패킷 도착률과 토큰 생성률의 통합 관리를 적용한 대기모형의 분석

정회원 최 두 일*, 김 태 성**°

Analysis of a Queueing Model with Combined Control of Arrival and Token Rates

Doo Il Choi*, Tae-Sung Kim**° ***Regular Members***

요 약

다양한 통신 서비스가 개발되면서, 네트워크 설계자들은 시간상관성, 폭주성, 예측 불가능한 트래픽의 통계적 변동성 때문에 발생하는 혼잡을 제어하기 위한 방안을 모색하여 왔다. 본 논문에서는 패킷 도착률과 토큰 생성률을 통합하여 관리하는 리키버킷 방식을 이용하여 네트워크의 혼잡을 예방하는 모형에 대하여 분석한다. 본 논문에서 다루는 모형에서는 패킷 도착률과 토큰 생성 시간간격을 대기중인 패킷수에 따라 제어함으로써 네트워크 혼잡을 예방하게 된다. 모형의 분석을 위하여 임베디드 마코프체인과 부가변수 기법을 사용하며, 대기중인 패킷수 확률분포, 패킷손실확률, 평균대기시간 등의 특성치를 구한다.

Key Words : Arrival Rate, Leaky Bucket, Queueing Analysis, Telecommunication Networks, Traffic Control

ABSTRACT

As the diverse telecommunication services have been developed, network designers need to prevent congestion which may be caused by properties of timecorrelation and burstiness, and unpredictable statistical fluctuation of traffic streams. This paper considers the leaky bucket scheme with combined control of arrival and token rates, in which the arrival rate and the token generation interval are controlled according to the queue length. By using the embedded Markov chain and the supplementary variable methods, we obtain the queue length distribution as well as the loss probability and the mean waiting time.

Ⅰ. Introduction

Queueing models have been widely studied for traffic control to support various traffic streams and to prevent congestion in telecommunication networks such as Asynchronous Transfer Mode (ATM)[1,2]. ATM networks support diverse traffics with different service characteristics such as voice, data and video. These traffics are statistically multiplexed and transmitted in very high speed. Unpredictable statistical fluctuation of traffic streams may cause congestion. An appropriate traffic control is required to prevent congestion and to gain bandwidth efficiency in ATM networks. Overload control is representative control scheme to prevent congestion. Some researchers have studied the control schemes by regulating service rates[3-5], others have studied the control schemes by

※ This work was supported by the Korea Research Foundation Grant funded by the Korean Government (KRF-2008-521-C00033).
* 한라대학교 응용수학과 (dichoi@halla.ac.kr), ** 충북대학교 경영정보학과/BK21사업단 (kimts@chungbuk.ac.kr), (° : 교신저자)
논문번호 : KICS2009-09-430, 접수일자 : 2009년 9월 29일, 최종논문접수일자 : 2010년 2월 19일

regulating arrival rates[6,7].

In this paper, combining two research streams, we consider the leaky bucket scheme with combined control of arrival and token rates. Almost all previous works have a static token generation interval and arrival rate. In our paper, the arrival rate and token generation interval are jointly controlled according to queue length. In other words, we place thresholds L_1 and $L_2\,(\geq L_1)$ on the buffer. According to whether the queue length exceeds the threshold L_1 or not, the arrival rate is controlled. Furthermore, if the queue length exceeds the threshold L_2, the token generation interval also is controlled.

The cells (packets) arrive according to Poisson processes, and they are stored 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 times T_1 or T_2. Each token allows a single cell to be transmitted, and the token following a transmission is removed from the token pool.

The arrival rate and the token generation interval are changed at only token generation instants. If the queue length $(Q(t))$ at token generation instant exceeds the threshold L_1 (i.e., $Q(t) \geq L_1$), the arrivals follow a Poisson process with rate λ_2. Otherwise, the arrivals follow another Poisson process with rate λ_1. Nevertheless, if the queue length exceeds upper threshold L_2, that is if $Q(t) \geq L_2$, the token generation interval is given by T_2. Otherwise, the token generation interval is $T_1\,(\geq T_2)$. By using the embedded Markov chain and the supplementary variable methods, we obtain the queue length distribution as well as the loss probability and the mean waiting time.

II. Analysis

2.1 System state distribution at token generation instants

Our model is analyzed by the embedded Markov chain method and the supplementary variable method. We first consider the system state distribution at time points just after the token generation instants.

Let $t_n\,(n \geq 1)$ be the n-th token generation epoch with $t_0 = 0$. We also introduce the notations:
B_n = the number of cells in buffer at time t_n+,
T_n = the number of tokens in token pool at time t_n+.

Since the arriving cells wait in buffer only if there is no token, we express the state of buffer and token pool as follows:

$$N_n \equiv B_n + M - T_n.$$

That is, if there are $i\,(0 < i < M)$ tokens in token pool ($B_n = 0$), then $N_n = M - i$. Also, if there are $i\,(0 < i < K-1)$ cells in buffer ($T_n = 0$), then $N_n = M + i$. Finally, the process $\{N_n, n \geq 0\}$ forms a Markov chain with finite state space $\{0, 1, \cdots, M+K-1\}$.

To obtain the system state distribution at token generation epochs, we need to know the number of cell arrivals during the token generation intervals T_1 or T_2. Thus, we introduce the following probabilities:

$$a_n^r = \Pr\{n \text{ cell arrivals by } \lambda_r \text{ during } T_1\} = \frac{(\lambda_r T_1)^n}{n!} e^{-\lambda_r T_1}, \quad r = 1, 2,$$

$$b_n^2 = \Pr\{n \text{ cell arrivals by } \lambda_2 \text{ during } T_2\} = \frac{(\lambda_2 T_2)^n}{n!} e^{-\lambda_2 T_2}.$$

And let

$$\overline{a}_k^r = \sum_{n=k}^{\infty} a_n^r, \quad r = 1, 2, \qquad \overline{b}_k^2 = \sum_{n=k}^{\infty} b_n^2$$

Then, the one-step transition probability matrix P of the Markov chain $\{N_n, n \geq 0\}$ is given by

$$P=\begin{vmatrix}
a_0^1+a_1^1 & a_2^1 & \cdots & a_{M+L_1}^1 & a_{M+L_1+1}^1 & \cdots & a_{M+L_2}^1 & a_{M+L_2+1}^1 & \cdots & a_{M+K-1}^1 & \bar{a}_{M+K}^1 \\
a_0^1 & a_1^1 & \cdots & a_{M+L_1-1}^1 & a_{M+L_1}^1 & \cdots & a_{M+L_2-1}^1 & a_{M+L_2}^1 & \cdots & a_{M+K-2}^1 & \bar{a}_{M+K-1}^1 \\
0 & a_0^1 & \cdots & a_{M+L_1-2}^1 & a_{M+L_1-1}^1 & \cdots & a_{M+L_2-2}^1 & a_{M+L_2-1}^1 & \cdots & a_{M+K-3}^1 & \bar{a}_{M+K-2}^1 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & a_1^1 & a_2^1 & \cdots & a_{L_2-L_1+1}^1 & a_{L_2-L_1+2}^1 & \cdots & a_{K-L_1}^1 & \bar{a}_{K-L_1+1}^1 \\
0 & 0 & \cdots & a_0^2 & a_1^2 & \cdots & a_{L_2-L_1}^2 & a_{L_2-L_1+1}^2 & \cdots & a_{K-L_1-1}^2 & \bar{a}_{K-L_1}^2 \\
0 & 0 & \cdots & 0 & a_0^2 & \cdots & a_{L_2-L_1-1}^2 & a_{L_2-L_1}^2 & \cdots & a_{K-L_1-2}^2 & \bar{a}_{K-L_1-1}^2 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & 0 & \cdots & a_1^2 & a_2^2 & \cdots & a_{K-L_2}^2 & \bar{a}_{K-L_2+1}^2 \\
0 & 0 & \cdots & 0 & 0 & \cdots & b_0^2 & b_1^2 & \cdots & b_{K-L_2-1}^2 & \bar{b}_{K-L_2}^2 \\
0 & 0 & \cdots & 0 & 0 & \cdots & 0 & b_0^2 & \cdots & b_{K-L_2-2}^2 & \bar{b}_{K-L_2-1}^2 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & b_1^2 & \bar{b}_2^2 \\
0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & b_0^2 & \bar{b}_1^2
\end{vmatrix}$$

Define the stationary probability distribution of the Markov chain $\{N_n,\, n\ge 0\}$ as

$$\pi_k=\lim_{n\to\infty}\Pr\{N_n=k\},\ k=0,1,\cdots,M+K-1.$$

Then, the stationary probability distribution $\pi=(\pi_0,\pi_1,\cdots,\pi_{M+K-1})$ for the system state at token generation epochs is given by solving the equations

$$\pi P=\pi,\qquad \pi e=1,\qquad e=(1,1,\cdots,1)^T.$$

2.2 System state distribution at arbitrary instants

In this section we derive the system state distribution at an arbitrary time. Let $N(t)$ indicate the system state at time t, and define the stationary probabilities

$$y_n=\lim_{t\to\infty}\Pr\{N(t)=n\},\qquad 0\le n\le M+K.$$

Then, the y_n is derived by using the supplementary variable method. We use the remaining token generation interval $\hat{T}$ and the elapsed token generation interval $\tilde{T}$ as supplementary variables. We also introduce the notation

$$\xi(t)=\begin{cases}1 & \text{if the token at time } t \text{ is generated by the interval } T_1,\\ 2 & \text{if the token at time } t \text{ is generated by the interval } T_2.\end{cases}$$

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

$$\alpha_{n,r}(x)dx=\lim_{t\to\infty}\Pr\{N(t)=n,\xi(t)=r,x<\hat{T}\le x+dx\},$$
$$r=1,2,$$

and the Laplace transform of $\alpha_{n,r}(x)$

$$\alpha_{n,r}^*(s)=\int_0^\infty e^{-sx}\alpha_{n,r}(x)dx.$$

To know the system state distribution at an arbitrary time, we must know the number of arrivals during the elapsed token generation interval. So, define the joint probability $\beta_r(n,x)dx$ as

$$\beta_r(n,x)dx = \lim_{t\to\infty} \Pr\{n \text{ arrivals by } \lambda_r \text{ during } \tilde{T},\ \xi(t)=r, x<\hat{T}\le x+dx\},\quad r=1,2,$$
$$\beta_3(n,x)dx = \lim_{t\to\infty} \Pr\{n \text{ arrivals by } \lambda_2 \text{ during } \tilde{T},\ \xi(t)=1, x<\hat{T}\le x+dx\},\quad n\ge 0.$$

We also define the Laplace transform $\beta_r^*(n,s)$ of $\beta_r(n,x)$:

$$\beta_r^*(n,s) = \int_0^\infty e^{-sx}\beta_r(n,x)dx.$$

Conditioning the system state at last token generation epoch before time t, $\alpha_{n,r}^*(s)$ satisfies the following equations:

For $0 \le n < M+K$,

$$\alpha_{n,1}^*(s) = \frac{T_1}{E}\left[\sum_{k=0}^{\min\{n,M+L_1-1\}} x_k\beta_1^*(n-k,s) + \sum_{k=M+L_1}^{\min\{n,M+L_2-1\}} x_k\beta_3^*(n-k,s)1_{\{n\ge M+L_1\}}\right],$$
$$\alpha_{n,2}^*(s) = \frac{T_2}{E}\sum_{k=M+L_2}^{n} x_k\beta_2^*(n-k,s)1_{\{n\ge M+L_2\}},$$

where

$$E = \sum_{n=0}^{M+L_2-1} x_k T_1 + \sum_{n=M+L_2}^{M+K-1} x_k T_2$$ is the mean token generation interval. As shown in Appendix, $\beta_r^*(n,s)$ is given as follows:

$$\beta_1^*(n,s) = \frac{1}{T_1}\left[\sum_{k=0}^{n} a_k^1 R_{n-k}^1(s) - e^{-sT_1}R_n^1(s)\right],$$
$$\beta_2^*(n,s) = \frac{1}{T_2}\left[\sum_{k=0}^{n} b_k^2 R_{n-k}^2(s) - e^{-sT_2}R_n^2(s)\right],$$
$$\beta_3^*(n,s) = \frac{1}{T_1}\left[\sum_{k=0}^{n} a_k^2 R_{n-k}^2(s) - e^{-sT_1}R_n^2(s)\right],$$

where

$$R_n^r(s) = (s-\lambda_r)^{-1}\{-\lambda_r(s-\lambda_r)^{-1}\}^n,$$
$$r = 1,2.$$

Finally, substituting $\beta_r^*(n,s)$ $(r=1,2,3)$ into above equations, we obtain the following results:

For $0 \le n < M+K$,

$$\begin{aligned} y_n &= \alpha_{n,1}^*(0) + \alpha_{n,2}^*(0) \\ &= \frac{1}{E}\left[\frac{1}{\lambda_1}\sum_{k=0}^{\min\{n,M+L_1-1\}} x_k\left\{1-\sum_{l=0}^{n-k}a_l^1\right\} + \frac{1}{\lambda_2}\sum_{k=M+L_1}^{\min\{n,M+L_2-1\}} x_k\left\{1-\sum_{l=0}^{n-k}a_l^2\right\}1_{\{n\ge M+L_1\}}\right. \\ &\quad \left. + \frac{1}{\lambda_2}\sum_{k=M+L_2}^{n} x_k\left\{1-\sum_{l=0}^{n-k}b_l^2\right\}1_{\{n\ge M+L_2\}}\right], \end{aligned}$$

and

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

Thus, by using the system state distribution at an arbitrary time, we obtain the following performance measures:

(a) The loss probability for an arbitrary arriving cell: $P_{loss} = y_{M+K}$.

(b) The mean queue length:

$$M = \sum_{i=M}^{M+K}(i-M)y_i$$

(c) By Little's law, we obtain the mean waiting time in the system:

$$W = \frac{M}{(\lambda_1+\lambda_2)(1-P_{loss})}.$$

Ⅲ. Conclusion

In this paper we analyzed a queueing model with combined control of arrival and token rates. The arrival rate of cells (or packets) and the time interval of token generation were jointly controlled to support QoS (Quality of Service) of traffic and to prevent congestion. The results of the paper can be applied for preventive congestion control in telecommunication networks such as ATM and next generation mobile systems.

Appendix

Suppose that the tokens are generated by the interval T_1 and the arrivals occur by the rate λ_1. Let $\hat{T}$ and $\tilde{T}$ be the remaining and the elapsed token generation intervals respectively. Since the token generation interval T_1 is finite and we consider a random point in this interval, we obtain

$$E\left[e^{-s\hat{T}_1}\right] = \int_0^{T_1} e^{-st}\frac{1}{T_1}dt = \left(1-e^{-sT_1}\right)/sT_1$$

From the definition of $\beta_1^*(n,s)$ and above equation, we can derive the following with $R_1(z) = (z-1)\lambda_1$:

$$\begin{aligned}\sum_{n=0}^{\infty}\beta_1^*(n,s)z^n &= E\left[e^{-s\hat{T}_1}e^{R_1(z)\tilde{T}_1}\right] \\ &= e^{R_1(z)T_1}E\left[e^{-(s+R_1(z))\hat{T}_1}\right] \\ &= \frac{1}{T_1}\left[e^{R_1(z)T_1}-e^{-sT_1}\right](s+R_1(z))^{-1} \\ &= \frac{1}{T_1}\left[\sum_{n=0}^{\infty}\sum_{k=0}^{n}a_k^1R_{n-k}^1(s)-\sum_{n=0}^{\infty}e^{-sT_1}R_n^1(s)\right]z^n,\end{aligned}$$

where $R_n^1(s) = (s-\lambda_1)^{-1}\{\lambda_1(\lambda_1-s)^{-1}\}^n$.
By coefficient comparison, we have

$$\beta_1^*(n,s) = \frac{1}{T_1}\left[\sum_{k=0}^{n}a_k^1R_{n-k}^1(s)-e^{-sT_1}R_n^1(s)\right].$$

By similar method, we obtain

$$\beta_2^*(n,s) = \frac{1}{T_2}\left[\sum_{k=0}^{n}b_k^2R_{n-k}^2(s)-e^{-sT_2}R_n^2(s)\right],$$

$$\beta_3^*(n,s) = \frac{1}{T_1}\left[\sum_{k=0}^{n}a_k^2R_{n-k}^2(s)-e^{-sT_1}R_n^2(s)\right],$$

where

$$R_n^2(s) = (s-\lambda_2)^{-1}\left\{\lambda_2(\lambda_2-s)^{-1}\right\}^n$$

References

[1] H. Takagi, *Queueing Analysis: A Foundation of Performance Evaluation, Vol.1*, Amsterdam, North-Holland, 1991.

[2] R. Jain, "Congestion control and traffic management in ATM networks: Recent advances and a survey," *Computer Networks and ISDN System*, Vol.28, pp.1723-1738, 1996.

[3] B. D. Choi and D. I. Choi, "Discrete-time analysis of the leaky bucket scheme with threshold-based token generation intervals," *IEE Proceedings of Communications*, Vol.143, pp.105-111, 1996.

[4] D. I. Choi, C. Knessl, and C. Tier, "A queueing system with queue length dependent service times, with application to cell discarding in ATM networks," *Journal of Applied Mathematics and Stochastic Analysis*, Vol.12, pp.35-62, 1999.

[5] K. Sriram and D. M. Lucantoni, "Traffic smoothing effects of bit dropping in a packet voice multiplexer," *IEEE Transactions on Communications*, Vol.37, pp.703-712, 1989.

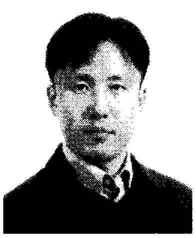
[6] D. I. Choi, T. S. Kim, and S. Lee, "Analysis of an MMPP/G/1/K queue with queue length dependent arrival rates, and its application to preventive congestion control in telecommunication networks," *European Journal of Operational Research*, Vol.187, pp.652-659, 2008.

[7] P. Yegani, "Performance models for ATM switching of mixed continuous-bit-rate and bursty traffic with threshold-based discarding," *Proceedings of IEEE ICC*, pp.1621-1627, 1992.

최 두 일 (Doo Il Choi) 정회원

1996년 2월 KAIST 응용수학 박사
1997년 2월 KAIST 전기 및 전자공학과 위촉연구원
1998년 3월~현재 한라대학교 교수
<관심분야> 큐잉이론의 연구 및 응용, 통신네트워크 등에 있어서의 트래픽제어에 대한 성능분석

김 태 성 (Tae-Sung Kim) 정회원



1991년 2월 한국과학기술원 (KAIST) 경영과학 학사
1993년 2월 한국과학기술원 (KAIST) 경영과학 석사
1997년 2월 한국과학기술원 (KAIST) 산업경영 박사
1997년 2월~2000년 8월 한국전자통신연구원 정보통신기술경영연구소 선임연구원
2005년 1월~2006년 2월 University of North Carolina at Charlotte 방문교수
2000년 9월~현재 충북대학교 경영정보학과 교수, 학과장
<관심분야> 통신 및 보안 분야의 경영 및 정책 의사결정