

웹서버 운용환경 하에서의 Traffic Model 분석

박경배* 홍창호 이승철
중앙대 전기공학과

A Study on the Traffic Analysis Model for Web Server Operation

Park, Kyung-bae Hong, Chang-ho Lee, Seung-chul
School of EE, Chung-Ang University

Abstract - 네트워크 데이터 traffic 특성은 네트워크의 성능분석과 설계에 중요한 요소이고 적절한 traffic modeling을 통해 load balancing, fault tolerance 및 전송속도 개선 등 시스템 성능을 높이는데 보다 유용한 자료를 제공할 수 있어 지금까지 많은 연구가 진행되어 왔다. 그러나 네트워크 traffic은 네트워크의 규모와 서비스 종류 등 여러 가지 이유로 어느 한가지 모델로 규정짓는 것이 용이하지 않다. 또한 근래 인터넷 사용자의 수가 급속히 증가함에 따라 네트워크 traffic의 대부분이 world wide web에 기인하는 특성을 볼 수 있다. 본 논문에서는 웹서버 운용환경 하에서 적용될 수 있는 traffic model들에 대해 고찰하고 실제 네트워크에서의 패킷 모니터링을 통한 traffic 분석 예를 보인다.

1. 서 론

근래 들어 우리들의 생활 속에서 인터넷의 이용 및 의존도가 높아감에 따라 인터넷의 역할이 점점 더 증가하고 있는 추세에 있다[1]. 이런 상황에서 인터넷 사용자나 인터넷 서비스 제공자(ISP) 모두 보다 안정된 서비스가 요구되고 있으며 네트워크의 bandwidth dimensioning이 문제가 되고 있다[2]. 이의 해결을 위해 네트워크 traffic 모델에 대한 연구가 꾸준히 진행되어 왔다.

네트워크 데이터 traffic의 특성은 네트워크의 성능분석과 설계에 있어 중요한 요소이다. 또한 네트워크 traffic의 분석은 bandwidth dimensioning이나 load balancing, server clustering 등의 연구에 기초자료를 제공하여 보다 안정성 있는 인터넷 서비스를 제공할 수 있다.

기존에 제안된 네트워크 traffic 모델들은 Possion process를 이용한 모델, Packet Train 모델, Self Similar 모델 및 Weibull 분포를 이용한 모델 등이 있으나 네트워크의 목적과 네트워크 서버의 서비스 종류에 따라 데이터 traffic의 특성이 다양하게 존재하므로 네트워크의 데이터 traffic을 어느 한가지 모델로 규정짓는 것은 용이하지 않다.

본 논문에서는 기존에 제안된 네트워크 traffic 모델들에 대해 간략히 고찰하고 웹서버 운용환경 하에서 네트워크 상의 traffic 분석을 위해 client의 서비스 request arrival을 Possion 분포로 가정하고 이를 이용하여 client의 request에 대한 웹서버에서의 서비스 time의 확률적인 분포와 outgoing packet의 waiting time 등에 대해 고찰한다.

2. 본 론

2.1 네트워크 traffic 모델

2.1.1 Possion 모델

데이터 traffic 연구에 많이 사용되는 모델로 통신망

에서 client의 서비스 요청은 드물게 발생한다. 즉, 시간적으로 비교적 긴 구간 뒤에 상대적으로 짧은 서비스 요청이 발생하는 형태의 모델이다. 이 모델은 다음과 같은 세 가지 가정을 한다.

1. client에 의해 발생하는 서비스 요청의 interarrival time은 exponential 분포를 따른다. 즉, 각 서비스 요청은 Possion process를 따른다.
2. client의 서비스 요청에 의한 서비스 time의 길이는 exponential 분포를 따른다.
3. client의 서비스 요청과 그에 대한 서비스 time의 길이의 각 분포는 시간에 관계없이 일정하고 서로 독립적이다.

2.1.2 Packet Train 모델

Packet Train 모델에서는 packet의 그룹이 함께 이동하는 것으로 간주한다. 각각의 packet이 독립적으로 이동하는 모델에서는 모든 packet이 자신의 길을 결정해야 하지만 train 모델에서는 같은 목적지를 갖는 packet의 그룹은 첫 번째 packet이 길을 정하면 나머지는 그 뒤를 따르게 되어 routing 결정에 의한 overhead를 줄일 수 있다.

train 모델은 모든 packet이 train 상태나 intertrain 상태에 있고 두 상태간에는 서로 상관관계가 없다고 가정한 two-state Markov 모델을 이용하여 모델링을 한다. 여기서 각 상태의 유지기간은 exponential 분포를 따르고 일반적으로 intertrain arrival time은 수 초 정도의 단위가 되고 train내 각 packet의 interarrival time은 milisecond 단위가 된다.

2.1.3 Self-Similar 모델

네트워크에서의 packet traffic은 강한 자기상관(auto-correlation)을 갖고 있으며 긴 구간에 걸쳐 상호 의존성이 존재한다고 가정하는 모델이다. 즉, packet들의 상관 관계는 packet 사이에 큰 지연(lag)이 존재해도 쉽게 소멸되지 않는 것으로 가정하는 모델로서 long-memory 모델이라고도 하며 cell 기반의 ATM 네트워크 traffic 연구 등에 이용되었다.

2.2 Possion process의 적용

2.1절에서 기술한 네트워크 traffic 모델 외에도 log normal 분포나 β parameter의 값에 따라 여러 가지 분포를 근사화할 수 있는 Weibull 분포 등을 이용한 traffic 분석이 있다.

본 논문은 웹서버에 대한 client의 서비스 request exponential 분포로 보고 서버의 서비스 시간의 분포와 queueing 시스템에서 각 packet의 서비스 waiting 시간의 분석을 위해 Possion process를 적용하였다.

Possion process는 renewal간의 시간이

exponential 분포를 갖는 renewal counting process이다. Possion process에서 구간 $(0, t]$ 에 발생하는 event의 수 $N(t)$ 는 파라미터 λt 를 갖는 Possion 분포를 갖는다. 즉, Possion 분포는

$$P_k(t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}$$

가 되고, 여기서,

λ : arrival rate

t : time interval

k : event의 수

이다. 그러므로 t 시간 내에 한 번 이상의 event가 발생할 확률은 $F(t) = 1 - e^{-\lambda t}$ 가 되고

$$f(t) = \frac{dF(t)}{dt} = \lambda e^{-\lambda t}$$

가 되어 지수분포가 된다. 그러므로 interarrival time 기대값은

$$E(t) = \int_0^{\infty} t f(t) dt = \int_0^{\infty} t \lambda e^{-\lambda t} dt = \frac{1}{\lambda}$$

가 된다.

웹서버를 concurrent 서버로 가정하고, 서비스에서 각 packet을 생성하는데 걸리는 시간을 서비스 time으로 간주하여 서비스 time을 일정하다고 할 경우 queuing 시스템을 M/D/1 시스템으로 모델링 할 수 있다. 따라서 queue에서 대기하고 있는 요청 수의 기대치는

$$E[N] = \rho + \frac{\rho^2}{2(1-\rho)}$$

이 되고[5], 여기서 ρ 는 서버 utilization(이용률)으로서

$$\rho = \lambda E[B] = \frac{\lambda}{\mu}$$

이다. $E[B]$ 는 서비스 time의 기대값이고 μ 는 서비스 rate를 나타낸다.

이때 queue에서의 waiting time W는

$$W = \frac{\rho}{2(1-\rho)} \times E[B]$$

가 된다.

2.3 TCP(Transmission Control Protocol)

Possion process를 이용하여 웹서버의 traffic을 분석하기 위해서는 client에서 웹서버로의 서비스 요청의 interarrival time과 서버의 서비스 time이 요구된다. 이 중 서비스 요청의 interarrival time은 client에서 서버로 보내는 packet의 header에 포함된 SYN flag의 bit이 1인 packet 간의 간격을 살펴봄으로써 얻을 수 있다.

2.3.1 TCP의 접속 성립과 종료

그림 1은 시간축 상에서 TCP의 접속 성립과 종료 과정을 나타낸다. TCP는 접속 지향적인 프로토콜이므로

client는 서버로부터 서비스를 받기 위해 우선 접속 요청인 SYN 신호를 서버에 보낸다. 이 신호를 받은 서버는 client에 SYN 신호와 ACK신호를 동시에 보낸다. 서버로부터 응답을 받은 client는 서버에 ACK 신호를 보내고 접속이 이뤄지게 된다. 서비스가 이뤄진 후 접속 종료 시 client는 서버에 FIN 신호를 보내고 서버는 응답으로 FIN과 ACK를 보낸다. 이에 대해 client는 ACK를 서버에 보내고 접속을 종료하게 된다.

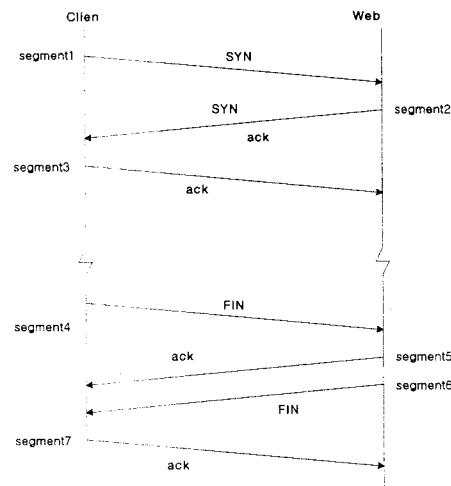


그림 1 시간축 상의 TCP 접속 성립과 종료

2.3.2 TCP datagram의 구조

그림 1에서 서버와 client로 전송되는 각 segment TCP header는 총 20byte로 되어 있고 SYN, ACK, 등의 flag이 있어 segment의 종류를 알 수 있다. 그럼에 TCP header의 구조를 보인다.

1	15 16	32
16bit source port number		16bit destination port number
32bit sequence number		
4bit header length	U A P R S F R G K H T N X	16bit window size
16bit TCP checksum		16bit urgent point
options(if any)		
data(if any)		

그림 2 TCP의 구조

상기 datagram에서의 주요 데이터들은 tcpdump를 이용하여 관측할 수 있다.

2.4 웹서버에서의 traffic 분석 예

본 논문에서는 traffic data의 수집을 위해 교내의 웹서버와 같은 Ethernet에 리눅스 서버를 설치하여 traffic 데이터를 수집하였다. freeware인 tcpdump를 사용하여 한 시간동안 약 150만여 개의 packet을 수신하였고 그 중 웹서버에 의한 traffic만을 분석하였다.

2.4.1 traffic 데이터의 수집

tcpdump는 네트워크 interface에 수신되는 packet

의 header를 보여준다. 그림 3은 "bash#tcpdump -i eth0 -ntt 'port www'" 명령을 이용해 수집한 데이터의 일부분이다. 그림에서 각 행에는 packet의 도착시간, 출발지의 ip와 포트, 목적지의 ip와 포트, TCP header의 flag, sequence number, acknowledge number, window size 등이 표시되어 있다.

```
972985755.104206 165.194.1.55. www > 165.194.82.111.2740: P 73426:74886(1460) ack 10890 win 8760 (DF)
972985755.105358 165.194.1.55. www > 165.194.82.111.2740: P 74886:76249(1363) ack 10890 win 8760 (DF)
972985755.106244 165.194.1.55. www > 165.194.82.111.2740: P 76249:77280(1031) ack 10890 win 8760 (DF)
972985755.107480 165.194.1.55. www > 165.194.82.111.2740: . 77280:78740(1460) ack 10890 win 8760 (DF)
972985755.108712 165.194.1.55. www > 165.194.82.111.2740: P 78740:80200(1460) ack 10890 win 8760 (DF)
```

그림 3 tcpdump를 이용한 packet의 수집 예

2.4.2 traffic 데이터의 분석

실제 웹서버의 traffic은 client의 서비스 요청에 의해 발생되므로 tcpdump에 의해 수집된 packet 중에서 서버에 들어오는 segment의 SYN flag이 1로 set된 packet들간의 interarrival time의 분포를 client의 서비스 요청 분포로 보고 이를 정리하여 그림 4에 그 분포를 보였다.

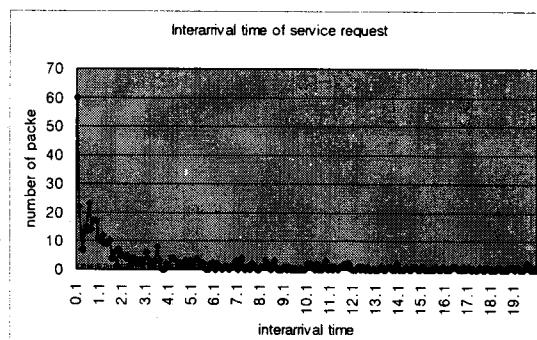


그림 4 웹서버에서의 서비스 요청 분포

그림 4에서 보는 바와 같이 client의 서비스 요청의 interarrival time은 exponential 분포로 나타났다. curve fitting 결과 서비스 요청의 arrival rate $\lambda = 0.3023$ 이고 client의 서비스 요청은 $1/\lambda$ 로 평균 3.3077초에 한번씩 도착함을 알 수 있다.

또, 서버의 서비스 time은 서버의 성능에 따라 그 차이가 있으므로 각 서버마다 traffic 분석의 결과는 달라지게 된다. 본 논문에서 사용된 서버에서는 그림 3에서 보면 연속된 데이터가 전송되는 시간간격이 packet의 생성 시간과 거의 같다고 볼 때 서버의 서비스 time $E[B]$ 는 약 1ms로 볼 수 있다. 따라서 서버 이용률 ρ 가

$$\rho = \lambda E[B] = 0.3023 \times 10^{-3}$$

이 되어 queue에서의 기대되는 대기 요청 수는

$$E[N] = \rho + \frac{\rho^2}{2(1-\rho)} = 0.3023 \times 10^{-3}$$

이 된다. 그러므로 client의 서비스 요청이 queue에서

기하는 waiting time W 는

$$W = \frac{\rho}{2(1-\rho)} \times E[B] = 0.15 \mu\text{s}$$

가 되는 것으로 나타났다.

3. 결 론

본 논문에서는 기존의 traffic 데이터 분석을 위한 모델들에 대해 고찰하고 Possion process를 적용하여 웹서버 운영환경 하에서의 데이터 traffic 분석 예를 보였다. 또, 웹서버에서의 서비스를 위한 packet 생성 시간을 서비스 time으로 보고 queuing 시스템을 M/D/1 시스템으로 모델링하여 client의 서비스 요청이 queue에서 대기하는 waiting time을 계산하였다. 향후 보다 정확한 traffic 분석을 위하여 다양한 서버 운영환경에서 장시간에 걸친 데이터의 수집을 행할 예정이다.

본 연구는 네트인텔리전스(주)의 지원에 의하여 수행되었음

(참 고 문 현)

- [1] G. Babic, B. Vandalore, and R. Jain, "Analysis and Modeling of Traffic in Modern Data Communication Networks", Ohio State Univ. tech. rep., Feb. 1998, <http://www.cis.ohio-state.edu/~jain/papers/tr-fmodel.htm>
- [2] Maurizio Molina, Paolo Castelli, and Gianluca Foddis, CSELT, "Web Traffic Modeling Exploiting TCP Connections' Temporal Clustering through HTML-REDUCE", IEEE Network, May/June, p46-55, 2000.
- [3] K.S.Trivedi, "Probability and Statistics with Reliability, Queuing, and Computer Science Applications", Prentice-Hall, 1982.
- [4] W. Richard Stevens, "TCP/IP Illustrated : Volume I, II", Addison-Wesley Publishing Company, 1994.
- [5] L. Kleinrock, "Queuing System Volume II: Computer Applications", New York: Wiley, 1976
- [6] "tcpdump, The Protocol Packet Capture and Dumper Program," <http://www.tcpdump.org/tcpdump-3.5.2.tar.gz>.