J. KSIAM Vol.7, No.1, 15-23, 2003

QUEUEING ANALYSIS OF THE HOL PRIORITY LEAKY BUCKET SCHEME

DOO IL CHOI

ABSTRACT. ATM networks provide the various kinds of service which require the different Quality of Services(QoS) such as loss and delay. By statistically multiplexing of traffics and the uncertainty and fluctuation of source traffic pattern, the congestion may occur. The leaky bucket scheme is a representative policing mechanism for preventive congestion control. In this paper, we analyze the HOL(Head-of-Line) priority leaky bucket scheme. That is, traffics are classified into real-time and nonreal-time traffic. The real-time traffic has priority over nonreal-time traffic for transmission. For proposed mechanism, we obtain the system state distribution, finally the loss probability and the mean waiting time of real-time and nonreal-time traffic. The simple numerical examples also are presented.

1. INTRODUCTION

ATM networks [1] provide the various kinds of services which require the different Quality of Services (QoS). Since the Broadband ISDN user terminals in ATM networks generate cells only when they have information to transmit and these cells are statistically multiplexed, the source traffic pattern has the uncertainty and fluctuates. In the ATM networks, the network resources are allocated through a negotiation between the user and the network during the connection establishment phase. To prevent the ATM network from reaching an unacceptable congestion level due to unexpected traffic variation or due to intentional excess of the negotiated parameters, it is necessary to monitor whether the traffic flow on every virtual channel connection conforms to the negotiated parameters or not. This function is called the policing or UPC (Usage Parameter Control) function.

The Leaky Bucket (LB) scheme [2] is a representative UPC mechanism taking into account the violation probability, dynamic reaction time and implementation complexity. However, the advent of fiber optic media and transmission technology increase the

AMS Mathematics Subject Classification : 60K25, 68M20

Key words and phrases. queueing analysis, HOL priority leaky bucket scheme, MMPP, ATM networks.

reliability of networks such that the needs for intensive error detections and corrections are no longer required. While congestion control is still needed to allocate the network resources fairly so that overload of one traffic does not deteriorate the performance of other traffics. Such congestion control must be a preventive control.

We analyze the Head of Line (HOL) priority leaky bucket scheme. This HOL priority leaky bucket scheme is more simple in implementation complexity than that of the leaky bucket scheme with QLT[6]. Also, this is support to the real-time traffic with distinct delay requirement.

In Section 2, we give the detailed description of the HOL priority leaky bucket scheme. In Section 3, we analyze the model when the arrival of the real-time traffic is Markov-modulated Poisson process (MMPP). Thus, the system state distribution is deriven by using the embedded Markov chain method. Finally, we obtain the loss probability and the mean waiting time. In Section 4, we give a special case when the arrivals of the real-time and nonreal-time traffic are Poisson processes. And the simple examples are given in Section 5.

2. Model description

There are two separate buffers to accommodate the cells of two types (that is, real-time and nonreal-time traffic) and a token pool to store tokens generated. The token is generated every constant interval T. The size of token pool is M. If the token pool is full, the newly generated tokens are discarded. The buffer size for real-time and nonreal-time traffic is assumed to be K_1 and K_2, respectively. The cells queued in each buffer are served on the first-come first-service basis.

The arrivals of real-time traffic is assumed to follow a Markov-modulated Poisson Process (MMPP) with representation (Q, Λ_1). This assumption is to reflect the burstiness and correlation of real-time traffic such as voice[3][4]. The $N \times N$ matrix Q is the infinitesimal generator of the underlying Markov process $J(t)$ with state space $1, 2, \cdots, N$. The $\Lambda_1 = \text{diag}(\lambda_i^1)$ is the arrival rate matrix. That is, if the underlying Markov process $J(t)$ is in state i $(i = 1, 2, \cdots, N)$, the arrival follows a Poisson process with rate λ_i^1. The MMPP arrival has been extensively used to model bursty traffics in ATM networks[5]. We assume the nonreal-time traffic such as data to follow a Poisson process with rate λ_2. We refer to real-time and nonreal-time traffic as type-1 and type-2.

The transmission of type-1 and type-2 is determined according to Head-of-Line priority schedule as follows (or see Fig. 1):

1) If there are any tokens in the token pool (in this case there are no cells waiting in the buffers), the cell arriving regardless of traffic type is transmitted, consuming a token.
2) If at least one type-1 cell exists in buffer, the cell of type-1 is transmitted.
3) The cells of type-2 are transmitted only if there is no cells of type-1.

Let π be the stationary probability vector of the underlying Markov process $J(t)$. Then π is given by solving the equations

$$\pi Q = 0 \qquad \pi e = 1$$

where e denotes a column vector with all elements equal to one.

3. Analysis

To derive the distribution of system state at token generation instants, we need to know the number of arrivals during the token generation interval T. Let $M(t)(M_1(t))$ be the number of total arrivals regardless of traffic type (type-1, respectively) during interval $(0, t]$. Now we define the conditional probabilities

$$\begin{aligned} p(n,t)_{ij} &= P\{M(t) = n,\ J(t) = j | M(0) = 0,\ J(0) = i\} \\ p_1(n,t)_{ij} &= P\{M_1(t) = n,\ J(t) = j | M_1(0) = 0,\ J(0) = i\} \\ & n \geq 0,\ 1 \leq i, j \leq N. \end{aligned}$$

By Chapman-Kolmogorov's forward equations, we have the following differential-difference equations for $N \times N$ matrices $P(n,t) \triangleq (p(n,t)_{ij})$ and $P_1(n,t) \triangleq (p_1(n,t)_{ij})$:

$$\begin{aligned} P'(n,t) &= P(n,t)(Q - \Lambda) + P(n-1,t)\Lambda \\ P_1'(n,t) &= P_1(n,t)(Q - \Lambda_1) + P_1(n-1,t)\Lambda_1 \end{aligned}$$

where $P(-1,t)$ and $P_1(-1,t)$ are the matrices 0 and $\Lambda = \Lambda_1 + \lambda_2 I$.
Then, it is easily shown that the matrices $P(n,t)$ and $P_1(n,t)$ have the probability generating functions

$$\begin{aligned} \overline{P}(z,t) &\triangleq \sum_{n=0}^{\infty} P(n,t) z^n \\ &= e^{R(z)t} \\ \overline{P}_1(z,t) &\triangleq \sum_{n=0}^{\infty} P_1(n,t) z^n \\ &= e^{R_1(z)t}, \qquad |z| \leq 1. \end{aligned}$$

where $R(z) = Q + (z-1)\Lambda$ and $R_1(z) = Q + (z-1)\Lambda_1$.

The effective arrival rate λ^* and λ_1^* by arrival Λ and Λ_1 are expressed by $\lambda^* = \pi\Lambda e$ and $\lambda_1^* = \pi\Lambda_1 e$ respectively.
Let $B_1(n)$ and $N_2(n)$ be the number of type-1 and type-2 cells queued in buffer just after the nth token generation epoch. Let $T(n)$ be the number of tokens in the token pool just after the nth token generation epoch. Since the cells wait in the buffer only if there is no token in the token pool, we can express the state of the buffer for type-1 cells and the token pool as follows[3]

$$N_1(n) \triangleq B_1(n) + M - T(n).$$

Then, the process $\{(N_1(n), N_2(n), J(n)), n \geq 0\}$ forms a 3-dimensional Markov chain with finite state space $\{(0,0,1), \cdots, (M-1,0,N), (M,0,1), \cdots, (M,0,N), (M,1,1), \cdots, (M+K_1-1, K_2, N)$. Let X_i be the first passage time to the ith cell arrival regardless of traffic type, that is

$$X_i = \inf\{t | M_1(t) + M_2(t) = i | M_1(0) + M_2(0) = 0\}$$

We also introduce the following joint probabilities:

$p_{m,n}(i,j) =$ the joint probability of m and n arrivals of type-1 and type-2 cells during interval $(0,T]$ and $J(T) = j$, given $J(0) = i$.

$p^k_{m,n}(i,j) =$ the joint probability that there are total k arrivals regardless of traffic type during interval $(0, X_i]$ and m and n arrivals of type-1 and type-2 cells during interval $(X_i, T]$ and $J(T) = j$, given that $J(0) = i$.

Let $p_{m,n}$ and $p^k_{m,n}$ be matrices with its (i,j)-element as $p_{m,n}(i,j)$ and $p^k_{m,n}(i,j)$, respectively. Then, $p_{m,n}$ and $p^k_{m,n}$ are given by

$$\begin{aligned}
p_{m,n} &= p_1(m,T)\frac{(\lambda_2 T)^n}{n!}e^{-\lambda_2 T} \\
p^k_{m,n} &= \int_0^T P\{t < X_i \leq t + dt, M_1(T-t) = m, M_2(T-t) = n\} \\
&= \int_0^T P(k-1,t)\Lambda P_1(m, T-t)\frac{\{\lambda_2(T-t)\}^n}{n!}e^{-\lambda_2(T-t)}dt
\end{aligned}$$

Let's introduce the following matrices:

$$A_k = P(k,T), \qquad A_i' = (p^i_{1,0} + p^i_{0,1}, p^i_{1,1} + p^i_{0,2}, \cdots, p^i_{1,K_2-1} + p^i_{0,\overline{K}_2}, p^i_{1,\overline{K}_2})$$

$$B^k_m = (p^k_{m,0}, p^k_{m,1}, \cdots, p^k_{m,K_2-1}, p^k_{m,\overline{K}_2}) \quad B^k_{\overline{K}_1} = (p^k_{\overline{K}_1,0}, p^k_{\overline{K}_1,1}, \cdots, p^k_{\overline{K}_1,K_2-1}, p^k_{\overline{K}_1,\overline{K}_2})$$

$$\mathbf{C}_0 = \begin{pmatrix} p_{0,0} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad C_1' = \begin{pmatrix} p_{0,1}+p_{1,0} & p_{0,2}+p_{1,1} & \cdots & p_{0,\overline{K_2}}+p_{1,K_2-1} & p_{1,\overline{K_2}} \\ p_{0,0} & p_{0,1}+p_{1,0} & \cdots & p_{0,K_2-1}+p_{1,K_2-1} & p_{1,\overline{K_2-1}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & p_{0,0} & p_{1,\overline{0}} \end{pmatrix}$$

$$C_k = \begin{pmatrix} p_{k,0} & p_{k,1} & \cdots & p_{k,\overline{K_2}} \\ 0 & p_{k,0} & \cdots & p_{k,\overline{K_2-1}} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_{k,\overline{0}} \end{pmatrix} \quad C_{\overline{k}} = \begin{pmatrix} p_{\overline{k},0} & p_{\overline{k},1} & \cdots & p_{\overline{k},\overline{K_2}} \\ 0 & p_{\overline{k},0} & \cdots & p_{\overline{k},\overline{K_2-1}} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_{\overline{k},\overline{0}} \end{pmatrix}$$

Then, the transition probability matrix $\overline{Q}$ of the Markov chain is given by

$$\overline{Q} =$$

$$\begin{pmatrix} A_0+A_1 & A_2 & A_3 & \dots & A_M & A_M' & B_2^M & \dots & B_{K_1-1}^M & B_{\overline{K_1}}^M \\ A_0 & A_1 & A_2 & \dots & A_{M-1} & A_{M-1}' & B_2^{M-1} & \dots & B_{K_1-1}^{M-1} & B_{\overline{K_1}}^{M-1} \\ 0 & A_0 & A_1 & \dots & A_{M-2} & A_{M-2}' & B_2^{M-2} & \dots & B_{K_1-1}^{M-2} & B_{\overline{K_1}}^{M-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & A_1 & A_1' & B_2^1 & \dots & B_{K_1-1}^1 & B_{\overline{K_1}}^1 \\ 0 & 0 & 0 & \dots & \mathbf{C}_0 & C_1' & C_2 & \dots & C_{K_1-1} & C_{\overline{K_1}} \\ 0 & 0 & 0 & \dots & 0 & C_0 & C_1 & \dots & C_{K_1-2} & C_{\overline{K_1-1}} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ & & & & & & & & C_1 & C_{\overline{2}} \\ & & & & & & & & C_0 & C_{\overline{1}} \end{pmatrix},$$

To derive the probability distribution of the system state just after the token generation epoch, let's define the following limiting probabilities

$$\begin{aligned} x_k(j) &\triangleq \lim_{n\to\infty} P\{N_1(n)=k, N_2(n)=0, J(n)=j\}, \\ &\qquad 0 \le k \le M-1, \\ x_{k,l}(j) &\triangleq \lim_{n\to\infty} P\{N_1(n)=k, N_2(n)=l, J(n)=j\}, \\ &\qquad M \le k < M+K_1,\ 0 \le l \le K_2, \\ x_k &\triangleq (x_k(1), x_k(2), \cdots, x_k(N)) \\ x_{k,l} &\triangleq (x_{k,l}(1), x_{k,l}(2), \cdots, x_{k,l}(N)) \\ x &\triangleq (x_0, \cdots, x_{M-1}, x_{M,0}, \cdots, x_{M,K_2}, \cdots, x_{M+K_1-1,K_2}) \end{aligned}$$

Then, the steady-state probability vector x of the Markov chain $\{(N_1(n), N_2(n), J(n)), n \geq 0\}$ is obtained by solving the equations

$$x\,\overline{Q} = x, \qquad x\,e = 1,$$

where e denotes a column vector with all elements equal to one

Next we derive the stationary probability distribution of the system state at an arbitrary time instant. Define the limiting probabilities $y_k(j)$ and $y_{k,l}(j)$ as follows

$$y_k(j) \triangleq \lim_{t\to\infty} P\{N_1(t) = k, N_2(t) = 0, J(t) = j\}, \quad 0 \leq k \leq M-1,$$
$$y_{k,l}(j) \triangleq \lim_{t\to\infty} P\{N_1(t) = k, N_2(t) = l, J(t) = j\}, \quad M \leq k \leq M+K_1,\ 0 \leq l \leq K_2,$$

where $N_1(t) = B_1(t) + M - T(t)$ and $N_2(t)$ is the number of type-2 cells at time t. Consider an arbitrary time τ, and let τ_T be the starting time of the token generation interval which includes the time τ. By considering the system state at the last token generation epoch before the time τ and the number of arrivals during the elapsed token generation interval $(0, \tau - \tau_T]$, we can obtain the probability distribution of the system state at an arbitrary time τ:
For $0 \leq n \leq M-1$,

$$y_n = \sum_{k=0}^{n} x_k U_{n-k},$$

where

$$\begin{aligned} U_m &= P\{\text{total } m\text{-arrivals during the elapsed token generation interval}(0, \tau - \tau_T] \\ &= \frac{1}{T}\int_0^T P(m,t)dt \end{aligned}$$

For $M \leq n_1 \leq M+K_1-1, \quad 0 \leq n_2 \leq K_2-1$,

$$\begin{aligned} y_{n_1,n_2} &= \sum_{k=0}^{M-1} x_k U^{M-k}_{n_1-M,n_2} + \sum_{k=M}^{n_1}\sum_{l=0}^{n_2} x_{k,l} U_{n_1-k,n_2-l}, \\ y_{n_1,K_2} &= \sum_{k=0}^{M-1} x_k U^{M-k}_{n_1-M,\overline{K_2}} + \sum_{k=M}^{n_1}\sum_{l=0}^{K_2-1} x_{k,l} U_{n_1-k,\overline{K_2-l}}, \\ y_{M+K_1,n_2} &= \sum_{k=0}^{M-1} x_k U^{M-k}_{\overline{K_1},n_2} + \sum_{k=M}^{M+K_1-1}\sum_{l=0}^{n_2} x_{k,l} U_{\overline{M+K_1-k},n_2-l} \\ &\qquad 0 \leq n_2 \leq K_2-1, \end{aligned}$$

where

$$U^i_{k,l} = P\{ \text{ total } i\text{-arrivals during interval}(0, X_i] \text{ and then } k\text{-and } l\text{-arrivals of type-1 and type-2 cells during interval } (X_i, \tau - \tau_T]$$
$$= \frac{1}{T}\int_0^T \int_0^u P(i-1,t)\Lambda P_1(k,u-t)\frac{\{\lambda_2(u-t)\}^l}{l!}e^{-\lambda_2(u-t)}dtdu$$

$$U_{k,l} = P\{k\text{-and } l\text{-arrivals of type-1 and type-2 cells during interval } (0,\tau - \tau_T]$$
$$= \frac{1}{T}\int_0^T P(k,t)\frac{(\lambda_2 t)^l}{l!}e^{-\lambda_2 t}dt$$
$$U_{\bar{j},l} = \sum_{m=j}^{\infty} U_{m,l}, \qquad U_{j,\bar{l}} = \sum_{m=l}^{\infty} U_{j,m},$$
$$U^i_{\bar{j},l} = \sum_{m=j}^{\infty} U^i_{m,l}, \qquad U^i_{j,\bar{l}} = \sum_{m=l}^{\infty} U^i_{j,m}.$$

$$y_{M+K_1,K_2} = \pi - \sum_{n=0}^{M-1} y_n - \sum_{n_1=M}^{M+K_1-1}\sum_{n_2=0}^{K_2} y_{n_1,n_2} - \sum_{n_2=0}^{K_2-1} y_{M+K_1,n_2}.$$

Using the above probability distribution, we obtain the following performance measures:

a. The loss probability

Type-1 cells : $P^1_{\text{loss}} = \sum_{n=0}^{K_2} y_{M+K_1,n}\ e,$

Type-2 cells : $P^2_{\text{loss}} = \sum_{n=M}^{M+K_1} y_{n,K_2}\ e.$

b. The mean queue length

Type-1 cells : $M_1 = \sum_{i=0}^{K_1}\sum_{n=0}^{K_2} i y_{M+i,n}\ e,$

Type-2 cells : $M_2 = \sum_{i=0}^{K_2}\sum_{n=0}^{K_1} i y_{M+n,i}\ e.$

c. By Little's law, we obtain the mean waiting time in buffer

$$\text{Type-1 cells} \quad : \quad W_1 = \frac{M_1}{\lambda_1^*(1 - P_{\text{loss}}^1)},$$

$$\text{Type-2 cells} \quad : \quad W_2 = \frac{M_2}{\lambda_2(1 - P_{\text{loss}}^2)}.$$

4. Special cases

In this subsection, we consider special case that arrival of real-time traffic (type-1) follows a Poisson process with rate λ_1. Let $\lambda = \lambda_1 + \lambda_2$. Then, the X_i follows a gamma distribution with parameter (i, λ). In this case, we obtain the following:

$$p_{m,n} = \frac{(\lambda_1 T)^m}{m!} e^{-\lambda_1 T} \frac{(\lambda_2 T)^n}{n!} e^{-\lambda_2 T}$$

$$\begin{aligned} p_{m,n}^k &= \int_0^T P\{t < X_i \le t + dt, M_1(T-t) = m, M_2(T-t) = n\} \\ &= \int_0^T \frac{\lambda(\lambda t)^{i-1}}{(i-1)!} e^{-\lambda t} \frac{\{\lambda_1(T-t)\}^m}{m!} e^{-\lambda_1(T-t)} \frac{\{\lambda_2(T-t)\}^n}{n!} e^{-\lambda_2(T-t)} dt \\ &= \frac{(\lambda T)^{i+m+n}}{(i+m+n)!} e^{-\lambda T} \binom{m+n}{m} (\frac{\lambda_1}{\lambda})^m (\frac{\lambda_2}{\lambda})^n, \qquad 1 \le i \le M. \end{aligned}$$

$$\begin{aligned} U_m &= \frac{1}{T} \int_0^T \frac{(\lambda t)^m}{m!} e^{-\lambda t} dt \\ &= \frac{1}{\lambda T} \{1 - e^{-\lambda T} \sum_{l=0}^{m} \frac{(\lambda T)^l}{l!}\}. \\ U_{k,l} &= \frac{1}{T} \int_0^T \frac{(\lambda_1 t)^k}{k!} e^{-\lambda_1 t} \frac{(\lambda_2 t)^l}{l!} e^{-\lambda_2 t} dt \\ &= \binom{k+l}{k} (\frac{\lambda_1}{\lambda})^k (\frac{\lambda_2}{\lambda})^l \frac{1}{\lambda T} (1 - e^{-\lambda T} \sum_{m=0}^{k+l} \frac{(\lambda T)^m}{m!}), \\ U_{k,l}^i &= \frac{1}{T} \int_0^T \int_0^u \frac{\lambda(\lambda t)^{i-1}}{(i-1)!} e^{-\lambda t} \frac{\{\lambda_1(u-t)\}^k}{k!} e^{-\lambda_1(u-t)} \frac{\{\lambda_2(u-t)\}^l}{l!} e^{-\lambda_2(u-t)} dt du \\ &= \binom{k+l}{l} (\frac{\lambda_1}{\lambda})^k (\frac{\lambda_2}{\lambda})^l \frac{1}{\lambda T} (1 - e^{-\lambda T} \sum_{m=0}^{i+k+l} \frac{(\lambda T)^m}{m!}), \end{aligned}$$

With above values, the system state distribution, the loss probability and the mean waiting time can be substituted.

5. Numerical Examples

In this Section, we give a simple numerical examples. For simplicity, let's assume the arrivals of real-time and nonreal-ime traffic to be Poisson process with rates $\lambda_1(=\lambda_1^*)$ and λ_2, respectively. In all numerical examples, we set the token generation interval (T) equal to 1 and $\lambda_1 = \lambda_2, K_2 = 10, M = 5$.

Figs. 2 and 3 show the loss probability and the mean waiting time of each traffic for various values of the buffer size K_1. From these Figures, we see that the loss probability of nonreal-time traffic is less than that of the real-time traffic, while the waiting time of the real-time traffic is less than that of nonreal-time traffic. Also, we can observe that the loss probability and the waiting time of the real-time traffic are more sensitive to the change of its buffer size K_1 than that of the nonreal-time traffic. These examples follows the reference [6] and the more and more many numerical examples for various performance measures also given in [6].

References

[1] Rathgeb E. P., "Modeling and performance comparison of policing mechanisms for ATM networks", *IEEE J. Select. Areas Commun.*, vol. 9, no. 3, pp. 325-334, April 1991.

[2] Choi B. D., and Choi D. I., Discrete time analysis of the leaky bucket scheme with threshold-based token generation intervals, *IEE Proc. Commun.* vol. 143, no. 2, April, 1996.

[3] Kim Y. H., Shin B. C., and Un C. K., Performance analysis of leaky bucket bandwidth enforcement strategy for bursty traffics in an ATM networks, *Computer Networks and ISDN Systems,* vol. 25, pp. 295-303, 1992.

[4] Yamanaka N., Sato Y., and Sato K., Performance limitation of the leaky bucket algorithm for ATM networks, *IEEE Trans. on Commun.,* vol. 43, no 8, pp. 2298–2300, Aug. 1995.

[5] Choi B. D., and Choi D. I., An analysis of $M, MMPP/G/1$ queues with QLT scheduling policy and Bernoulli schedule, *IEICE Trans. Commun.* vol. E81-B(1), pp. 13-22, 1998.

[6] Choi D. I., Choi B. D. and Sung. D. K., Performance analysis of priority leaky bucket scheme with queue-length-threshold scheduling policy, *IEE P. - Communications* vol. 145(6), pp. 395-401, 1998.

Department of applied Mathematics
Halla University
220-712 San 66, Heungup-li
Heungup-myon, Wonju-shi, Kangwon-do
South Korea