Performance Analysis of LAN Interworking Unit with PS Schedule in Internet Environment

Chul Geun Park*, Dong Hwan Han** ***Regular Members***

ABSTRACT

We analyze some queueing models to discuss capacity dimensioning of access links of a LAN interworking unit connected to the Internet backbone network. We assume that the IWU has a processor sharing scheduling algorithm to transmit WWW documents to the Internet. In order to analyze the LAN IWU with PS schedule, we use G/M/1/PS, M/M/1/PS and M/G/1/PS queueing models. But we use both the heavy-tailed input and service time models and the Markovian input and service time models in numerical examples for simple performance comparisons. As performance measures, we obtain the mean sojourn times and variances. We present some numerical results to show the effect of heavy-tail characteristics on the mean sojourn time and the access link capacity.

Ⅰ. Introduction

A rapid and enormous growth of the Internet has a great impact on the Internet infrastructures such as a backbone network, an access line and an access network. For an efficient design of such infrastructures including capacity dimensioning of the Internet access network, we need a study on the characteristics of Internet traffic as well as the performance analysis of the IWU (InterWorking Unit) such as routers and servers in a LAN (Local Area Network) as the Internet access network.

In this paper, we focus on the analysis of the LAN IWU such as a proxy server and a cache server with PS(Process Sharing) schedule. To discuss the dimensioning methodology of the Internet access link, we use several queueing models. Queueing analysis requires accurate models of the network and the system under study in order to obtain the useful data. Accordingly, we need the better model of Internet traffic as the input model of queueing system. Fortunately, many studies have been done on characterizing Internet traffic.

M. Nabe et al.[1] have shown that WWW(World Wide Web) traffic is generating a major part of Internet traffic. They have studied the characteristic of WWW traffic using the log data of access links and have found that the document size follows a log-normal distribution. They have also found that the request inter-arrival time follows a log-normal distribution during a whole day and an exponential distribution during the busy hour. Based on these results, they have built an M/G/1/PS queueing model to discuss a dimensioning methodology of the Internet access link. Currently, Internet traffic are not adequately captured by conventional Erlangian stochastic model[2]. Given the fundamental differences in application traffic patterns and usages between Internet and other services such as voice service and cellular mobile radio service, the lack of more suitable methods is giving capacity and performance problems. Several papers[1,2] have indicated that the document sizes of WWW traffic have a heavy-tailed distribution. S. Deng[3] have shown that the inter-arrival time of WWW documents follows a Pareto distribution and the size of documents also follows a Pareto distribution,

* 선문대학교 전자정보통신공학부 통신망연구실(cgpark@sunmoon.ac.kr), ** 선문대학교 수학과
논문번호 : K01062-0206, 접수일자 : 2001년 2월 6일
※ 이 논문은 2000년도 한국학술진흥재단의 지원에 의하여 연구되었음(KRF-2000-003-E00264)
This work was supported by Korea Research Foundation Grant.(KRF-2000-003-E00264)

which is one of heavy-tailed distributions.

Futhermore, as for the arrival pattern of the source traffic, all documents of Internet traffic have to be packetized in IP(Internet Protocol) layer. This generation process of IP packets may have different traffic characteristics compared with generation process of the documents. Recent studies [4,5,6] on Internet traffic characteristics have revealed that the arrival process of IP packets is bursty in nature. Especially, Z. Niu et al.[5] has used a versatile pont process, an MAP(Markovian Arrival Process), to model the bursty nature of IP packet arrivals in an ATM-LAN IWU. Park et al.[6] used an MMPP(Markov Modulated Poisson Process), which is treated as a special case of an MAP to model the bursty nature of IP packet arrivals in ATM environment. In this paper, however, we will use a generation process of the source traffic in the TCP(Transmission Control Protocol) layer in order to analyze the performance of the LAN IWU such as a proxy server and a cache server for capacity dimensioning of Internet access links con- nected to the Internet access backbone. This analysis model is useful for both ATM-LAN and conventional LAN environments

We assume that the IWU under consideration has a PS scheduling algorithm to serve the WWW requests through the backbone network. We also assume that the inter-arrival time of WWW requests follows the heavy-tailed distribution such as log-normal and Pareto in [1,3]. Therefore we use a general probability distribution to model the source traffic in the TCP layer. We scrutinize the effect of the heavy-tail characteristic on the performance of the IWU for dimensioning the access link.

The overall organization of this paper is as follows. In Section II, we describe the system model of the considered IWU with PS algorithm and study the performance analysis of this system by using the G/M/1/PS, M/M/1/PS and M/G/1/PS queueing models for some simple comparisons. In Section III, we give some numerical examples for the perform- ance measures of our queueing models. We show the effect of the heavy-tail characteristic on the performance of the IWU. We finally have a conclusion in Section IV.

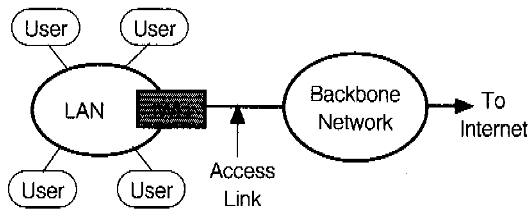


Fig. 1 IWU and access network

Ⅱ. System Model and Queueing Analysis

1. System Model

The access link model of the LAN IWU under consideration are depicted in Fig. 1. The IWU is connected to the Internet backbone network such as an ATM network through the access link. In the access network, each user is connected to the IWU such as a proxy server and a cache server through the LAN. The request documents are returned to user via the access line and the access network.

In this paper, we focus on the performance analysis of the access line and the IWU, which regarded as a queueing system, where the server and customers correspond to the access line and the WWW documents, respectively. This means that we need not consider the backbone network and customer lines (intra-LAN), or that these have enough capacities, which is usual in the current Internet.

Especially, we know that the access line is actually one main cause of the congestion, while it is also an issue how congestion of backbone network and access network are relived. Therefore, for improving the perform- ance of the access network, it becomes important to identify the performance of the IWU with the access lines. For the service discipline, we adopt the PS scheduling algorithm, because multiple WWW documents can share the access line by virtue of underlying TCP. But we neglect the other detail behaviors of TCP protocol for simplicity. We assume that arrivals of the WWW documents have the heavy-tailed distribution. So we use the general probability distributions as an input traffic models or as a service time model. Moreover, we use the exponential probability distributions as inter-arrival time or service time models for comparisons of the queueing models.

2. Queueing Analysis

In this subsection, we consider single server queues G/M/1/PS and M/G/1/PS in a random environment in which the server works according to the processor sharing discipline. We calculate the performance measure such as mean sojourn times and their variances in the queueing systems. The sojourn time of a document is the amount of time from the epoch the document arrives to the epoch of its departure and is the measure of interest here. Let $F(x)$ denote the inter-arrival time distribution with mean $1/\lambda$ and Laplace-Stieltjes transform(LST) $F^*(s)$. Let $G(x)$ denote the service time distribution with $G(0)=0$, mean $1/\mu$ and LST $G^*(s)$. Let the random variables S and W denote the generic service time and sojourn time respectively. Let $W(x)$ be the distribution of W and $W^*(s)$ its LST. Let $\rho=\lambda/\mu$. Assume that $\rho<1$, that is, the considering queues are stable.

For the G/M/1/PS queue, the service time distribution is $G(x)=1-e^{-\mu x}$, for $x>0$. From Theorem 1 of Jagerman and Sengupta[7], for $0<y<1$, we define the sequence $\{f_n(y)\}_{n=0}^{\infty}$ by the recursion $f_0(y)=g(y)$ and

$$f_{n+1}(y)=g(y)+\frac{b(y)}{y}f_n(y)-\frac{1}{y}\int_0^y b'(u)f_n(u)\,du \quad (1)$$

for $n\geq 0$, where the prime denotes the first derivative,

$$b(y)=F^*(s+\mu(1-y)),$$

$$g(y)=\frac{\mu}{y}\int_0^y \frac{1-b(u)}{(1-\mu)(s+\mu(1-u))}\,du.$$

Then $f(y)=\lim_{n\to\infty} f_n(y)$ exists with respect to the recursion (1) and the LST $W^*(s)$ of the sojourn time W is given by

$$W^*(s)=(1-\eta)f(\eta), \quad (2)$$

where η is the smallest positive root of the equation

$$\eta=F^*(\mu(1-\eta)).$$

Note that for $s>0$, alternative represen- tations (integral and power series) for $f(y)$ exist which are also complicated (see Jagerman and Sengupta[7] for details). From Ramaswami[8], we have

$$EW=\frac{1}{\mu(1-\eta)}, \quad (3)$$

$$EW^2=\frac{4}{[2+\mu F^{*\prime}(\mu(1-\eta))]\mu^2(1-\eta)^2}. \quad (4)$$

As a special case of G/M/1/PS, we have $\eta=\rho$ and $F^*(s)=\frac{\lambda}{\lambda+s}$, for M/M/1/PS. From the above equations (3) and (4), we have

$$EW=\frac{1}{\mu(1-\rho)},\quad EW^2=\frac{4}{(1-\rho)^2(2-\rho)\mu^2}.$$

Turning our attention now to the M/G/1/PS queue, we first note that the inter-arrival time distribution is $F(x)=1-e^{-\lambda x}$, for $x>0$. From Theorem 2.5 of Ott[9], we have

$$W^*(s)=\int_0^{\infty}\frac{(1-\rho)dG(x)}{(1-\rho)\phi_1(s,x)+s\phi_2(s,x)} \quad (5)$$

where ϕ_1 and ϕ_2 satisfy the followings, for $\sigma>0$,

$$\int_0^{\infty} e^{-\sigma x}\phi_1(s,x)\,dx=\frac{1-\rho\,\overline{G}^*(\sigma)}{\sigma(1-\frac{s}{\sigma}-\rho\,\overline{G}^*(\sigma))},$$

$$\int_0^{\infty} e^{-\sigma x}\phi_2(s,x)\,dx=\frac{\rho(1-\overline{G}^*(\sigma))}{\sigma^2(1-\frac{s}{\sigma}-\rho\,\overline{G}^*(\sigma))},$$

$$\overline{G}^*(\sigma)=\frac{\mu(1-G^*(\sigma))}{\sigma}.$$

Therefore, we have

$$EW=\frac{1}{\mu(1-\rho)}, \quad (6)$$

$$EW^2=\frac{2}{(1-\rho)^2}\left[\mu_2-\int_0^{\infty}\int_0^{x}(x-y)H(y)\,dy\,dG(x)\right], \quad (7)$$

where μ_2 is the second moment of the service time and

$$\int_0^\infty e^{-sx} dH(x) = \frac{1-\rho}{1-\rho\, \overline{G}^*(s)}.$$

Ⅲ. Numerical Examples

In this section we present some numerical results to show the effect of the mean service time(in fact, the link capacity) and the heavy-tail characteristic on the mean sojourn time of WWW traffic, where the sojourn time of a document is the amount of time from the epoch the document arrives to the epoch of its departure in the system. We give some examples to have a clue for capacity dimensioning of access link of backbone network to the Internet. To investigate the effect of the heavy-tailed characteristic on the mean sojourn time in the G/M/1/PS queueing models, we use at first both Pareto distribution and log-normal distributions as input traffic models[1,3], i.e., inter-arrival time distributions of documents. As service time models (document size distributions), we use also both Pareto and log-normal distributions in the M/G/1/PS queueing models. The Pareto and log-normal distributions are defined respectively as

$$F(x) = 1-\left(\frac{k}{x}\right)^{\alpha},\ x \geq k,$$
$$F(x) = \int_0^x \frac{1}{\sqrt{2\pi}\sigma y} \exp\left[\frac{-(\log y-\zeta)^2}{2\sigma^2}\right] dy,$$

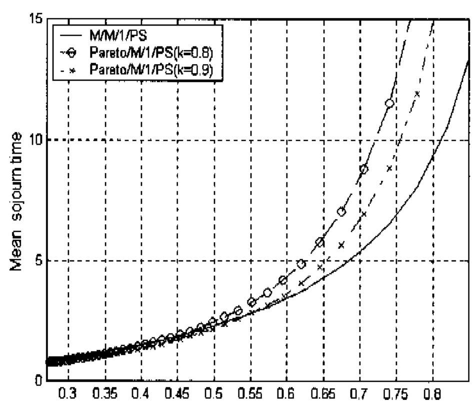


Fig. 2 Mean sojourn time vs. traffic intensity ρ

where if the random variable Y=log$X$ has a normal distribution, then X is said to have a log-normal distribution.

In Figs. 2 and 3, we show the mean sojourn times and variances in the G/M/1/PS queue with heavy-tailed Pareto distributions and the M/M/1/PS queue to investigate the mean delay of documents in the LAN IWU with a PS schedule when the service rate varies from 0.51 to 1.6 by using the equations (3) and (4). We use two Pareto distributions with parameter pairs (k, α)=(0.8, 1.53) and (0.9, 1.639) and an exponential distribution with the corresponding mean λ =0.433 to two parameter pairs in order to compare the [Pareto]/M/1/PS queue and the M/M/1/PS queue. From these figures, we see that the Pareto inter-arrival distribution over-estimates the mean sojourn time and variance compared with the exponential inter-arrival distribution. So this fact leads to require the large volume of access link capacity. We can obtain the useful fact that we have to find out a more accurate distribution from many other distributions especially including the heavy-tailed distributions with many different parameters to estimate the more accurate volume of access link capacity.

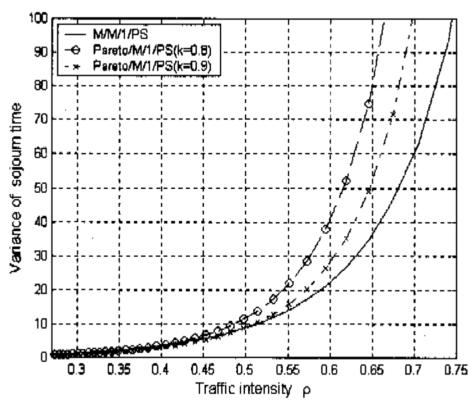


Fig. 3 Variance vs. traffic intensity ρ

Especially from Fig. 3, we can conceive that large documents meet with the larger delays by the definition of variance and the service discipline of PS schedule when the traffic load becomes higher.

In Figs. 4 and 5, we show the mean sojourn times and variances in the [Log- normal]/M/1/PS queue

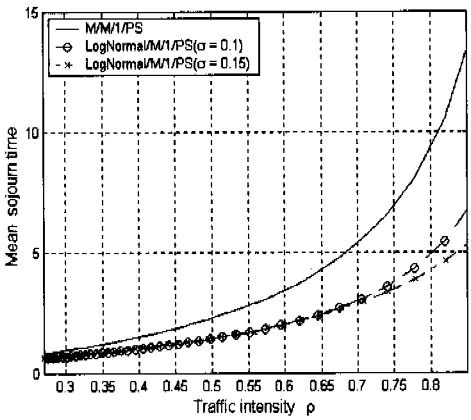


Fig. 4 Mean sojourn time vs. traffic intensity ρ

and M/M/1/PS queue. We use log-normal distributions with the pairs (σ ,ζ)=(0.1, 0.832) and (0.15, 0.826) and an exponential distribution with the corres- ponding mean λ =0.433 to two parameter pairs. From these figures, we see that the log-normal distribution under-estimates the mean sojourn times and variances compared with the exponential distribution, this fact leads to require the small volume of access link capacity.

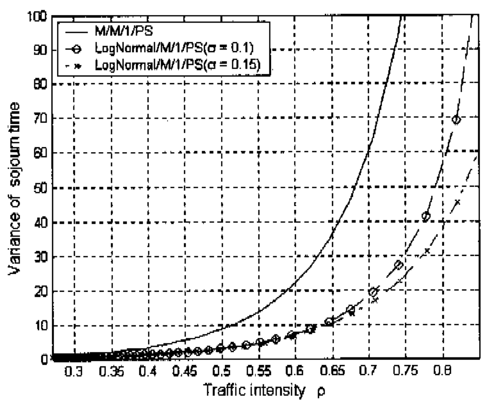


Fig. 5 Variance vs. traffic intensity ρ

In Fig. 6, we show the mean sojourn times(EW) and variances in the M/[Log- normal]/1/PS queue and M/M/1/PS queue by using the equations (6) and (7). We use a log-normal distribution with parameters σ =0.1, ζ =0.3 and an exponential distribution with the corresponding mean λ =0.737 to the parameter pair σ and ζ . In this figure. we see the two mean sojourn times are the same and the log-normal distribution under- estimates their variances compared with the exponential distribution.

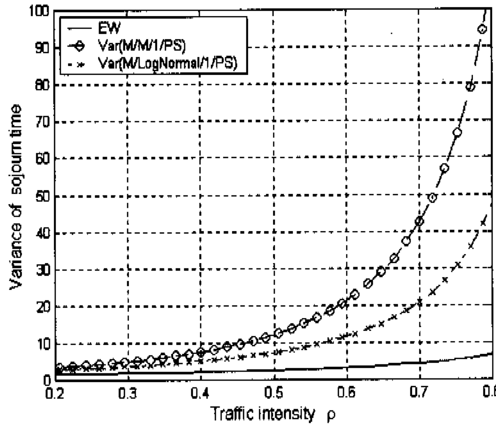


Fig. 6 Mean and variance vs. traffic intensityρ

From the above examples, we can recognize that the Pareto inter-arrival time of documents and the Pareto service time have negative effects on the IWU performance in comparison with the Poisson input and the exponential service time. We can also observe that the Pareto distribution may over-estimate the delay of the documents and this leads to estimate the large volume of access link capacity. On the contrary, we can observe that the log-normal distribution may under-estimate the delay of the documents and this leads to estimate the small volume of access link capacity.

Ⅳ. Conclusions

In this paper, we have built some types of queueing model for a LAN interworking unit to discuss capacity dimensioning of access links. We use 5 queueing models such as [Pareto]/M/1/PS, [Log-normal]/M/1/PS, M/M/1/PS, M/[Pareto]/1/PS and M/[Log- normal]/1/PS to compare the performances of a LAN IWU in different conditions. We have given some numerical results to show the effect of heavy-tail characteristics on the mean delay time of documents in the LAN IWU with the PS schedule.

From the numerical examples, we have recognized that the Pareto inter-arrival time of documents and the Pareto service time have negative effects on the IWU performance in comparison with the Poisson input and the exponential service time. We have also conceived that large documents meet with larger delay by the definition of variance and the service discipline of PS schedule when the traffic load becomes higher. We can observe that the Pareto distribution may over-estimate the delay of the documents and this leads to estimate the large volume of access link capacity. On the contrary, we can observe that the log-normal distribution may under-estimate the delay of the documents and this leads to estimate the small volume of access link capacity. Thus we can obtain the useful fact that we should find out a more accurate distribution from many other distributions especially including the heavy-tailed distributions with many different parameters to estimate the more accurate volume of access link capacity, which is for further study.

Acknowledgement

The authors wish to thank that this work was supported by Korea Research Foundation Grant. (KRF-2000-003-E00264)

References

[1] M. Nabe, M. Murata, and H. Miyahara, "Analysis and modeling of World Wide Web traffic for capacity dimensioning of Internet Access lines", *Performance Evaluation*, Vol. 34, pp. 249-271, 1998

[2] B. A. Mah, "An Empirical Model of HTTP Network Traffic", *IFIP ATM '97*, pp. 593-601, 1997

[3] S. Deng, "Empirical model of WWW Document arrivals at access link", Proc. ICC'96, pp. 1797-1802, 1996

[4] V. Paxson and S. Floyd, "Wide Area Traffic the Failure of Poisson Modeling", *Proc. SIGCOMM'94*, pp. 257-268, 1994

[5] Z. Niu, Y. Takahashi, and N. Endo, "Performance Evaluation of SVC-based IP-Over-ATM Networks", *IEICE*, Vol. E81-B, No. 5, pp. 948-957, 1998

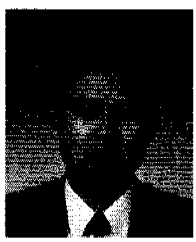
[6] C. G. Park and D. H. Han, "Performance Analysis of LAN Interworking Unit for Capacity Dimensioning of Internet Access Links", *Journal of Korean Institute of Communication Science*, Vol 25, No.4B, pp. 692-699, 2000. 4.

[7] D. L. Jagerman and B. Sengupta, "A functional equation arising in a queue with a gating mechanism", *Probability in the Engineering and Informational Sciences*, Vol 3, pp. 417-433, 1989

[8] V. Ramaswami, "The sojourn time in the GI/M/1 queue with processor sharing, *J. App. Prob.*, Vol 21, pp. 445-450, 1984

[9] T. J. Ott, "The sojourn time distribution in the M/G/1 queue with processor- sharing", *J. App. Prob.*, Vol 21, pp. 360-378, 1984

박 철 근(Chul-geun Park) 정회원



1983년 2월 : 부산대학교 수학과 졸업(이학사)

1986년 2월 : 한국과학기술원 응용수학과 졸업 (이학석사)

1995년 8월 : 한국과학기술원 수학과 졸업(이학박사)

1986년 4월~1997년 2월 : 한국통신 통신망연구소 (선임연구원)

1997년 3월~현재 : 선문대학교 전자정보통신공학부 정보통신공학전공(부교수)

<주관심 분야>트래픽공학, 통신망해석, 큐잉이론

한 동 환(Dong-hwan Han) 정회원

1987년 2월 : 서강대학교 수학과 졸업(이학사)
1989년 2월 : 한국과학기술원 응용수학과 졸업 (이학석사)
1993년 8월 : 한국과학기술원 수학과 졸업 (이학박사)
1994년 3월~현재 : 선문대학교 자연과학부 수학전공(부교수)
<주관심 분야> 확률과정론, 큐잉이론, 통신망해석