

논문 2004-41TC-5-5

KT의 Ntopia 가입자 망 트래픽 분석 및 모델링

(Analysis and Modeling of Traffic at Ntopia Subscriber Network of Korea Telecom)

주 성 돈*, 이 채 우**

(Sung-Don Joo and Chae-Woo Lee)

요 약

인터넷이 발전하면서 이전에 존재하던 응용프로그램과는 다른 새로운 응용프로그램들이 등장하고 있으며, 이에 따라 트래픽의 특징 또한 변하고 있다. 그 결과 새로운 응용프로그램이 네트워크 성능에 미치는 영향도 변하고 있다. 현재 널리 사용되고 있는 P2P(Peer to Peer) 응용프로그램이 발생시키는 트래픽은 기존의 웹(HTTP)이나 FTP(File Transfer Protocol) 응용프로그램의 트래픽과 다른 통계적 특징을 보여주고 있다. 본 논문은 P2P 트래픽의 특징을 분석하기 위해 KT(Korea Telecom)의 가입자망 트래픽을 측정하였다. 측정된 트래픽을 플로우별로 분석을 하고, P2P 트래픽의 자기 유사성을 측정하고 웹 트래픽과 비교하였다. 분석 결과 P2P 트래픽은 기존의 웹 트래픽과 비교해 매우 버스트(burst)하였으며, P2P 응용프로그램들이 업스트림 트래픽과 다운스트림 트래픽을 대칭적으로 만들고 있었다. 네트워크에서 발생하는 패킷 손실과 지연 등 QoS 관련 파라미터를 예측하기 위해, 본 논문에서는 잘 알려진 자기 유사 트래픽 모델을 이용하여 P2P 트래픽을 모델링하고 트래픽 모델의 정확성을 검증하기 위해 SSQ(Single Server Queue)를 이용하여 손실확률과 평균 지연시간을 비교하였다. 시뮬레이션 결과 자기 유사 트래픽 모델은 P2P 트래픽의 성능을 잘 예측할 수 있었으며, 네트워크를 설계하거나 성능을 측정할 때 입력 트래픽으로서 이 트래픽 모델들을 사용할 수 있다.

Abstract

As Internet technologies are mature, many new applications that are different characteristics are emerging. Recently we see wide use of P2P(Peer to Peer) applications of which traffic shows different statistical characteristics compared with traditional application such as web(HTTP) and FTP(File Transfer Protocol). In this paper, we measured subscriber network of KT(Korea Telecom) to analyze P2P traffic characteristics. We show flow characteristics of measured traffic. We also estimate Hurst parameter of P2P traffic and compare self-similarity with web traffic. Analysis results indicate that P2P traffic is much burstier than web traffic and makes both upstream traffic and downstream traffic be symmetric. To predict parameters related QoS such as packet loss and delays we model P2P traffic using two self-similar traffic models and predict both loss probability and mean delay then compare their accuracies. With simulation we show that the self-similar traffic models we derive predict the performance of P2P traffic accurately and thus when we design a network or evaluate its performance, we can use the P2P traffic model as reference input traffic.

Keywords : P2P, Traffic Model, Self-Similarity, Heavy-tailed ON/OFF, M/Pareto

I. 서 론

인터넷의 발달과 함께, 새로운 응용프로그램이 끊임

* 학생회원, 아주대학교 전자공학과
(School of Electronics Engineering, Ajou University)

** 정회원, 아주대학교 전자공학과
(School of Electronics Engineering, Ajou University)

※ 본 논문은 한국과학재단에서 지원하는 연구 지원사업(R01-2003-000-10724-0)의 연구 결과입니다.

접수일자: 2004년2월2일, 수정완료일: 2004년4월30일

없이 나타나고 있다. 특히 메신저, 당나귀, 소리바다, Napster 등 P2P 응용프로그램이 현재 널리 사용되고 있다^[1]. 이러한 응용프로그램들은 콘텐츠 제공 방식을 바꾸고 있다. P2P 응용프로그램이 나오기 이전의 네트워크 트래픽은 비대칭이었다. 즉, P2P 응용프로그램이 등장하기 이전은 정보의 대부분이 IDC(Internet Data Center)에 위치하고 있어, 업스트림 트래픽보다 다운스트림 트래픽이 더 많았다^[2]. 그러나 P2P 응용프로그램의 경우, 각 사용자는 서버와 클라이언트로 동작할 수

있어 사용이 늘어날수록 ISP(Internet Service Provider)의 트래픽 분포가 더욱 대칭적으로 나타난다. 일반적인 ISP 네트워크는 '일반 사용자들은 IDC에 위치한 서버의 정보를 다운로드하고, 사용자들끼리의 트래픽은 매우 작다'는 가정을 바탕으로 설계되었다. 그러나 P2P 사용이 늘어나면, 이전의 네트워크 설계 가정으로는 P2P 사용자의 QoS를 만족시킬 수 없다. 따라서 인터넷 접속기술과 네트워크 구조는 이러한 트래픽의 변화를 수용할 수 있도록 바뀌어야만 한다.

인터넷에서 QoS는 매우 중요한 문제이다. 효과적인 QoS 지원을 위해 네트워크 관리자는 QoS를 요구하는 트래픽 특성을 이해해야 한다. 본 논문에서는 최근 트래픽의 특징과 네트워크의 특징을 조사하기 위해 KT의 초고속 가입자망의 트래픽을 측정하고, 트래픽을 업스트림과 다운스트림으로 분류하여 플로우를 분석하였다. 그리고 P2P 트래픽의 버스트(burst)를 알아보기 위해 자기 유사성을 측정해보고 이 결과를 웹 트래픽과 비교하였다. 분석 결과 P2P 응용프로그램은 기존의 웹(HTTP) 트래픽보다 더욱 버스트함을 알 수 있었다. 트래픽 모델은 네트워크에서 지연과 손실을 측정할 수 있기 때문에 응용프로그램의 QoS와 네트워크 성능을 평가하는 중요한 요소이다. 본 논문은 M/Pareto 모델과 Heavy-tailed ON/OFF 모델을 이용하여 P2P 트래픽을 모델링하고, SSQ(Single Server Queue)를 이용하여 트래픽 모델의 정확성을 평가하였다.

본 논문은 구성은 다음과 같다. II장에서는 데이터 트래픽의 자기 유사성을 알아보고, III장에서는 자기 유사 트래픽 모델인 Heavy-tailed ON/OFF 모델과 M/Pareto 모델을 소개하였으며, SSQ를 이용하여 모델의 정확성을 평가하는 방법을 나타내었다. IV장은 트래픽이 측정된 위치와 방법을 소개하였으며, V장에서는 플로우의 분석을 통해 트래픽의 중요한 통계적 결과를 나타내었다. VI장에서는 P2P 트래픽의 자기 유사성을 측정하고, III장에서 소개한 모델을 이용하여 P2P 트래픽을 모델링하였다. 그리고 SSQ를 이용하여 손실확률과 평균 지연시간을 측정함으로써 모델의 정확성을 검증하였다. 마지막 VII장은 본 논문의 결과를 정리하고 결론을 내렸다.

II. 트래픽의 자기 유사성

트래픽의 자기 유사성은 모든 시간의 범위에서 트래픽의 버스트가 사라지지 않는 현상을 말한다^[1-3]. 이 특

성은 1993년 벨코어 연구원들이 이더넷 트래픽에서 발견하였다[2]. 또한 그들은 트래픽 스트림이 다중화(Multiplexing)되더라도 버스트한 특징이 사라지지 않고 그대로 남아 있다는 사실을 보여주었다. 그 이후 트래픽의 자기 유사한 현상은 이더넷 환경뿐만 아니라 ATM, 압축된 디지털 비디오 트래픽, 웹 트래픽, WAN 트래픽 등에서도 관찰이 되었다^{[3][4][6][7]}. 이 같은 현상은 수학적으로 통계적 개념을 이용하여 정의할 수 있다. 정상상태 시계열 X 에 대해 m -aggregated 시계열 $X^{(m)} = \{X_{mk}, k=0, 1, 2, \dots\}$ 은 인접한 m 크기의 블록을 겹치지 않게 원래의 시계열을 합함으로써 정의한다^[3-4].

$$X^{(m)} = \frac{1}{m} (X_{mt-m+1} + X_{mt-m+2} + \dots + X_{tm}) \quad (1)$$

시계열 X 가 모든 $m = 0, 1, 2, 3, \dots$ 에 대해 다음 조건을 만족하면 정확히 자기 유사하다(exactly self-similar).

$$\text{Var}[X^{(m)}] = \frac{\text{Var}[X]}{m^\beta} \quad \text{Variance} \quad (2)$$

$$R_{X^{(m)}}(k) = R_X(k) \quad \text{Autocorrelation}$$

마르코프 가정을 바탕으로 하는 모델은 $\beta = 1$ 이며, 시간 평균의 분산은 $1/m$ 비율로 감소한다. 그러나 자기 유사한 프로세스에서 분산은 $1/m$ 보다 더 느린 비율 ($1/m^\beta, 0 < \beta < 1$)로 감소한다. 그리고 시계열 X 가 충분히 큰 k 에 대해 다음과 같다면, 근사적으로 자기 유사(asymptotically self-similar)하다.

$$R_{X^{(m)}}(k) \rightarrow R_X(k) \text{ as } m \rightarrow \infty \quad (3)$$

트래픽의 자기 유사성을 평가하기 위해 Hurst 파라미터를 이용한다. Hurst 파라미터는 0.5~1 사이의 값을 가지며, 그 값이 1에 가까워질수록 자기 유사성이 강해진다. Hurst 파라미터 값을 얻는 방법은 분산-시간 그래프(variance-time plot), R/S plot, Periodogram, IDC plot, Whittle estimate 등이 있다^{[4][6]}. 본 논문에서는 P2P 트래픽의 자기 유사성을 측정하기 위해 육안으로 쉽게 자기 유사성을 확인할 수 있는 분산-시간 그래프를 이용하였다^[6]. 분산-시간 그래프는 m -aggregated된 프로세스 $X^{(m)}$ 의 분산을 m 에 관하여 로그함수로 나타내면 그래프의 기울기가 -1 보다 작은 값을 가진다. 식 (4)에서 나타낸 것처럼 기울기(β)와 Hurst 파라미터는

$\beta = 2 - 2H$ 의 관계를 가진다.

$$\begin{aligned} \text{Var} [X^{(m)}] &\sim \frac{\text{Var} [X]}{m^\beta} \\ \log (\text{Var} [X^{(m)}]) &\sim \log (\text{Var} [X]) \\ &\quad - \beta \log (m), H = 1 - \beta/2 \end{aligned} \quad (4)$$

트래픽의 자기 유사성은 긴 지연과 많은 패킷의 손실을 발생시켜 QoS에 나쁜 영향을 준다. 현재 네트워크의 상당부분을 P2P 트래픽이 차지하고 있기 때문에 P2P 트래픽의 버스트를 조사하고, 네트워크 성능과 QoS에 미치는 영향을 조사하는 것은 매우 중요하다.

III. 트래픽 모델

트래픽의 자기 유사성이 발표된 이후 이 트래픽을 묘사하기 위해 많은 트래픽 모델들이 제안되었다. 본 논문에서는 비교적 쉽게 자기 유사 트래픽을 발생시키는 Heavy-tailed ON/OFF 모델과 M/Pareto 모델을 이용하였다. 다음은 두 트래픽 모델을 소개한다.

1. Heavy-tailed ON/OFF 모델

일반적인 ON/OFF 모델에서 ON/OFF 상태의 지속시간을 Heavy-tailed 분포를 갖게 하여 이를 중첩시키거나 다중화하면 자기 유사 트래픽을 만들어낼 수 있다^[7]. 본 논문에서는 ON/OFF 모델의 ON/OFF 상태가 Heavy-tailed 분포를 갖도록 하기 위해 카오틱 맵(chaotic map)을 이용하였다. 그림 1은 카오틱 맵을 이용하여 일반적인 ON/OFF 모델을 묘사하는 방법을 나타내고 있다. 카오틱 맵의 상태 변수 x_n 은 식(5)에 나타난 것처럼 시간에 따라 비선형 맵 $f_1(x), f_2(x)$ 를 따라 그림 1의 (b)처럼 변하게 된다.

$$x_{n+1} = \begin{cases} f_1(x_n), & 0 \leq x_n \leq d \\ f_2(x_n), & d \leq x_n \leq 1 \end{cases} \quad (5)$$

이처럼 $f_1(x), f_2(x)$ 를 따라 움직인 결과는 ON/OFF 상태의 지속시간을 나타내게 된다.

카오틱 맵 $f_1(x), f_2(x)$ 는 초기 조건이 결정되면 나머지 모든 결과가 결정되기 때문에 확률적이지 못하다. 그러나 $f_1(x), f_2(x)$ 를 따라 움직인 결과로 얻은 ON/OFF 상태 지속시간의 값은 확률적인 결과로 얻은 지속시간의 값과 유사한 결과를 얻을 수 있다^[7]. 카오틱 맵을 이용한 Heavy-tailed ON/OFF 모델의 자세한 내용은 [7]에서 확인할 수 있다.

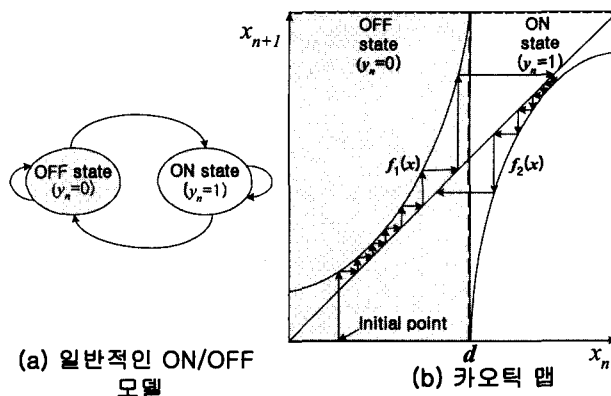


그림 1. ON/OFF 모델과 카오틱 맵의 비교
Fig. 1. Comparison of ON/OFF model and Chaotic map

본 논문에서는 $f(x) = \epsilon + x + cx^m$ 인 비선형 맵을 이용하여 Heavy-tailed ON/OFF 모델을 만들었으며, ON/OFF 상태를 결정짓는 d 의 값과 자기유사성을 표현하는 m 값이 필요하게 된다. m 의 값은 측정된 트래픽의 Hurst 값으로부터 구할 수 있으며, 본 논문에서는 맵의 간략화를 위해 ϵ 의 값을 0으로 설정하고, d 의 값은 0.5로 하여 기본적인 Heavy-tailed ON/OFF 모델이 되도록 하였다. 이렇게 얻은 Heavy-tailed ON/OFF 모델을 50개 이상을 중첩하여 자기 유사 트래픽을 발생시켰다.

2. M/Pareto 모델

M/Pareto 모델은 M/G/ ∞ 프로세스의 특별한 경우이며, Heavy-tailed ON/OFF 모델^[7]의 특별한 경우라 할 수 있다^[8]. M/Pareto 모델은 트래픽이 발생하는 구간은 heavy-tailed 분포를 가지며, 트래픽을 발생시키는 각 세션의 간격은 지수분포를 갖는다.

버스트 세션의 도착은 도착률이 λ 인 Poisson 프로세스로 표현한다. 그러므로 인접 버스트 세션 사이의 도착 시간은 평균이 $1/\lambda$ 인 지수 분포를 갖는다. M/Pareto 모델에서 각 버스트 세션의 지속시간(d)은 상호 독립적이며, Pareto 확률 변수를 가진다. Pareto 분포는 LRD(Long Range Dependence)특징을 갖고 있어 M/Pareto 모델에 자기 유사성을 갖게 한다. M/Pareto 모델을 이용하여 트래픽을 발생시키기 위해서 다음 네 가지 파라미터를 요구한다.

- Poisson arrival rate: λ
- Arrival rate of work within a burst: r
- Starting point of Pareto tail: δ
- Decreasing factor of Pareto: γ

M/Pareto 모델의 파라미터는 측정된 통계적 결과인 평균, 분산, Hurst 파라미터를 이용하여 얻을 수 있다. 그러므로 세 가지 값으로 네 가지 값을 얻어야 하므로 자기 유사성은 변화시키지 않으면서 최적의 λ, r 값을 찾아야 한다. 이 모델의 자세한 설명은 [8]에서 찾아 볼 수 있다.

3. SSQ(Single Server Queue)

트래픽 모델은 단순히 측정된 통계적 결과를 이용하여 트래픽을 모델링 하므로 모든 통계적 특성을 정확하게 나타낼 수 없다. 그러므로 모델의 정확성을 평가할 기준이 필요하다. 네트워크는 하나 이상의 SSQ로 묘사할 수 있으므로 트래픽 모델의 정확성은 SSQ에서의 성능을 비교함으로써 평가될 수 있다^[9]. 본 논문은 FIFO (First In First Out) 이산 시간 큐 모델을 사용하였다. A_n 은 n 번째 표본화 구간동안 시스템에 들어오는 일의 양, C 는 서버의 서비스율을 표시한다. 시스템의 버퍼가 무한 버퍼일 경우 n 번째 구간의 남은 일의 양을 Q_n 이라고 표시하면 Q_n 의 프로세스는 다음과 같다.

$$Q_n = \max(0, Q_{n-1} + A_n - C) \quad (6)$$

위의 프로세스를 이용하여 무한 버퍼를 사용할 경우, 시스템의 버퍼의 길이를 구할 수 있고, 유한 버퍼를 사용할 경우 시스템의 손실확률과 평균 지연 시간을 구할 수 있다. 그림 2는 실제 트래픽과 트래픽 모델을 이용하여 얻은 트래픽을 SSQ에 넣어 모델의 정확성을 평가하는 과정을 나타내고 있다.

본 논문은 서버 이용률 변화에 따라 SSQ에서 발생하는 손실확률과 평균 지연시간을 이용하여 트래픽 모델의 정확성을 평가할 것이다.

IV. 트래픽 측정

최종 가입자 단의 트래픽 특성 파악을 위해 본 논문에서는 KT와 함께 Ntopia 가입자 수 400명이 넘는 아파트 단지의 가입자 트래픽을 2003년 9월 15일부터 16일까지 24시간동안 측정하였다. 그림 3은 네트워크 구성과 측정위치를 표시하고 있다. 본 논문에서 이용한 트래픽은 Tcpcdump^[10]를 이용하여 가입자 백본 스위치에서 미러링 한 것이다. 그리고 덤프의 양을 줄이기 위해 이더넷 프레임당 68Bytes로 제한하였다. 각 프레임의 첫 68Bytes는 기본적인 트래픽을 분석하기 위한 IP 헤

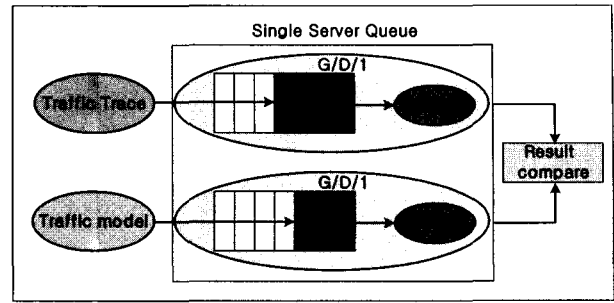


그림 2. SSQ 결과 비교
Fig. 2. Comparison of SSQ result

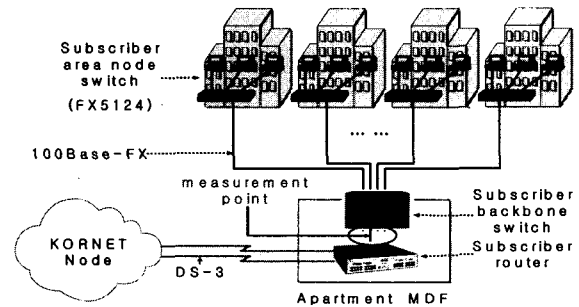


그림 3. 네트워크 구조와 트래픽 측정 위치
Fig. 3. Network configuration of high speed Internet access for apartment complex subscribers

더, TCP 헤더의 모든 정보를 제공하고 있다.

V. 플로우 분석

본 논문은 Coralreef^[11]를 이용하여 플로우를 분석하였다. 이 도구는 플로우를 <소스 IP, 소스 포트, 목적지 IP, 목적지 포트, 프로토콜>이 같은 패킷 스트림을 하나의 플로우로 정의한다. 그리고 64초 동안 패킷이 도착하지 않으면, 플로우가 종료되었다고 판단하였다^[11]. 플로우의 분석 결과는 <IP주소, 프로토콜, 포트번호, 패킷수, 바이트수, 플로우수, 플로우 시작시간, 플로우 종료시간>을 포함하고 있다.

이러한 정보를 이용하여 그림 4에 30분 간격으로, 발생된 플로우의 수와 전송된 바이트 수를 표현하였다. 그림 4에서 놀라운 결과는 업스트림이 다운스트림보다 더 많다는 것이다. 이 결과는 많은 응용프로그램이 기존의 웹 응용프로그램등과 같은 클라이언트-서버의 형태가 아님을 의미하고 있다. 이와 같은 관측 결과는 이용률이 높은 일반 비대칭 회선 가입자단의 네트워크에서는 관찰하기 힘든 결과이다. 그림 4에서 관측된 플로우의 수와 전송된 바이트의 수가 비례하지 않고 있다. 특히, 한밤중에는 트래픽의 플로우는 매우 작으나 트래픽 크기는 매우 크게 나타났다. 이러한 이유를 분석하

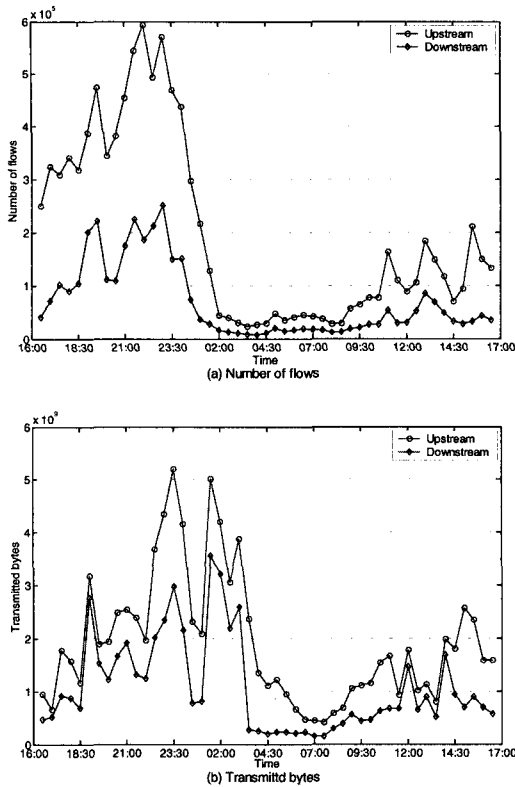


그림 4. 하루동안 전송된 바이트 수와 플로우의 수
Fig. 4. The number of flows and bytes transmitted during one day

기 위해 본 논문은 이 결과를 주요 응용프로그램별로 분류하였다.

응용프로그램의 분류는 프로토콜과 포트 번호를 기준으로 분류하였다. 응용프로그램 분류에서 E-Donkey, 소리바다, V-share, Napster, KazaA, Genie messenger, MSN-messenger 등을 P2P로 분류하였다. 그러나 구루구루 등 포트 번호를 동적으로 할당할 수 있는 P2P 응용프로그램은 P2P로 분류하지 못했다. 그림 5는 24시간 동안 발생한 업스트림과 다운스트림의 트래픽을 응용프로그램별로 분류한 결과이다.

응용프로그램 분석 결과, 다운 스트림은 기타로 분류된 트래픽(65%)을 제외하면 웹 트래픽(25%)이 대부분이었다. 그러나 업스트림 트래픽은 기타로 분류된 트래픽(64%)과 P2P트래픽(32%)이 대부분을 차지하고 있었다. 특히, 기타로 분류된 트래픽은 동적 포트 번호 할당이 가능한 P2P 응용프로그램과 P2P 형식의 프로그램들이 포함되어 있기 때문에 P2P 트래픽이 네트워크에서 차지하는 비율은 관측 결과보다 더 높을 것이다. 다운스트림의 P2P 트래픽의 플로우 수를 관찰하면, 트래픽의 크기에 비해 매우 큰 플로우 수를 나타내고 있다. 이 같은 이유는 P2P 플로우의 78%가 단일 패킷

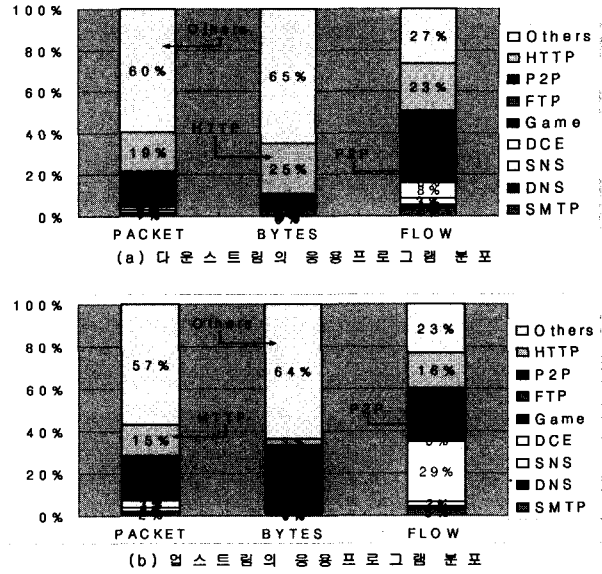


그림 5. 다운스트림과 업스트림의 응용프로그램 분류
Fig. 5. Classification of application in both upstream and the downstream traffic

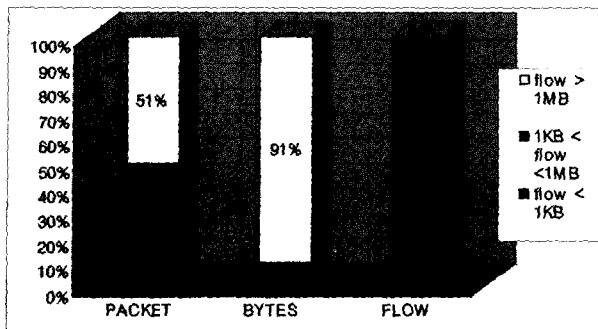
으로 플로우를 구성하고 있으며, 일부 P2P 응용프로그램(소리바다)이 UDP를 통해서 Hello message를 전달하고 있기 때문이다.

플로우의 통계적 특징을 알아보기 위해 하루 동안 측정된 결과를 분석하였다. 그림 6에서 업스트림 플로우의 94% 이상이 1Kbytes이하로 나타났다. 그러나 실제 네트워크에 부하를 주는 트래픽의 대부분은 전체 트래픽의 91%를 차지하는 1Mbytes이상인 플로우들이다.

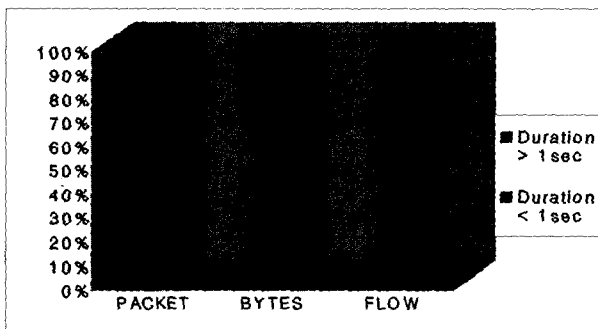
1Mbytes이상의 업스트림 플로우의 경우 “기타”로 분류된 응용프로그램이 66%이며, P2P 응용프로그램은 33%이었다. 플로우의 지속시간 역시 비슷한 분포를 갖고 있다. 1초 이하의 플로는 전체 플로우의 81% 이상이며, 트래픽 발생량의 91%가 1초 이상의 플로우로 구성되어 있다. 1초 이상의 플로우중 “기타”로 분류된 응용프로그램은 64%, P2P 응용프로그램은 32%를 차지하고 있다. 또한 다운스트림 트래픽은 P2P 응용프로그램보다 웹 응용프로그램이 더 많은 분포를 차지하는 것을 제외하면 비슷한 결과를 보이고 있다.

VI. P2P 트래픽의 분석과 모델링

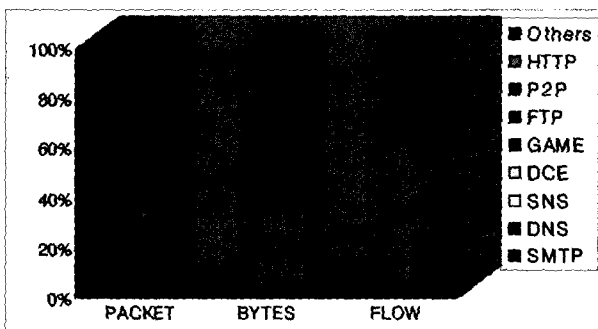
트래픽의 버스트는 자기 유사성으로 평가될 수 있기 때문에 표본화한 트래픽의 Hurst 파라미터를 측정하였다. 여기에서는 자기 유사성을 가진 트래픽으로 잘 알려진 웹 트래픽^[3]과 P2P 트래픽을 비교하였다. 그리고 III장에서 언급한 Heavy-tailed ON/OFF 모델과



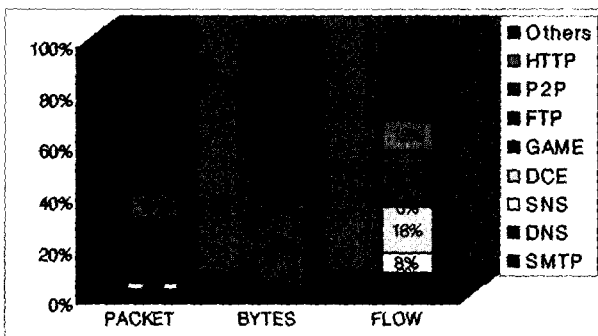
(a) 업스트림의 플로우 크기에 따른 패킷, 플로우 수 분포



(b) 업스트림의 플로우 지속시간에 따른 패킷, 플로우 수 분포



(c) 업스트림의 1MB이상의 응용프로그램 분포

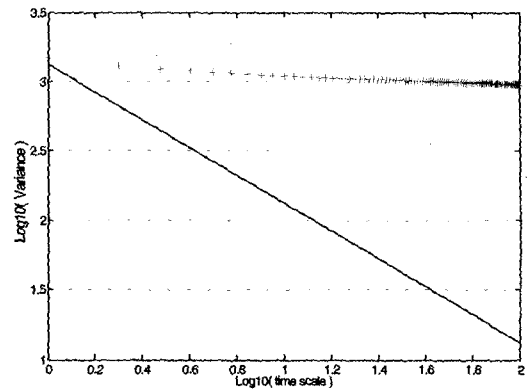


(d) 업스트림의 1초 이상의 응용프로그램 분포

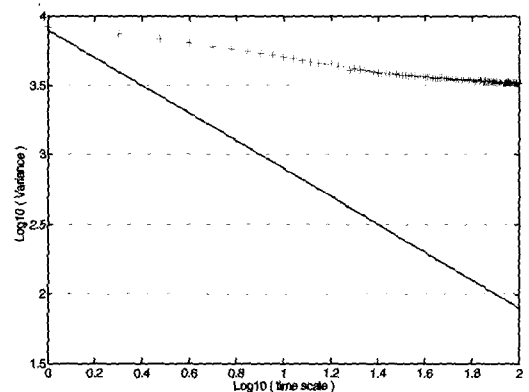
그림 6. 24시간동안 관측된 업스트림 분포

Fig. 6. Upstream flow distribution of observed in 24 hours.

M/Pareto 모델을 이용하여 P2P 트래픽을 모델링하고, SSQ를 이용하여 P2P 트래픽의 손실확률과 평균 지연 시간을 예측하고 이결과를 P2P 트래픽과 비교함으로써



(a) variance-time plot of P2P traffic : H=0.963



(b) Variance-time plot of web traffic : H=0.898

그림 7. P2P 트래픽과 웹(HTTP) 트래픽의 분산-시간 그래프 비교 결과

Fig. 7. Comparison of variance-time plot of both P2P traffic and web(HTTP) traffic

트래픽 모델의 정확성을 평가하였다.

1. 트래픽의 자기 유사성 측정

Hurst 파라미터를 찾기 위해, 본 논문은 인터넷 사용이 가장 많은 11:00-12:00p.m.의 전송된 트래픽의 양을 표본화하였다. 그림 7은 II장에서 소개한 분산-시간 그래프를 이용하여 전체 P2P 응용프로그램들의 Hurst 파라미터를 측정한 결과이다. 예상대로 웹 트래픽은 H=0.898로 강한 자기 유사성을 보이고 있었다. 그러나 P2P 트래픽은 H=0.963으로 웹 트래픽보다 자기 유사성이 더 강하였다. 이것은 P2P 트래픽이 웹 트래픽보다 더욱 버스트하게 나타나고 있음을 의미한다^[2]. P2P 트래픽이 네트워크 트래픽의 상당수를 차지하고 있으며, 다른 트래픽보다 더욱 자기 유사하기 때문에 앞으로 P2P 트래픽이 증가할 경우 이 트래픽은 네트워크 성능을 더욱 저하시킬 것이다.

네트워크의 구조가 사용자간의 대칭적인 트래픽을 수용할 수 없도록 설계되어 있기 때문에 국내의 많은

표 1. 트래픽 모델 파라미터
Table 1. Traffic model parameters

M/Pareto 모델	λ	r	γ	δ
Heavy-tailed ON/OFF 모델	평균	분산	m	d
	78.75	1480.58	1.9311	0.5

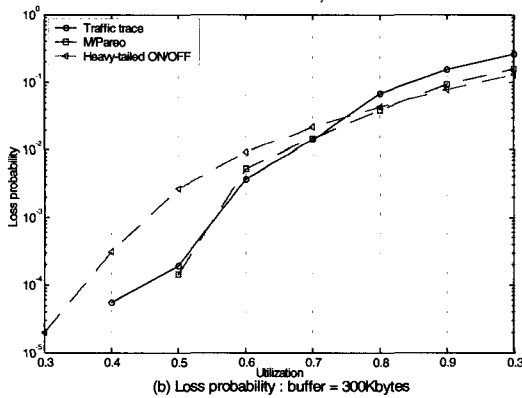
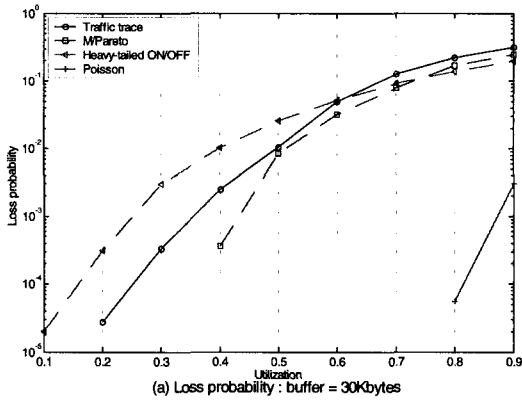


그림 8. 손실확률을 이용한 트래픽 모델의 정확성 평가.
Fig. 8. Verification of traffic model accuracy using loss probability.

ISP들은 큰 혼잡이 발생하지 않도록 사용자의 업스트림 트래픽 전송속도를 제한하고 있다. 비단 ISP 업체들 뿐만 아니라 학내 캠퍼스 망도 비슷하다. 현재 P2P 응용프로그램뿐만 아니라 P2P 응용 프로그램의 사용자 또한 지속적으로 늘어나고 있다. ISP 업체들은 다른 응용프로그램의 QoS를 감소시키지 않으면서, P2P 트래픽을 얼마나 잘 관리하는가가 당분간은 사용자단의 네트워크 관리의 중요한 문제점이 될 것이다.

2. 트래픽 모델링

인터넷 사용이 높은 시간대(11:00-12:00p.m.)의 트래

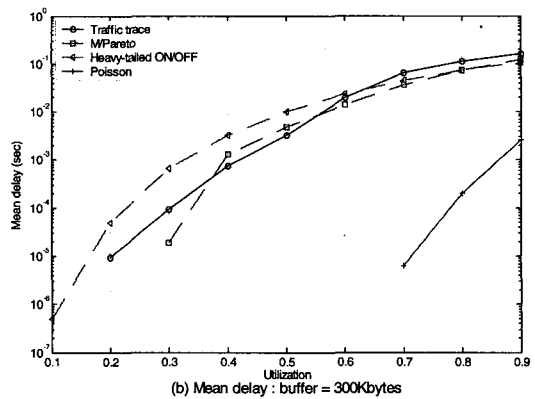
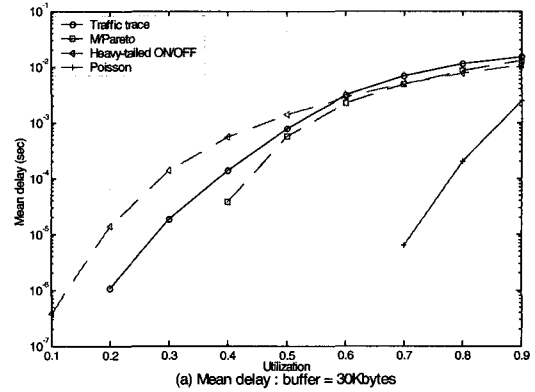


그림 9. 평균 지연시간을 이용한 트래픽 모델의 정확성 평가.

Fig. 9. Verification of traffic model accuracy using mean delay.

픽을 모델링하기 위해 본 논문은 트래픽 측정 결과를 업/다운 구분 없이 0.1초 간격으로 전송된 트래픽의 양을 샘플링 하고, 트래픽을 묘사하였다. 측정된 트래픽의 통계적 결과는 평균: 78.75Kbytes/0.1sec, 분산: 1480.58 Kbytes², Hurst 파라미터 H=0.963로 나왔다. 이 세 가지 파라미터를 이용하여 M/Pareto 모델과 Heavy-tailed ON/OFF 모델 그리고 기존의 Poisson 모델을 사용하여 트래픽을 묘사하였다. 두 트래픽 모델의 모델링 파라미터는 표 1에 나타내었다.

트래픽 모델의 정확성을 평가하기 위해 SSQ의 이용률 변화에 따른 손실확률과 평균 지연시간을 조사하였다. 그림 8은 버퍼 크기변화에 따른 두 트래픽 모델의 손실확률을 보여주고 있다. 기존의 Poisson 모델은 실제 트래픽을 거의 묘사하지 못하고 있음을 확인할 수 있다. 반면, M/Pareto 모델은 서버의 이용률이 50% 이상일 경우 정확히 트래픽의 손실확률을 나타내고 있음을 알 수 있다. Heavy-tailed ON/OFF 모델 또한 이용률이 50%이하에서는 실제 트래픽과 차이를 보이고 있으나 Poisson 모델 보다 매우 정확히 모델링하고 있음을 알 수 있다. 그림 8

의 (b)에서 Poisson 모델 결과가 나타나지 않은 것은 손실확률 결과가 모든 범위에서 0으로 나타났기 때문에 표시하지 않았다.

그림 9에서는 버퍼 크기 변화에 따른 평균 지연시간의 예측 결과를 나타내고 있다. Poisson 모델링 결과는 실제 트래픽을 거의 모델링 할 수 없음을 보이고 있다. Heavy-tailed ON/OFF 모델은 앞서 살펴보았던 평균 손실확률의 경우와 같이 모든 이용률의 범위에서 Poisson 모델보다 매우 정확히 모델링하고 있다. 반면 M/Pareto 모델은 이용률이 40%이상일 경우 실제 트래픽과 매우 유사하게 모델링하고 있음을 확인할 수 있다. 지금까지의 시뮬레이션 결과는 기존의 Poisson 모델보다 자기 유사 트래픽 모델인 M/Pareto 모델과 Heavy-tailed ON/OFF 모델이 P2P 트래픽을 매우 정확히 묘사할 수 있음을 보여주고 있다.

본 논문의 모델링 결과, 네트워크를 설계하거나 성능을 분석할 때, 기존의 단순한 Poisson 모델을 이용하여 나타낼 경우 실제 트래픽의 환경을 거의 반영하지 못함을 알 수 있다. 그러나 M/Pareto 모델과 Heavy-tailed ON/OFF 모델은 위의 결과에서 확인할 수 있듯이 실제 트래픽을 정확히 모델링하여 네트워크의 설계와 성능을 분석하기 위한 입력 트래픽으로 사용할 수 있을 것이다.

VII 결 론

본 논문은 최근 가입자 트래픽을 측정하였으며, 이를 바탕으로 빠르게 성장하고 있는 P2P 트래픽의 특징을 분석하고, 이를 간단한 트래픽 모델을 이용하여 모델링하였다. 분석 결과, P2P 응용프로그램의 많은 사용으로 네트워크 트래픽의 특징이 변하고 있었다. P2P 트래픽은 매우 버스트하게 나타났으며, 웹 응용프로그램처럼 서버-클라이언트 관계의 응용프로그램들은 비대칭적인 트래픽을 발생시키지만, P2P 응용프로그램은 대칭적인 트래픽을 발생시킨다. 이러한 새로운 변화가 네트워크에 미치는 영향은 다음과 같다. 첫째, P2P 응용프로그램의 사용이 늘어나면, 현재의 ADSL, 케이블 모뎀 같은 비대칭 인터넷 접속 기술은 개선이 되어야 한다. 이런 기술은 P2P 응용프로그램이 요구하는 업스트림 대역폭을 충족시켜줄 수 없다. 둘째, ISP의 네트워크 구조가 바뀌어야 한다. 현재 ISP 네트워크 구조는 '정보의 대부분이 IDC에 있다'는 가정으로 네트워크가 설계되었다. 그렇기 때문에 사용자들 사이의 P2P 트래픽의 전달

은 비효율적이다. 그러므로 앞으로 ISP의 네트워크 구조는 메쉬 구조로 바뀌어야 한다. 셋째, P2P 트래픽은 기존의 트래픽보다 더욱 버스트하고, 앞으로 트래픽의 대부분이 P2P 트래픽일 것이다. 그러므로 P2P 트래픽을 제한하지 않는다면, 혼잡 상황은 더욱 쉽게 발생할 것이며, QoS를 더욱 낮추는 결과를 야기할 것이다.

본 논문은 또한 Heavy-tailed ON/OFF 모델과 M/Pareto 모델을 이용하여 측정된 트래픽을 묘사하였다. 이 모델들은 기존의 Poisson 모델보다 매우 정확히 트래픽의 손실확률과 평균 지연시간을 정확히 모델링할 수 있었다. 그러므로 이 두 모델은 P2P 응용프로그램의 종단간 QoS 예측과 다른 응용프로그램의 영향을 알 수 있다. 가까운 미래는 P2P 응용프로그램과 같은 형태의 응용프로그램의 트래픽이 대부분의 대역폭을 점유하게 될 것이다. P2P 응용프로그램의 QoS를 보장해주고 효율적인 네트워크 서비스를 제공하기 위해 네트워크 트래픽의 특징을 이해하는 것은 매우 중요한 일이며, 더 많은 연구가 필요할 것이다.

참 고 문 헌

- [1] 김명섭, 강훈정, 홍원기, "Flow Grouping을 통한 P2P 트래픽 분석방법에 관한 연구", Proc. of KNOM 2003 Conference, Daejeon, Korea, May 22-23, 2003, pp. 210-218.
- [2] Kevin Thompson, Gregory J. Miller and Rick Wilder " Wide-Area Internet Traffic Patterns and Characteristics," IEEE Network, vol. 11(6) pp. 10-23, Nov-Dec. 1997.
- [3] Kihong Park, Walter Willinger "Self-Similar Network Traffic and Performance Evaluation," JOHN WILEY & SONS, pp. 1-38, 2000.
- [4] W. E. Leland, M. S. Taqqu, W. Willinger and D. V. Wilson, "On the self similar nature of Ethernet traffic (extension version)," IEEE/ACM Transactions on Networking, vol. 2(1), pp. 1-15, 1994.
- [5] Mark E. Crovella and Azer Bestavros "Self-Similarity in World Wide Web Traffic Evidence and Possible Causes," IEEE ACM Transactions on Networking, pp.835-846 December 1997.
- [6] O. Rose, "Estimation of the Hurst Parameter of Long-Range Dependent Time Series," Research Report, 1996.
- [7] P. Pruthi and A. Erramilli, "Heavy-tailed ON/OFF Source Behaviour and Self-similar

Traffic," IEEE ICC'95, pp. 445-150, 1995.

[8] Ronald G. Addie, timothy D. Neame, Moshe Zukerman, "Modeling Superposition of Many Sources Generating Self Similar Traffic," IEEE ICC'99, pp. 387-391, 1999.

[9] Ronald G. Addie, Moshe Zukerman and Tim Neame "Fractal Traffic : Measurements, Modeling and Performance Evaluation," IEEE ICC'95, vol. 3, pp. 471-476, April 1998.

[10] <http://www.tcpcdump.org/>

[11] <http://www.caida.org/tools/measurement/coralreef/>

저 자 소 개



주 성 돈(학생회원)
 2003년 2월 아주대학교
 전자공학과 졸업
 2003년 3월~현재 아주대학교
 대학원 전자공학과 석사
 과정.

<주관심분야: Traffic engineering, Wireless Communication>



이 채 우(정회원)
 1985년 서울대학교 제어계측
 학사졸업
 1988년 한국과학기술원
 전자공학과 석사
 1995년 University of Iowa 박사
 1985년 1월~1985년 12월 (주)금성
 통신 연구원.

1988년 9월~1999년 3월 한국통신 선임연구원
 1999년 3월~2001년 9월 Lucent Technologies
 Korea 이사.

2001년 9월~2002년 2월 한양대학교 겸임교수
 2002년 3월~현재 아주대학교 전자공학과 조교수
 <주관심분야: 광대역 통신망, Ubiquitous network-
 king, Traffic Engineering>

