state of network). According to whether the queue length exceeds the threshold L or not, the arrivals are controlled. We assume the arrivals to be an MMPP by considering the bursty and correlated input traffic.

2. Model description

There is a buffer to accommodate the arriving cells, and a token pool to store tokens generated. The cells arrive according to an MMPP, and they are queued in buffer with finite capacity K if no tokens are available. The token pool has a finite capacity M , so that the newly generated tokens are discarded when the token pool is full. Tokens are generated at every constant time T . Each token allows a single cell to be transmitted, and the token following a transmission is removed from the token pool. We place a threshold L on the buffer and control the arrival according to buffer state (that is, queue length). The arrival rate is changed at only token generation instant. In other words, if the queue length at token generation instant exceeds the threshold L , the arrivals follow a MMPP with representation (Q, Λ_2) , where $\Lambda_2 = \text{diag}(\lambda_i^2)$. Otherwise, the arrivals follow another MMPP with representation (Q, Λ_1) , where $\Lambda_1 = \text{diag}(\lambda_i^1)$, $i = 1, 2, \dots, N$. The square matrix Q with size N is the infinitesimal generator of the underlying Markov process $J(t)$ with state space $\{1, 2, \dots, N\}$. The steady-state probability vector Π of the underlying Markov process $J(t)$ is given by solving the equations:

$$\Pi Q = 0, \quad \Pi e = 1,$$

where e and 0 are vectors of size N consist of all ones and zeros, respectively.

Let $M_1(t)(M_2(t))$ be the number of arrivals by $\Lambda_1(\Lambda_2)$ during the interval $(0, t]$. Now we define the conditional probabilities

$$p_{i,j}^r(n, t) = Pr\{M_r(t) = n, J(t) = j \mid M_r(0) = 0, J(0) = i\},$$

$$r = 1, 2, n \geq 0.$$

By the Chapman-Kolmogorov's forward equations, we have the differential-difference equations for the matrices $P_r(n, t) \triangleq (p_{i,j}^r(n, t))_{1 \leq i, j \leq N}$:

$$P_r'(n, t) = P_r(n, t)(Q - \Lambda_r) + P_r(n-1, t)\Lambda_r, \quad r = 1, 2, \quad n \geq 0,$$

where $P_r(-1, t)$ is the matrix 0. Then, it is easily shown that the matrix $P_r(n, t)$ has the probability generating function

$$\begin{aligned} \bar{P}_r(z, t) &\triangleq \sum_{n=0}^{\infty} P_r(n, t)z^n \\ &= e^{[Q+(z-1)\Lambda_r]t}, \quad |z| \leq 1, \quad r = 1, 2. \end{aligned}$$

3. Analysis

3.1. System state distribution at token generation instants

We consider the system state at token generation instants $0, T, 2T, \dots$. Let $B(n)$ ($T(n)$) be the number of cells (tokens) in buffer (token pool, respectively) just after the n th token generation instant. Since the arriving cells wait in buffer only if there is no token, we so express the state of buffer and token pool as follows

$$N(n) \triangleq B(n) + M - T(n).$$

That is, if there are i tokens in token pool ($B(n) = 0$), then $N(n) = M - i$. Also, if there are i cells in buffer ($T(n) = 0$), then $N(n) = M + i$. Finally, the process $\{(N(n), J(n)), n \geq 0\}$ forms a 2-dimensional Markov chain with finite state space $\{(0, 0), \dots, (0, N), (1, 0), \dots, (M + K - 1, N)\}$. In this paper, we consider the stationary probability distribution of the system state. Define the matrix A_k^r as

$$A_k^r = P_r(k, T), \quad k = 0, 1, \dots, \quad r = 1, 2.$$

Then, the transition probability matrix \bar{Q} of the Markov chain $\{(N(n), J(n)), n \geq 0\}$ is given by

$$\bar{Q} = \begin{pmatrix} A_0^1 + A_1^1 & A_2^1 & \dots & A_{M+L}^1 & A_{M+L+1}^1 & A_{M+L+2}^1 & \dots & A_{M+K-1}^1 & \bar{A}_{M+K}^1 \\ A_0^1 & A_1^1 & \dots & A_{M+L-1}^1 & A_{M+L}^1 & A_{M+L+1}^1 & \dots & A_{M+K-2}^1 & \bar{A}_{M+K-1}^1 \\ 0 & A_0^1 & \dots & A_{M+L-2}^1 & A_{M+L-1}^1 & A_{M+L}^1 & \dots & A_{M+K-3}^1 & \bar{A}_{M+K-2}^1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & A_1^1 & A_2^1 & A_3^1 & \dots & A_{K-L}^1 & \bar{A}_{K-L+1}^1 \\ 0 & 0 & \dots & A_0^2 & A_1^2 & A_2^2 & \dots & A_{K-L-1}^2 & \bar{A}_{K-L}^2 \\ 0 & 0 & \dots & 0 & A_0^2 & A_1^2 & \dots & A_{K-L-2}^2 & \bar{A}_{K-L-1}^2 \\ & & & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ & & & 0 & 0 & 0 & \dots & A_1^2 & \bar{A}_2^2 \\ & & & 0 & 0 & 0 & \dots & A_0^2 & \bar{A}_1^2 \end{pmatrix}$$

where $\bar{A}_k^r = \sum_{n=k}^\infty A_n^r$.

It is shown easily the \bar{Q} is a stochastic matrix. The steady-state probability vector $x = (x_0, x_1, \dots, x_{M+K-1})$ of the Markov chain $\{(N(n), J(n)), n \geq 0\}$ finally is given by solving the equations

$$x \bar{Q} = x, \quad x e = 1.$$

3.2. System state distribution at an arbitrary instants

In this subsection we derive the system state distribution at an arbitrary time. Let \tilde{T} (\hat{T}) be the elapsed (remaining, respectively) token generation time. Define the limiting probabilities and the vectors

$$y_n(j) = \lim_{t \rightarrow \infty} Pr\{N(t) = n, J(t) = j\},$$

$$y_n = (y_n(1), y_n(2), \dots, y_n(N)).$$

We furthermore define the joint probability distribution of the system state and the remaining token generation time at arbitrary time τ as

$$\alpha_{n,j}(t)dt = Pr\{N(\tau) = n, J(\tau) = j, t < \hat{T} \leq t + dt\},$$

and the Laplace transform of $\alpha_{n,j}(t)$

$$\alpha_{n,j}^*(s) = \int_0^\infty e^{-st} \alpha_{n,j}(t)dt, \quad j = 1, 2, \dots, N.$$

Let $\alpha_n^*(s) = (\alpha_{n,1}^*(s), \alpha_{n,2}^*(s), \dots, \alpha_{n,N}^*(s))$. In order to derive $\alpha_{n,j}^*(s)$, we define the conditional probability $\beta_r(n, j_1, j_2, t)dt$ as

$$\beta_r(n, j_1, j_2, t)dt = Pr\{ n \text{ arrivals by } \Lambda_r \text{ during } \tilde{T}, J(\tau) = j_2, \\ t < \hat{T} \leq t + dt \mid J(\bar{\tau}) = j_1\}, \quad n \geq 0,$$

where $\bar{\tau}$ is the starting time of the token generation interval which includes the time τ . We also define the Laplace transform $(\beta_r^*(n, j_1, j_2, s))$ of $\beta_r(n, j_1, j_2, t)$ and the matrix $\beta_r^*(n, s)$ with $\beta_r^*(n, j_1, j_2, s)$ as (j_1, j_2) -element:

$$\beta_r^*(n, j_1, j_2, s) = \int_0^\infty e^{-st} \beta_r(n, j_1, j_2, t) dt,$$

$$\beta_r^*(n, s) = (\beta_r^*(n, j_1, j_2, s))_{1 \leq j_1, j_2 \leq N}, \quad n \geq 0, \quad r = 1, 2.$$

Then, the vectors $\alpha_n^*(s)$ satisfy the following equations:

For $0 \leq n < M + L$

$$\alpha_n^*(s) = \sum_{k=0}^n x_k \beta_1^*(n-k, s).$$

For $M + L \leq n < M + K$

$$\alpha_n^*(s) = \sum_{k=0}^{M+L-1} x_k \beta_1^*(n-k, s) + \sum_{k=M+L}^n x_k \beta_2^*(n-k, s).$$

It is easy to show that $\beta_r^*(n, s)$ is given by

$$\beta_r^*(n, s) = \frac{1}{T} \left[\sum_{k=0}^n A_k^r R_{n-k}^r(s) - e^{-sT} R_n^r(s) \right],$$

where $R_n^r(s) = (sI - \Lambda_r + Q)^{-1} \{ \Lambda_r (\Lambda_r - sI - Q)^{-1} \}^n$, $r = 1, 2$. Finally, substituting $\beta_r^*(n, s)$ ($r = 1, 2$) into above equations $\alpha_n^*(s)$, we obtain $y_n = \alpha_n^*(0)$:

For $0 \leq n < M + L$,

$$y_n = \frac{1}{T} \sum_{k=0}^n x_k \left[\sum_{l=0}^{n-k} A_l^1 (Q - \Lambda_1)^{-1} \{ \Lambda_1 (\Lambda_1 - Q)^{-1} \}^{n-k-l} \right. \\ \left. - (Q - \Lambda_1)^{-1} \{ \Lambda_1 (\Lambda_1 - Q)^{-1} \}^{n-k} \right].$$

For $M + L \leq n < M + K$

$$y_n = \frac{1}{T} \sum_{k=0}^{M+L-1} x_k \left[\sum_{l=0}^{n-k} A_l^1 (Q - \Lambda_1)^{-1} \{ \Lambda_1 (\Lambda_1 - Q)^{-1} \}^{n-k-l} - (Q - \Lambda_1)^{-1} \{ \Lambda_1 (\Lambda_1 - Q)^{-1} \}^{n-k} \right] + \frac{1}{T} \sum_{k=M+L}^n x_k \left[\sum_{l=0}^{n-k} A_l^2 (Q - \Lambda_2)^{-1} \{ \Lambda_2 (\Lambda_2 - Q)^{-1} \}^{n-k-l} - (Q - \Lambda_2)^{-1} \{ \Lambda_2 (\Lambda_2 - Q)^{-1} \}^{n-k} \right],$$

and

$$y_{M+K} = \Pi - \sum_{n=0}^{M+K-1} y_n.$$

Thus, by using the stationary queue length distribution y_n , we obtain the following loss probability (P_{loss}) for arbitrary arriving cell:

$$P_{\text{loss}} = \frac{y_{M+K} (\Lambda_1^* + \Lambda_2^*) e}{\sum_{k=0}^{M+K} y_k (\Lambda_1^* + \Lambda_2^*) e} = \frac{y_{M+K} (\Lambda_1^* + \Lambda_2^*) e}{\Pi (\Lambda_1^* + \Lambda_2^*) e},$$

where $\Lambda_1^* = \sum_{k=0}^{M+L-1} x_k e \Lambda_1$, $\Lambda_2^* = \sum_{k=M+L}^{M+K-1} x_k e \Lambda_2$.

REMARK. If there is no change of arrival (i.e., $\Lambda_1 = \Lambda_2 = \Lambda$, also $L = K$), then the system state distribution at arbitrary time is given by

$$y_n = \frac{1}{T} \sum_{k=0}^n x_k \left[\sum_{l=0}^{n-k} A_l (Q - \Lambda)^{-1} \{ \Lambda (\Lambda - Q)^{-1} \}^{n-k-l} - (Q - \Lambda)^{-1} \{ \Lambda (\Lambda - Q)^{-1} \}^{n-k} \right],$$

where $A_l^1 = A_l^2 = A_l$.

This is just the corresponding result of the ordinary LB scheme without threshold [4].

4. Waiting time distribution

We derive the distribution of waiting time (W) for a cell which is not blocked. Let's a tagged cell arrive at time t . Then, the tagged cell may find the system in one of the following states:

- 1) There are n ($1 \leq n \leq M$, i.e., $0 \leq N(t) \leq M - 1$) tokens in the token pool. In this case, the tagged cell is transmitted immediately and the waiting time is zero.

- 2) There is no token and no cell queued in the buffer (i.e., $N(t) = M$). In this case, the tagged cell will be transmitted at the next token generation instant, the waiting time, thus, is \hat{T} .
- 3) There are n ($1 \leq n \leq K - 1$, i.e., $M + 1 \leq N(t) \leq M + K - 1$) cells queued in the buffer. In this case, the tagged cell may have to wait until these cells are transmitted. Thus, the waiting time is $\hat{T} + nT$.

By considering above cases, we can derive the Laplace transform of the waiting time for the cell which is not blocked:

$$\begin{aligned}
 W^*(s) &= E[e^{-sW}] \\
 &= \frac{1}{1 - P_{\text{loss}}} \left[\sum_{n=0}^{M-1} y_n + \alpha_M^*(s) + \sum_{n=M+1}^{M+K-1} (e^{-sT})^{n-M} \alpha_n^*(s) \right] \\
 &\quad \times (\Lambda_1^* + \Lambda_2^*)e / \left(\sum_{n=0}^{M+K} y_n (\Lambda_1^* + \Lambda_2^*)e \right).
 \end{aligned}$$

Then, the mean waiting time of a cell is given by differentiating this Laplace transform:

$$E[W] = (-1) \left. \frac{d}{ds} W^*(s) \right|_{s=0}.$$

5. Numerical examples

In this section, we give some numerical examples to show the effects of our proposed LB scheme. The followings are assumed for numerical examples. As an arrival of cells, we use a two-state MMPP with

$$Q = \begin{bmatrix} -\sigma_{12} & \sigma_{12} \\ \sigma_{21} & -\sigma_{21} \end{bmatrix}, \Lambda_1 = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \Lambda_2 = \Lambda_1/2.$$

In all numerical examples, we take buffer sizes $K = 10$, $\sigma_{12}(= \sigma_{21}) = 0.1$ and $\lambda_2/\lambda_1 = 6$. We also set the token generation interval(T) equal to 1. The total effective arrival rate λ^* is given by

$$\lambda^* = \Pi(\Lambda_1^* + \Lambda_2^*)e.$$

Figs. 1 and 2 illustrate the loss probability and the mean waiting time, respectively, as a function of total effective arrival rate when $M = 5$. From the figures, for various values of the threshold value L , we can

observe that the loss probability and the mean waiting time increase as the threshold value L becomes large.

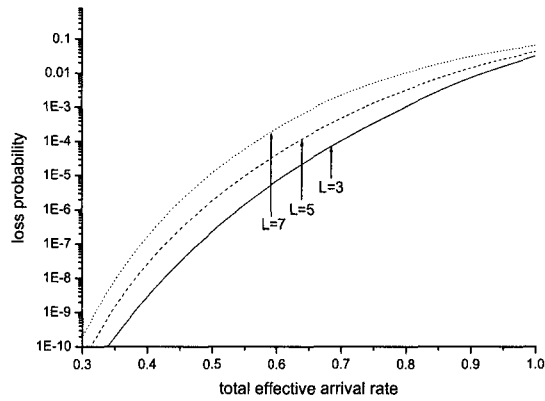


FIGURE 1. Loss probability against total effective arrival rate

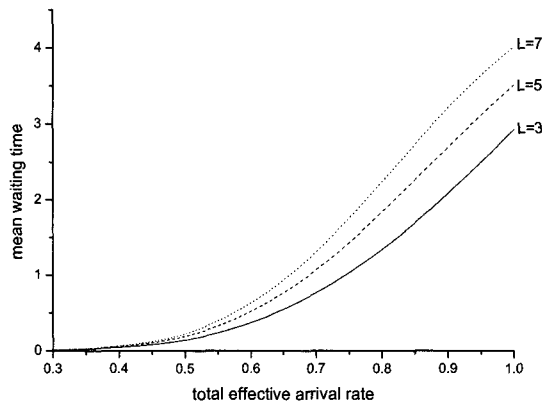


FIGURE 2. Mean waiting time against total effective arrival rate

For various values of the token pool size M when $L = 5$, Figs. 3 and 4 illustrate the loss probability and the mean waiting time, respectively, as a function of total effective arrival rate. From the figures, we can

observe that the loss probability and the mean waiting time decrease as the token pool size M becomes large.

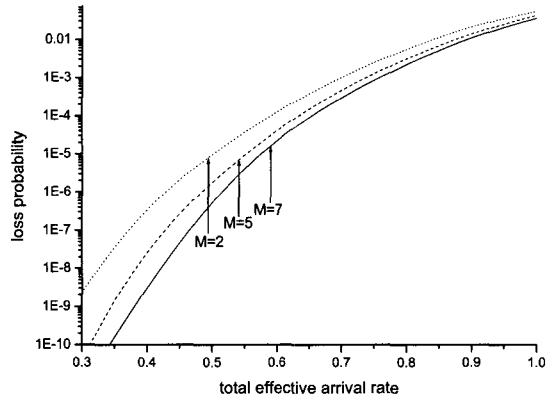


FIGURE 3. Loss probability against total effective arrival rate

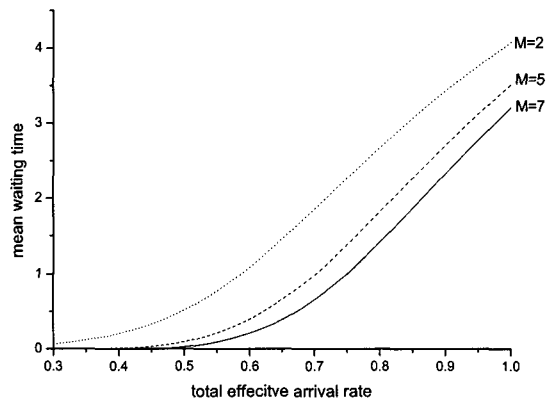


FIGURE 4. Mean waiting time against total effective arrival rate

In Figs. 5 and 6, we compare the loss probability and the mean waiting time of MMPP arrivals with those of poisson arrivals. From the figures, we can observe that the loss probability and the mean waiting

time with MMPP arrivals are larger than those of poisson arrivals. This comes from the burstiness of MMPP arrivals. It indicates the importance of modelling for arrival process of input traffic.

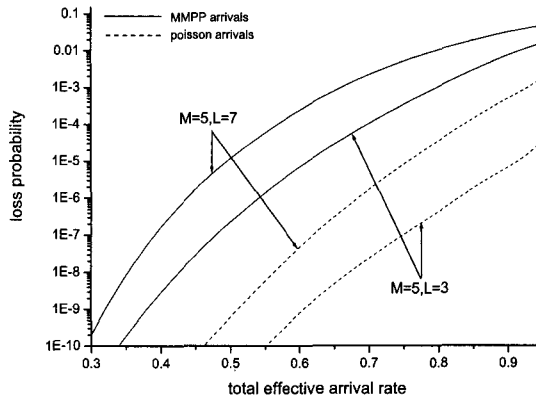


FIGURE 5. Loss probability against total effective arrival rate

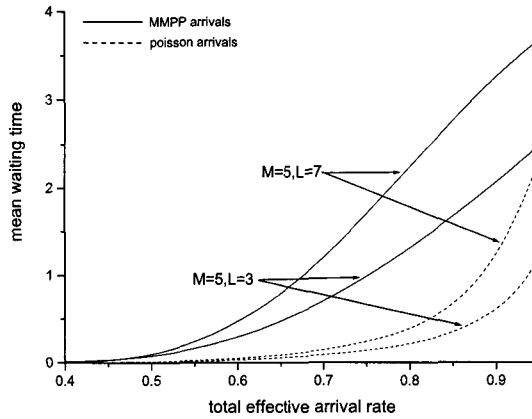


FIGURE 6. Mean waiting time against total effective arrival rate

When $L = K (= 10)$, our proposed LB scheme is just the ordinary LB scheme without threshold. In Figs. 7 and 8, we compare the our

proposed LB scheme with the ordinary LB scheme without threshold. We can certify that the our proposed LB scheme has the fairly enhanced performance than the ordinary LB scheme without threshold.

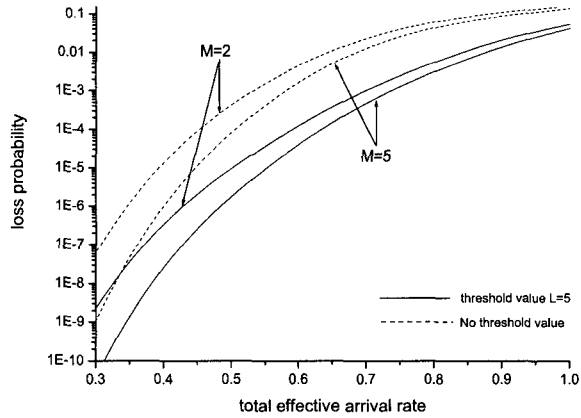


FIGURE 7. Loss probability against total effective arrival rate

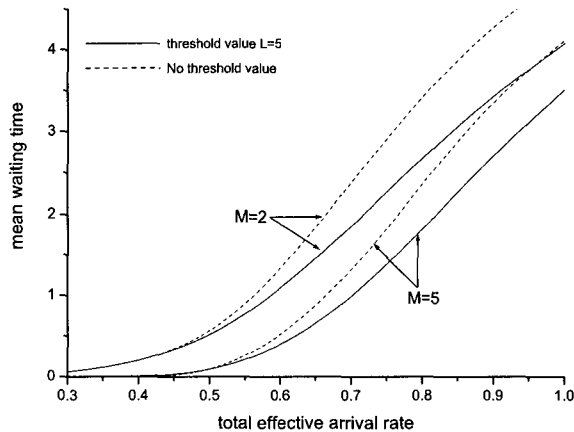


FIGURE 8. Mean waiting time against total effective arrival rate

References

- [1] H. Ahmadi, R. Guerin, and K. Sohraby, *Analysis of leaky bucket access control mechanism with batch arrival process*, IEEE Globecom'90 (1990), 344–349.
- [2] M. Butto, E. Cavallero, and A. Tonietti, *Effectiveness of the leaky bucket policing mechanism in ATM networks*, IEEE J. Sel. Areas Commun. **9** (1991), no. 3, 335–342.
- [3] B. D. Choi and D. I. Choi, *Discrete-time analysis of the leaky bucket scheme with threshold-based token generation intervals*, IEE Proc. Commun. **143** (1996), 105–111.
- [4] ———, *The discrete-time analysis of the leaky bucket scheme with dynamic leaky rate control*, Comm. Korean. Math. Soc. **13** (1998), no. 3, 603–627.
- [5] D. I. Choi, B. D. Choi, and D. K. Sung, *Performance analysis of priority leaky bucket scheme with queue-length-threshold scheduling policy*, IEE Proc. Commun. **145** (1998), no. 6, 395–401.
- [6] Y. H. Kim, B. C. Shin, and C. K. Un, *Performance analysis of leaky bucket bandwidth enforcement strategy for bursty traffics in an ATM networks*, Computer Networks ISDN Syst. **25**(1992), 295–303.
- [7] J. Y. Lee and C. K. Un, *Performance of dynamic rate leaky bucket algorithm*, IEE Electron. Lett. **29** (1993), no. 17, 1560–1561.
- [8] C. G. Park, B. D. Choi, and D. K. Sung, *Performance analysis of leaky bucket scheme with a threshold in the data buffer*, Computer Networks ISDN Syst. **29** (1997), 781–795.
- [9] M. Sidi, Z. Lin, I. Cidon, and I. Gopal, *Congestion control through input rate regulation*, Proceedings of IEEE Globecom'89 (1989).
- [10] K. Sohraby and M. Sidi, *On the performance of bursty and modulated sources subject to leaky bucket rate based access control schemes*, IEEE Trans. on Commun. **42** (1994), 477–487.
- [11] J. Tunner, *New directions in communications(or Which way to the information age)*, IEEE Communi. Mag. **24** (1986), 17–24.
- [12] G. L. Wu and J. W. Mark, *Discrete time analysis of leaky bucket congestion control*, Computer Networks ISDN Syst. **26** (1993), 79–94.
- [13] J. Zeng and L. F. Merakos, *Analysis of a priority leaky bucket regulator for virtual path traffic in ATM networks*, Proceedings of IEEE Globecom'94. **1994**, 606–611.

DOO IL CHOI, DEPARTMENT OF APPLIED MATHEMATICS, HALLA UNIVERSITY, KANGWON-DO 220-712, KOREA

E-mail: dichoi@halla.ac.kr

SANG MIN LEE, DEPARTMENT OF INDUSTRIAL ENGINEERING, KAIST 305-701, DAEJON-SHI, KOREA

E-mail: oobob@kaist.ac.kr