

P2P 트래픽의 특성 분석과 트래픽 모델링

준회원 주성돈*, 정회원 이채우**

Emerging P2P Traffic Analysis and Modeling

Sung-don Joo* Associate Member, Chae-woo Lee** Regular Member

요 약

현재 급속히 늘어나고 있는 P2P(Peer to Peer) 애플리케이션은 버스트(burst)한 트래픽을 많이 발생시켜 네트워크에 큰 부하를 주고 있다. 따라서 P2P 트래픽의 특성을 분석하고 이것이 네트워크 및 QoS에 미치는 영향을 이해하는 것은 매우 중요하다. 본 논문에서는 P2P 애플리케이션의 사용현황을 알아보기 위해 캠퍼스 네트워크의 트래픽을 측정하고, 측정된 트래픽의 플로우를 분석하고, P2P 트래픽의 자기 유사성을 측정하여 웹 트래픽과 비교 분석하였다. 분석 결과 P2P 트래픽은 기존의 웹 트래픽보다 더 버스트하였으며, 이는 P2P 트래픽이 네트워크 성능에 악영향을 미치고 있음을 알 수 있었다. 측정된 P2P 트래픽을 네트워크 성능평가 및 예측에 용이하게 이용할 수 있게 하기 위해, 이를 자기 유사 트래픽 모델을 이용하여 모델링하였다. 유도한 트래픽 모델의 정확성을 평가하기 위해, SSQ(Single Server Queue)를 이용하여 네트워크 성능 평가에서 매우 중요한 파라미터인 큐의 길이와 손실을 예측하였으며, 이것을 측정된 트래픽의 그것과 비교하여 트래픽 모델의 정확성을 평가하였다. 본 논문에서 사용한 자기 유사 트래픽 모델은 P2P 트래픽을 잘 모델링하고 있었으며, 이 결과는 P2P 트래픽의 특성을 고려한 네트워크 설계와 성능을 평가하고 예측할 때 사용될 수 있다.

ABSTRACT

Rapidly emerging P2P(Peer to Peer) applications generate very bursty traffic, which gives a lot of burden to network, and the amount of such traffic is increasing rapidly. Thus it is becoming more important to understand the characteristics of such traffic and reflect it when we design and analyze the network. To do that we measured the traffic in a campus network and present flow statistics and traffic models of the measured traffic, and compare them with those of the web traffic. The results indicate that P2P traffic is much burstier than web traffic and as a result it negatively affects network performance. We modeled P2P traffic using self-similar traffic model to predict packet delay and loss occurred in network which are very important to evaluate network performance. We also predict queue length distribution and loss probability in SSQ(Single Server Queue). To assess accuracy of traffic model, we compare the SSQ statistics of traffic models with that of the traffic trace. The results show that self-similar traffic models we use can predict P2P traffic behavior in network precisely. It is expected that the traffic models we derived can be used when we design network capacity and predict network performance and QoS of the P2P applications.

Key Words : Traffic model, P2P, Self-similarity, FGN, M/Pareto

* 아주대학교 전자공학과 멀티미디어 네트워킹 연구실(sungdon@ajou.ac.kr), ** 아주대학교 (cwlee@ajou.ac.kr)

논문번호 : 030526-1125, 접수일자 : 2003년 11월 25일

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

I. 서론

최근, 인터넷을 보면 과거에는 볼 수 없었던 e-donkey, 소리바다와 같은 P2P(Peer to Peer) 애플리케이션들이 많이 이용되고 있다. 기존의 웹 애플리케이션 등은 대부분의 컨텐츠가 IDC (Internet Data Center)와 같이 네트워크의 중심에서 있었고, 일반 사용자들은 이것을 주로 다운로드하였다. 따라서 네트워크의 하향 트래픽이 상향 트래픽보다 훨씬 많았다. 그러나 P2P의 등장으로 인해 일반 사용자들간의 트래픽이 많아지고 있다. 이처럼 P2P 애플리케이션의 사용으로 트래픽의 발생 패턴이 달라지고 있음을 알 수 있다. 이와 같은 P2P 응용은 여러 가지 형태로 네트워크에 전과는 다르게 영향을 주고 있습니다. 애플리케이션에 따라 트래픽의 형태가 다르며, 트래픽에 따라 네트워크의 성능과 애플리케이션의 QoS가 달라진다. 현재 P2P가 많이 이용되고 있으며, 앞으로 지속적으로 증가할 것이므로 이를 이해하는 것은 매우 중요하다.

본 논문에서는 최근 P2P 트래픽의 특성을 파악하기 위해 캠퍼스 네트워크 트래픽을 측정하고, 그 결과를 플로우별로 분석하였다. 또한 측정된 트래픽을 네트워크 성능과 QoS를 분석하는데 이용하기 위해, 트래픽 모델을 이용하여 측정된 P2P 트래픽을 모델링하였다. 이렇게 유도된 트래픽 모델의 정확성을 검증하기 위해 SSQ를 이용하여 큐의 길이와 손실을 예측하고, 이 결과를 측정된 P2P 트래픽의 결과와 비교하였다.

본 논문의 구성은 다음과 같다. 2장에서 트래픽의 버스트를 평가할 수 있는 데이터 트래픽의 자기 유사성(Self-Similarity)에 대해서 설명한다. 3장에서, M/Pareto 모델과 FGN 모델의 특징에 대해 설명한다. 4장에서 큐(queue) 모델을 이용하여 트래픽 모델의 정확성을 평가하는 방법을 소개한 후, 5장에서 P2P 트래픽을 측정된 캠퍼스 네트워크 망을 간단히 살펴보고, 측정 결과를 정리하였다. 6장에서는 웹 트래픽과 P2P 트래픽인 소리바다 트래픽을 3장에서 소개한 트래픽 모델을 이용하여 모델링하였다. 마지막으로 7장에서 본 논문을 정리한다.

II. 트래픽의 자기 유사성

트래픽의 버스트는 네트워크 성능 평가에 매우 중요한 요소이다. 이러한 트래픽의 버스트를 측정할

수 있는 방법으로 자기 유사성을 이용할 수 있다. 즉, 자기 유사성이 강하면 트래픽의 버스트 정도가 심해진다.

일반적으로 트래픽의 자기 유사성은 모든 시간의 범위에서 트래픽의 버스트가 사라지지 않는 현상을 의미한다. 이 같은 현상은 통계적 개념을 이용하여 정의할 수 있다. 정상상태 시계열(time series) X 에 대해 m -aggregated 시계열 $X^{(m)} = \{X_{mt}, k=0,1,2,\dots\}$ 은 인접한 m 크기의 블록을 겹치지 않게 원래의 시계열을 합함으로써 정의한다^{[2],[4]}.

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

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

$$\begin{aligned} \text{Var}[X^{(m)}] &= \frac{\text{Var}[X]}{m^\beta} \text{ Variance} \\ R_{X^{(m)}}(k) &= R_X(k) \text{ Autocorrelation} \end{aligned} \quad (2)$$

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

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

지금까지 통계적 개념을 이용하여 자기 유사성을 정의하였다. 2차 모멘트 이상의 통계적 값을 사용하면 더욱 정확히 묘사할 수 있지만, 복잡성의 문제로 인해 2차 모멘트까지만 이용한다.

트래픽의 자기 유사성은 Hurst 파라미터를 이용하여 평가할 수 있다^{[2],[4],[9]}. Hurst 파라미터는 0.5 ~ 1 사이의 값을 가지며, $H=0.5$ 는 자기 유사성이 존재하지 않는 경우이며, $H=1.0$ 은 정확히 자기 유사한 경우이다. 즉, Hurst 파라미터 값이 1에 가까워질수록 자기 유사성이 강해지는 것이다. Hurst 파라미터 값을 얻는 방법은 분산-시간 그래프, R/S 그래프, Periodogram, IDC-그래프, Whittle estimate 등이 있다. 본 논문에서는 육안으로 손쉽게 자기 유사성을 확인할 수 있는 분산-시간 그래프와 비교적 정확한 측정 방법인 Periodogram을 이용하였다^[9]. 분

산-시간 그래프는 m-aggregated된 시계열 $X^{(m)}$ 의 분산을 m 에 관하여 로그함수로 나타내면 그래프의 기울기가 -1 보다 큰 값을 가진다. 최소 제곱 직선 근사법 (least square line fitting)을 이용하여 분산 값의 기울기(β)를 구하면 Hurst 파라미터 값을 구할 수 있다^{[4][9]}. 기울기(β)와 Hurst 파라미터는 $\beta=2-2H$ 의 관계를 가진다.

$$\text{Var}[X^{(m)}] \sim \frac{\text{Var}[X]}{m^\beta}, \quad (4)$$

$$\log(\text{Var}[X^{(m)}]) \sim \log(\text{Var}[X]) - \beta \log(m)$$

Periodogram은 스펙트럼 밀도의 측정을 바탕으로 한다. LRD(Long Range Dependence)의 특성은 주파수가 0으로 접근을 할 때 스펙트럼의 분포에 영향을 주게 된다. 스펙트럼 밀도는 자기 공분산의 푸리에 변환 형태이기 때문에, 자기 공분산 함수를 이용하여 스펙트럼 밀도를 얻을 수 있다. 이산 시간에서 정의되는 시계열 $X(t)$ 는 구간 N 에 대해서 푸리에 변환을 통해서 식 (5)처럼 스펙트럼 밀도를 얻을 수 있다.

$$I(w) = \frac{1}{2\pi N} \sum_{k=1}^N |x_k e^{jkw}|^2 \quad (5)$$

이 값을 Periodogram이라 한다. 주파수가 0으로 접근함에 따라 스펙트럼은 $I(w) \sim w^{1-2H}$ 의 형태를 갖는다. Hurst 파라미터 값은 로그함수로 표현하여 기울기를 이용해서 얻을 수 있다.

III. 자기 유사 트래픽 모델

본 논문에서는 비교적 간단한 개념을 이용하여 트래픽의 자기 유사성을 나타낼 수 있는 트래픽을 모델을 사용하여 P2P 트래픽을 모델링 하였다. 다음절은 M/Pareto 모델과 FGN 모델을 설명하고 있다.

1. M/Pareto 모델

M/Pareto 모델은 Heavy-tailed ON/OFF 모델^[15]의 특별한 경우라 할 수 있다. 즉, 트래픽이 발생되는 세션의 길이는 Heavy-tailed 분포를 가지며, 트래픽을 발생시키는 각 세션의 간격은 지수분포를 갖는다. 그림 1은 M/Pareto 모델의 개념을 보이고 있다^{[16][17]}. 세션의 도착은 도착율이 λ 인 포와송 프로세스로 표현된다. 그러므로 인접 버스트 세션 사이의 도착 시간은 평균이 $1/\lambda$ 인 지수 분포가 된다. 길이가 T인 시간간격동안 도착하는 버스트 세션의

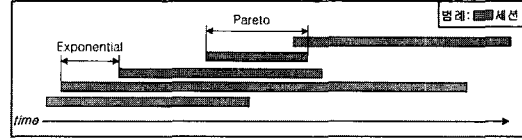


그림 1. M/Pareto 모델의 세션 관계

수는 평균이 λT 인 포와송 프로세스가 된다^{[16][17]}.

M/Pareto 모델에서 각 버스트 세션의 지속시간(d)은 상호 독립적이며, Pareto 확률 분포를 가진다. Pareto 분포는 LRD(long range dependence)특징을 갖고 있어 M/Pareto 모델에 자기 유사성을 갖게 한다. Pareto complementary 분포 함수는 다음과 같다.

$$\Pr(d > x) = \begin{cases} (\frac{x}{\delta})^{-\gamma}, & x \geq \delta \\ 1, & otherwise \end{cases} \quad (6)$$

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

측정된 트래픽을 통해 얻을 수 있는 통계적 결과인 평균, 분산, Hurst 파라미터를 이용하여 M/Pareto 모델의 파라미터를 얻어야 한다. 그러므로 자기 유사성은 변화시키지 않으면서 λ, r 값을 이용하여 SSQ에서 성능을 가장 잘 나타내는 값을 찾아야 한다. M/Pareto 모델의 자세한 내용은 [16],[17]에서 찾아볼 수 있다.

2. FGN (Fractional Gaussian Noise) 모델

FBM은 정확히 자기 유사한 특징을 갖고 있으며, 이 함수의 증가함수 역시 자기 유사한 특징을 갖고 있다. FGN 함수는 FBM의 증가함수로서 나타낼 수 있다. FGN은 일정 시간 $(0, t]$ 동안 들어온 트래픽의 양을 $A(t)$ 으로 정의하여 다음과 같이 나타낸다^{[10]-[14]}.

$$A(t) = mt + \sqrt{ma}Z(t) \quad (7)$$

FGN 모델은 $m, a, Z(t)$ 의 세 가지 파라미터를 사용하여 자기 유사성을 가진 트래픽을 나타낸다. 파라미터 m 은 평균, a 는 IDC(Index of Dispersion Count), $Z(t)$ 는 Hurst 파라미터를 이용하여 자기 유사성을 표현하는 FGN 함수이다. $a, Z(t)$ 는 트래픽의 형태를 결정하며, m 은 트래픽의 양을 결정하는

파라미터이다. 이처럼 FG N 모델은 트래픽을 발생시키는 소스의 특징을 묘사하지 않고 중첩된 (aggregation) 트래픽에 자기 유사성을 추가하는 방법을 사용하고 있다. 그러나 m , a , H 세 가지 파라미터를 이용하여 쉽게 자기 유사성을 가진 트래픽을 만들 수 있는 장점이 있다.

트래픽 모델링은 측정된 트래픽에서 얻은 통계적 결과 값을 이용하여 앞서 설명한 트래픽 모델에 필요한 파라미터 값을 찾는 것이다.

IV. SSQ(Single Server Queue)를 이용한 트래픽 모델의 정확성 분석

트래픽 모델의 주요한 응용은 트래픽이 네트워크를 통과할 때 트래픽의 특성을 예측하는 것이다. 이상적 트래픽 모델은 실제 트래픽의 모든 통계적 특성을 정확하게 나타낼 수 있어야 한다. 그러나 그런 모델은 문자 그대로 이상적인 것이다. 실제 트래픽 모델은 몇 가지 통계적 결과만을 이용하여 모델링한다. 그러므로 모델의 정확성을 평가할 기준이 필요하다. 네트워크는 하나 이상의 SSQ (single server queue)로 묘사할 수 있다. 그러므로 트래픽 모델의 정확성은 SSQ에서의 성능을 통해 평가될 수 있다^[8]. 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.

$$Q_n = \max(0, Q_{n-1} + A_n - C) \quad (8)$$

식 (8)을 이용하여 무한 버퍼를 사용할 경우, 시스템의 버퍼의 길이를 구할 수 있고, 유한 버퍼를 사용할 경우 시스템의 손실확률을 구할 수 있다. 그림 2는 SSQ를 이용하여 실제 트래픽과 모델을 통해 발생된 트래픽을 비교하는 과정을 나타내고 있다. 본 논문에서는 위에서 언급했던 방법을 통해서 서버의 이용률 변화에 따른 모델의 정확성을 평가할 것이다. 여기서 이용률은 입력 트래픽의 양과 서버에서 처리할 수 있는 양의 비율로서 표현한다.

V. 트래픽 측정위치 및 결과

기존의 논문에서는 P2P 트래픽의 분석과 네트워

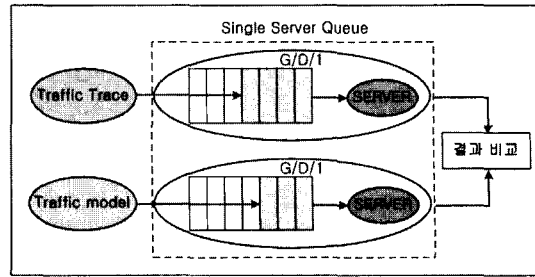


그림 2. SSQ 결과 비교

크 성능 평가에 이용될 수 있는 입력 트래픽의 모델링은 이루어지지 않았다. 본 논문에서는 현재 많이 사용되고 있는 P2P 트래픽 특성을 이해하기 위해 2003년 8월 26일부터 28일까지 48시간동안 캠퍼스 망 관리팀의 도움을 얻어 캠퍼스 네트워크 망을 미러링 하였다. 측정이 이루어진 곳은 캠퍼스 망과 KT(Korea Telecom)의 KORNET망이 연결되는 라우터이며, 측정 도구로써 Tcpdump를 이용하여 이더넷 프레임의 첫 68bytes를 미러링 하였다. 측정된 패킷은 IP 헤더와 TCP 헤더 등을 포함하고 있어 분석에 필요한 기본 정보는 모두 포함하고 있다. 그림 3은 캠퍼스의 네트워크 망을 간략화 한 구조를 보여주고 있다. 측정이 이루어진 곳은 각 단과 대학 건물을 기가 이더넷으로 연결하고 있으며 교내의 모든 네트워크가 이곳으로 연결되어 있다.

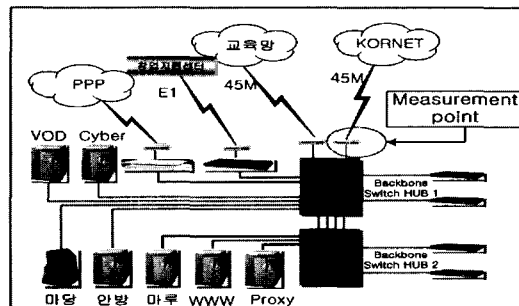


그림 3. 캠퍼스 네트워크 구성도

플로우의 분석을 위해 Coralreef를 사용하였으며, 플로우를 <소스 IP, 소스 포트, 목적지 IP, 목적지 포트, 프로토콜>이 같은 단 방향 스트림으로 정의한다^[1]. 그리고 64초 동안 패킷이 도착하지 않으면 플로우가 종료되었다고 판단하였다^[1]. 애플리케이션의 분류는 프로토콜 번호와 포트 번호를 기준으로 분류되었다. P2P 트래픽은 E-Donkey, 소리바다, Napster, V-share, MSN-messenger, Genie-messenger

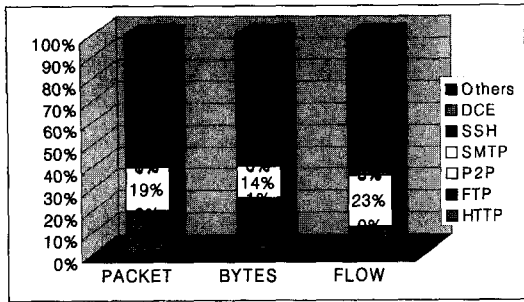


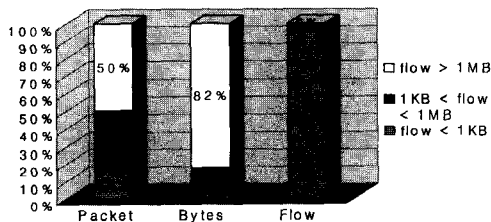
그림 4. 캠퍼스 트래픽의 플로우 분석 결과

등 비교적 잘 알려진 애플리케이션들을 포함하고 있다. 그러나 구루구루 (Guruguru) 처럼 동적으로 포트 번호할당이 가능한 P2P 애플리케이션은 분류하지 못하였다. 측정 결과, 전체 트래픽의 약 25%는 웹 트래픽이 차지하고 있었으며, 14% 이상이 P2P 트래픽이었다. 본 논문에서 P2P 트래픽의 특성 파악을 위해 소리바다 트래픽(4%) 이용하였다. 소리바다 트래픽은 모든 P2P 트래픽을 나타낼 수 없지만 일반인들에게 널리 알려져 있으며, P2P의 일반 형태를 갖기 때문에 P2P 트래픽의 특성파악에 도움이 될 것이다. 분류되지 못한 60%의 애플리케이션은 P2P 트래픽을 포함하여 게임, 스트림 방송 등을 포함하고 있다. 이처럼 분류되지 못한 트래픽이 많은 근본적인 이유는 잘 알려진 포트번호를 기준으로

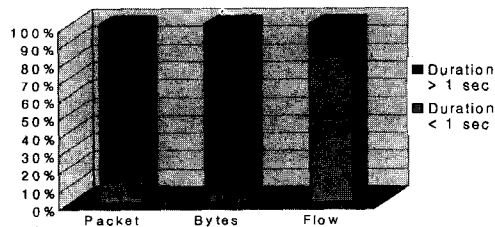
분류를 하였기 때문에 모든 애플리케이션의 분류는 불가능하다.

플로우를 크기와 지속시간을 기준으로 분류한 결과를 그림 5에 나타내었다. 이 플로우 분석 결과에서 전체 플로우의 90% 이상이 1Kbytes이하로 나타났다. 그러나 실제 네트워크에 부하를 주는 트래픽의 대부분은 전체 트래픽의 82%를 차지하는 1Mbytes 이상인 플로우들이다. 플로우의 지속시간 역시 비슷한 분포를 갖고 있다. 1초 이하의 플로우는 전체 플로우의 80% 이상이며, 트래픽 발생량의 95%가 1초 이상의 플로우로 구성되어 있다. 이처럼 네트워크 트래픽의 대부분을 차지하는 일부 플로우를 구분하여 플로우 기반 스위칭을 한다면 고속의 전송이 가능할 것이다.

트래픽의 버스트 정도를 알기 위해, 2003년 8월 26일 15:30분부터 1800초 동안 0.1초 간격으로 웹 트래픽과 소리바다 트래픽을 표본화한 트래픽의 Hurst 파라미터를 측정하였으며, 이 결과들을 비교하였다^[5]. 표 1은 표본화한 트래픽의 통계적 결과를 보이고 있다. Hurst 파라미터는 II절에서 언급한 분산-시간 그래프와 Periodogram을 이용하여 그림 6에 나타내었다. 웹 트래픽은 분산-시간 그래프에서 $H=0.835$, Periodogram에서 $H=0.938$ 이 나왔으며, P2P 트래픽인 소리바다 트래픽은 분산-시간 그래프에서 $H=0.914$, Periodogram에서 $H=1.023$ 이 나왔다. Hurst 파라미터는 1보다 클 수 없다. 그러나 Periodogram의 측정값이기 때문에 1보다 큰 경우가 발생하나 이는 매우 큰 자기 유사성을 가졌음을 의미하고 있다.



(a) 플로우 크기 에 의한 분류



(b) 플로우 지속시간 에 의한 분류

그림 5. 플로우 크기와 지속시간에 따른 분포 분석

표 1. 소리바다와 웹 트래픽의 통계적 결과

	web traffic	soribada traffic
평균	75.11 Kbytes/0.1sec	19.46 Kbytes/0.1sec
분산	633.996 Kbytes ²	139.58 Kbytes ²
H	0.835	0.914

표 1에 언급한 Hurst 파라미터는 분산-시간 그래프를 이용하여 측정된 결과를 나타내었다. Hurst 파라미터 측정 결과, 웹 트래픽은 강한 자기 유사성을 갖고 있음을 알 수 있었다. 그러나 P2P 트래픽인 소리바다 트래픽은 이보다 더욱 강한 자기 유사성을 가짐을 알 수 있다. 이는 소리바다 트래픽이 기존의 웹 트래픽보다 더욱 버스트하여 라우터에 더

욱 큰 부하를 주고 있음을 보여주고 있다⁴⁾. 그렇기 때문에 P2P 트래픽의 관리가 더욱 중요해지고 있다.

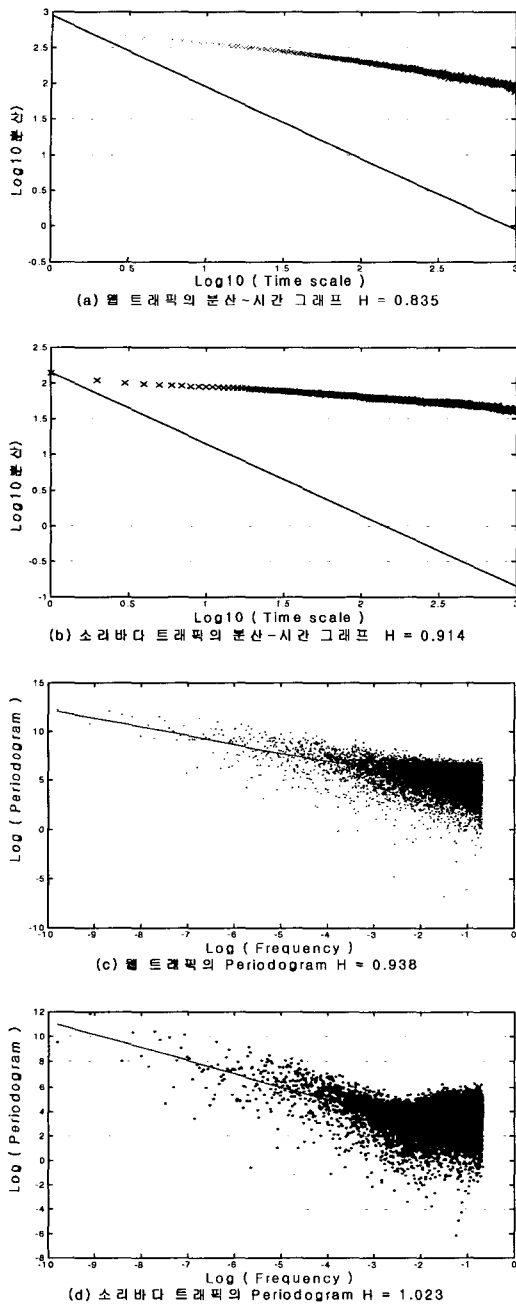


그림 6. 두 트래픽의 분산-시간 그래프와 Periodogram.

VI. 트래픽 모델링

네트워크에서 P2P 트래픽의 특성을 분석 및 예측

하기 위해 앞 절에서 언급한 M/Pareto 모델과 FGN 모델을 이용하여 P2P 트래픽인 소리바다 트래픽을 모델링 하였으며, 이 결과를 비교하기 위해 웹 트래픽을 모델링 하였다. 트래픽 모델의 정확성은 SSQ의 이용률을 변화시키면서 SSQ의 버퍼 길이 분포와 손실확률을 구하여 측정된 트래픽의 그것과 비교하여 검증하였다. 네트워크에서의 지연은 버퍼의 길이와 비례관계에 있으므로 버퍼 길이 분포는 네트워크에서 패킷의 지연을 유추할 수 있다. 그리고 유한 버퍼를 사용할 경우 SSQ의 버퍼길이를 웹 트래픽은 300Kbytes와 700Kbytes, 소리바다 트래픽은 100Kbytes와 200Kbytes로 달리하여 SSQ의 손실확률을 구하였다.

1. M/Pareto 트래픽 모델링

M/Pareto 모델을 사용하기 위해 Pareto 분포의 시작점을 0.1초로 하여 버스트가 최소한 한 표본화 구간동안 지속되도록 하였다. M/Pareto 모델은 III절에서 설명한 바와 같이 측정된 세 가지 통계적인 값을 이용하여 네 가지 트래픽 모델 파라미터를 얻어야 한다. 그렇기 때문에 위에서 설명한 방법으로 λ, r 을 변화시키면서 가장 적절한 파라미터 값을 얻어야 한다. 여기서 Pareto 분포를 결정하는 δ, γ 파라미터 값은 변화시키지 않으므로 트래픽의 자기 유사성은 변하지 않는다. 표 2는 SSQ의 이용률이 60%일 때 각각의 트래픽에 대해 최적으로 구한 M/Pareto 모델의 파라미터를 보여주고 있다.

트래픽 모델 결과에서 우리는 M/Pareto 모델이 Pareto 분포 함수의 사용으로 인해 다소 불안정한 결과를 보이는 경우를 접할 수 있었다. 이는 실제 대부분의 세션 길이는 제한되지만 Pareto 분포 확률 변수 값은 무한한 분산 값을 갖기 때문에 세션의 길이가 지나치게 커지는 경우를 만들게 된다. 캠퍼스 플로우 분석 결과 일반적인 플로우의 대부분(82%이상)은 1초를 넘지 못하였다. 따라서 본 논문에서는 자기 유사성의 변화를 주지 않으면서 세션의 길이를 제한하였다. 세션 길이를 제한하는 방법은 매우 긴 세션 길이(일반적으로 100초 이상)를 표현하지 못하지만, Pareto 분포의 확산을 제한하고

표 2. SSQ의 이용률이 60%일 때 모델 파라미터

	λ	r	δ	γ
Web	2.034	291.1	0.1	1.332
Soribada	340.82	7.95	0.1	1.172

결과를 안정시킬 수 있다.

Pareto 분포 값을 제한한 이후 더욱 안정된 결과를 얻을 수 있었다. 본 논문에서 사용한 Pareto 분포 제한 값은 반복적인 시뮬레이션을 통해 얻었다.

그림 7의 트래픽 모델링 결과에서, SSQ의 이용률이 70%를 넘게 되면, SSQ의 큐 길이 분포가 실제 트래픽과 매우 유사한 결과를 보이고 있다. 그러나 소리바다 트래픽 모델링 결과는 다소 차이가 발생하고 있으나 실제 트래픽의 특성을 잘 따르고 있다. 그림 8은 SSQ의 버퍼의 길이에 따른 손실확률 결과를 보여주고 있다. 버퍼의 길이 변화에 불구하

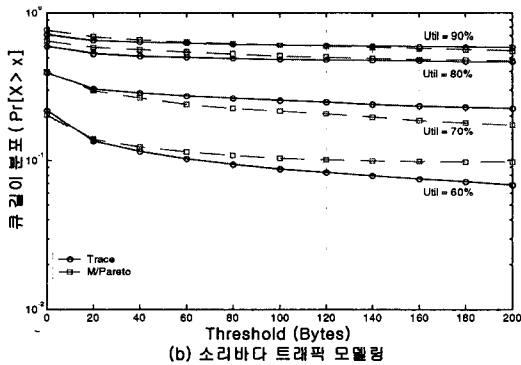
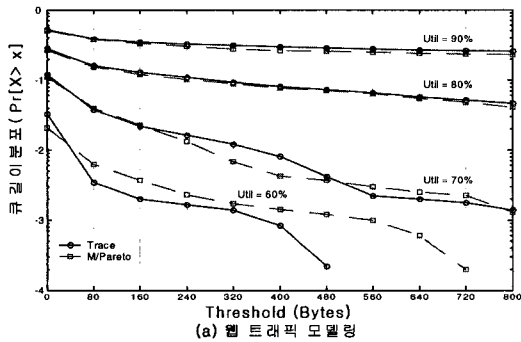


그림 7. SSQ의 이용률에 따른 M/Pareto 모델의 큐 길이 분포

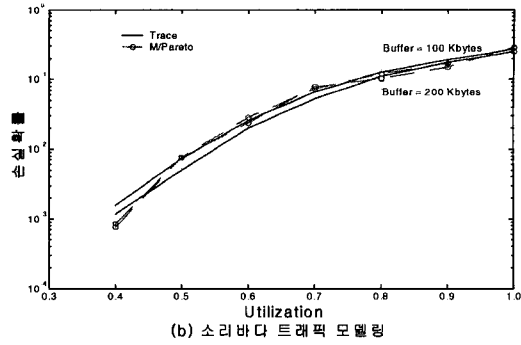
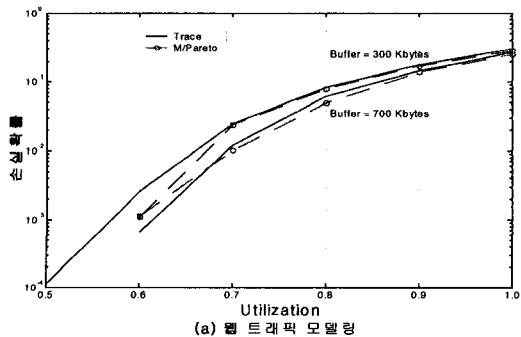


그림 8. SSQ의 버퍼길이에 따른 M/Pareto 모델의 손실확률

고 실제 손실확률은 크게 차이가 발생하지 않았다. M/Pareto 모델은 웹 트래픽의 경우 매우 유사한 결과를 얻을 수 있었으며, 소리바다 트래픽 모델링 결과를 살펴보면 버퍼 길이를 변경하여도 실제 발생된 손실은 유사하게 나왔다. 일반적으로 서버의 이용률이 높아질수록 M/Pareto 모델의 손실확률은 실제 트래픽의 손실확률과 매우 유사하였다. 이처럼 M/Pareto 모델은 측정된 트래픽을 큰 제약 없이 잘 묘사하고 있음을 관찰할 수 있었다.

2. FGN 트래픽 모델링

실제 관측된 트래픽은 정확히 자기 유사하지 않기 때문에 정확히 자기 유사한 FBM보다 FGN 모델을 이용하여 트래픽을 모델링하였다. FGN 트래픽 모델을 이용하여 P2P 트래픽을 나타내기 위해 우선 FGN 함수의 발생이 필요하다. FGN 함수는 FBM 증가함수를 통해서 얻을 수 있기 때문에 본 논문에서는 FBM을 먼저 얻고 이 함수의 증가함수를 발생시키는 간접적인 방법을 이용하여 FGN을 얻었다^[13]. 이 결과는 증가함수이기 때문에 확률 변수가 음의 값을 가질 수 있다. 트래픽의 양을 표현하는 값이 0이하의 값을 가질 수 없기 때문에 0의 값을 갖도록 하였다^[14]. 본 논문에서 사용한 FGN 모델의 파라미터는 표 1의 트래픽 측정 결과에서 얻을 수 있다. 표 3은 FGN 모델의 파라미터를 나타내고 있다.

표 3. FGN 트래픽 모델 파라미터

	Mean(<i>m</i>)	IDC(<i>a</i>)	Hurst(<i>H</i>)
Web	75.110	8.441	0.835
Soribada	19.463	7.172	0.914

FGN 모델을 이용하여 발생된 트래픽은 실제 트래픽의 평균과 5%이상의 차이를 보이는 경우가 발생하였다. 이러한 원인은 증가함수를 이용하여 트래픽을 나타내기 때문에 보이는 현상이다. 그림 9에서 확인할 수 있듯이, FGN 모델의 큐 분포 곡선은 M/Pareto 모델에 비해 SSQ의 이용률이 낮아질수록 부정확하였다. 그러나 Hurst 파라미터가 큰 소리바다 트래픽의 경우 FGN 모델은 이용률에 따른 SSQ의 큐 분포 곡선을 웹 트래픽 보다 잘 나타내고 있었다. 이 결과는 FGN 모델이 트래픽을 발생시키는 소스의 특징을 고려하지 않고 단순히 수학적으로 LRD 특징을 나타낼 수 있는 한 방법이기 때문이다.

그림 10은 FGN 모델의 SSQ의 버퍼 크기에 따른 손실확률을 보여주고 있다. 이 결과는 소리바다와 웹 트래픽 모두 M/Pareto 모델에 비해 부정확하였다. 그러나 M/Pareto 모델에서와 같이 SSQ의 이용률이 높아질수록 정확성이 높아지는 결과를 확인할 수 있다. 또한 소리바다 트래픽을 모델링 한 결과는 웹 트래픽보다 더 정확할 뿐만 아니라 SSQ의 이용률이 70%이상에서는 실제 트래픽과 매우 유사한 손실확률을 보여주고 있다.

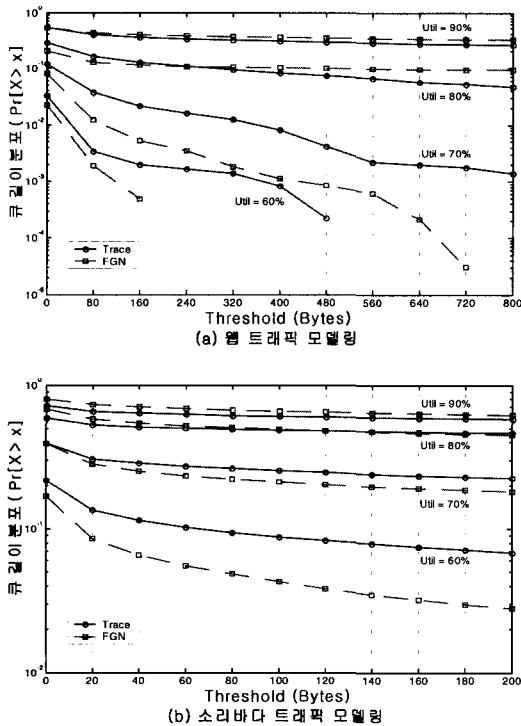


그림 9. SSQ의 이용률에 따른 FGN 모델의 큐 길이 분포

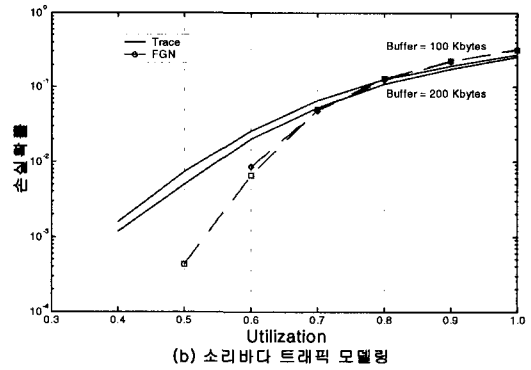
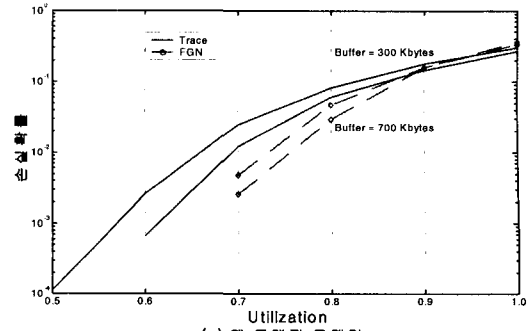


그림 10. SSQ의 버퍼길이에 따른 FGN 모델의 손실확률

지금까지 본 논문에서는 M/Pareto, FGN 모델을 이용하여 P2P 애플리케이션인 소리바다 트래픽과 웹 트래픽을 모델링하였다. 또한 모델의 정확성 검증에 위해 네트워크에서의 트래픽 특성인 큐 길이 분포와 손실확률을 이용하였다. M/Pareto 모델은 네트워크에서의 트래픽의 큐 길이 분포와 손실확률을 잘 묘사하고 있었으며, FGN 모델은 M/Pareto 모델보다 정확성이 떨어지나 제한된 범위에서는 잘 묘사하고 있었다.

지금까지 이러한 시뮬레이션 결과, 우리가 유도한 트래픽 모델은 측정된 P2P 트래픽을 정확히 묘사하고 있어, 네트워크에서 발생하는 P2P 트래픽의 지연과 손실을 예측하기 위한 입력 트래픽으로 사용될 수 있음을 의미한다.

VII. 결론

본 논문에서는 P2P 트래픽의 특성을 분석하기 위해 캠퍼스 트래픽을 측정하였다. 측정된 트래픽의 분석 결과, P2P 트래픽은 웹 트래픽보다 Hurst 파라미터가 더 높게 나타나 P2P 트래픽이 더 버스트함을 알 수 있었다. 전체 트래픽의 약 14%가 P2P

트래픽이었으나, 포트 번호를 동적으로 할당할 수 있는 P2P 애플리케이션들은 포함되지 않은 결과이다. 이러한 분류되지 못한 P2P 트래픽을 포함한다면, P2P 트래픽은 네트워크 대역폭의 더욱 많은 부분을 차지할 것이다.

본 논문에서 얻은 P2P 트래픽 모델링 결과, M/Pareto 모델과 FGN 모델은 네트워크에서 P2P 트래픽의 특성을 매우 잘 표현할 수 있으며, 측정된 P2P 트래픽에 의한 네트워크 성능을 평가하기 위한 입력 트래픽으로 사용할 수 있음을 보였다. P2P 트래픽은 매우 버스트하고 미래엔 트래픽의 대부분을 차지할 것으로 예상되어 네트워크 성능에 많은 영향을 줄 것으로 예측되기 때문에, 향후 P2P 트래픽에 대한 많은 연구가 이루어져야 할 것이다.

참 고 문 헌

- [1] Kevin Thampson, Gregory J. Miller, Rick Wilder, "Wide-Area Internet Traffic Patterns and Characteristics," *IEEE Network*, vol. 11(6), pp. 10-23, Nov-Dec 1997.
- [2] William Stalling, "High-Speed Networks: TCP/IP and ATM Design Principle," Prentice Hall, pp. 182-207, 1998.
- [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. Taquq, W. Willinger, D. V. Wilson, "On the self-similar nature of Ethernet traffic (extension version)," *IEEE/ACM Transaction on Networking*, vol. 2, no. 1, pp. 1-15, Feb 1994.
- [5] Mark E. Corvella, Azer Bestavros, "Self-similarity in World Wide Web Traffic Evidence and Possible Causes," *IEEE/ACM Transaction on Networking*, pp. 835-846, Dec 1997.
- [6] Kevin Zafer Sahinoglu, Sirin Tekinay, "Self-Similar Traffic and Network Performance," *IEEE Communications Magazine*, vol. 37, pp. 48-52, Jan 1999.
- [7] 이동철, 김기문, 김동일, " Self-Similar 특성을 이용한 데이터 트래픽 특성에 관한 연구," *한국해양정보통신학회 춘계종합학술대회*, pp. 173-178, 05. 2001.
- [8] 김동일, "자기 유사성을 갖는 World Wide Web 트래픽 소스 모델링에 관한 연구," *한국해양정보통신학회 춘계종합학술대회*, pp. 104-107, 05. 2002.
- [9] O. Rose, "Estimation of the Hurst Parameter of Long-Range Dependent Time Series," *Research Report No. 137*, Feb 1996.
- [10] Abdelanser Adas, "Traffic Models in Boradband Networks," *IEEE Communications Magazine*, pp. 82-89, July 1997.
- [11] I. Norros, "A storage model with self-similar input," *Queueing systems*, vol. 16, pp. 387-396, 1994.
- [12] I. Norros, "Studies ofr a Model for Connectionless traffic, based on Fractional Brownian Motion," Conference on Applied Probability in Engineering , *Computer and Communication Science*, Paris, June 1993.
- [13] Wing-cheong Lau, Ashok Erramilli, Jonathan L. Wang, Walter Willinger, "Self-similar traffic Generation: The Random Midpoint Displacement Algorithm and Its Properties," *IEEE ICC'95*, vol. 1, pp. 466-472, 1995.
- [14] Stefano Giordano, Michele Pagano, Riccardo Pannocchia, Franco Russo, " A new call admission control scheme based on the self similar nature of multimedia traffic," *IEEE ICC'96*, vol. 3, pp. 23-27, June 1996.
- [15] P. Pruthi, A. Erramilli, "Heavy-tailed ON/OFF Source Behaviour and Self-Similar Traffic," *IEEE ICC'95*, pp. 445-450, June 1995.
- [16] Timothy D. Neame, Moshe Zukerman, Ronal G. Addie, "Application of the M/Pareto Process to Modeling Broadband Traffic Streams," *ICON'99*, pp. 53-58, Oct. 1999.
- [17] Ronald G. Addie, Timothy D. Neame, Moshe Zukerman, "Modeling Broadband Traffic Streams," *IEEE Gobecom'99*, pp. 387-391, Dec. 1999.
- [18] Ronald G. Addie, Moshe Zukerman, Timothy D. Neame, "Fractal Traffic: Measurements, Modeling and Performance Evaluation," *IEEE ICC'95*, vol. 3, pp. 471-476, April 1998.

주 성 돈(Sung-don Joo)

준회원



2003년 2월 : 아주대학교

전자공학과 졸업.

2003년 3월~현재 : 아주대학교

대학원 전자공학과

석사과정.

<주관심 분야> Traffic engineering, Wireless LAN

이 채 우(Chae-woo Lee)

정회원



1985년 : 서울대학교

제어계측 학사.

1988년 : 한국과학기술원

전자공학과 석사.

1995년 : University of Iowa

박사.

1985년 1월~1985년 12월 (주)금성통신 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.

<주관심 분야> 광대역 통신망, Ubiquitous
networking, Traffic Engineering.