방화벽이 존재하는 캠퍼스 망에서의 P2P 트래픽 측정 및 분석

정회원 이 영 석*

Measurement and Analysis of P2P Traffic in Campus Networks Under Firewall

Youngseok Lee* *Regular Members*

요 약

본 논문은 P2P 트래픽을 차단시키는 방화벽이 존재하는 고속의 캠퍼스 망에서 세 개의 P2P 응용 트래픽을 장기간 측정하고 분석한 결과를 제시한다. P2P 트래픽을 탐지하기 위하여 다양한 방법들이 제안되고 있지만, 현재 가장 간단하고도 비용이 저렴한 방법은 포트 번호를 이용한 방화벽이다. 이 방화벽이 설치되고 난 이후 대량의 P2P 트래픽이 줄어들 것으로 예측되었지만, 8개월간의 트래픽 측정 결과 세 개의 새로운 P2P 응용 트래픽(30% 업로드, 5.6% 다운로드 트래픽)과 포트번호를 숨기는 eDonkey P2P 응용 트래픽(6.7% 업로드 트래픽, 4% 다운로드 트래픽)으로 인하여 본 연구에서 판별한 P2P트래픽 양이 다시 증가하였다. 본 논문에서 수행한 장기간 트래픽 측정결과는 포트번호를 이용하는 트래픽 필터링 기법이 P2P 응용 탐지에 효과적이지 않다는 것을 보여주었고, 본 캠퍼스 망에서 관찰된 세가지 P2P 트래픽은 P2P응용 프로그램의 보상 체제와 고속의 캠퍼스 망 연결 등의 이유 때문에 외부로 향하는 업로드 트래픽의 양이 크다는 것을 보여주었다.

Key Words : P2P, traffic measurement, eDonkey, flow, firewall.

ABSTRACT

This paper reports on the study of P2P traffic behaviors in a high-speed campus network under a simple firewall which drops packets with default port numbers for the well-known P2P applications. Among several ways of detecting P2P traffic, the easiest method is to filter out packets with the default port number of each P2P application. After deploying the port-based firewall against P2P-traffic, it is expected that the amount of P2P traffic will be decreased. However, during the eight-month measurement period, three new commercial P2P applications have been identified and their traffic usages have reached up to 30/5.6% of the total outbound/inbound traffic volumes at the end of the measurement period. In addition, the most famous P2P application, eDonkey, has adapted and has escaped detection through port hopping. The measurement result shows that the amount of eDonkey traffic is around 6.7/4.0% of the total outbound/inbound traffic volume. From the measurement results, it is observed that the port-based firewall is not effective to limit the usage of P2P applications and that the P2P traffic is steadily growing due to not only the evolution of existing P2P applications such as port hopping but also appearances of new P2P applications.

* 충남대학교 전기정보통신공학부 컴퓨터전공 데이터네트워크 연구실(lee@cnu.ac.kr)
논문번호 : KICS2005-06-260, 접수일자 : 2005년 6월 26일
※이 논문은 2004년도 충남대학교 자체연구비의 지원과 정보통신부 및 정보통신연구진흥원의 대학 IT연구센터 지원사업의 결과로 수행된 것임

Ⅰ. 서 론

최근에 들어 P2P(Peer-to-Peer) 응용들이 인터넷 트래픽 및 사용 패턴을 급격하게 바꾸면서, 네트워크 설계, 운영 및 관리에 큰 영향을 끼치고 있다. P2P 응용들은 대개 클라이언트와 서버기능을 동시에 수행하면서 음악, 동영상과 같은 정보를 공유하면서 대량의 트래픽을 장시간 발생시킨다. 이러한 대량의 P2P 트래픽은 값비싼 인터넷 사용기관의 외부 인터넷 연결 회선의 대역폭을 점유하여 이메일, 원격 접속 및 웹과 같은 비 P2P 응용들의 성능 저하에 영향을 끼칠 수 있다. 따라서, 많은 기관과 인터넷서비스제공업체들은 과도한 트래픽을 유발하는 P2P 응용들을 차단하는 다양한 네트워크 정책을 고려하기 시작했다. 대표적으로 방화벽이나 트래픽 제어기[1]를 외부 인터넷 연결회선에 설치하여 P2P 트래픽을 탐지하고 제어하도록 하고 있다. 게다가, 인터넷서비스제공업체들은 P2P트래픽 제어를 위하여 정액제에서 종량제의 요금제도도 검토하기 시작하였다.

P2P 트래픽의 탐지를 위해서는 다양한 방법들이 제안되고 있다. P2P 응용들은 일반적으로 서버(또는 수퍼피어)에게 자신의 수신 포트 번호와 IP주소를 알려주어 다른 사용자들이 접속하게 한다. 예를 들면, eDonkey와 같은 프로그램은 4662번을 이용하여 다른 사용자들이 접속하도록 하며, 4661번의 포트번호는 수퍼피어의 디폴트 포트번호로 사용되어 클라이언트가 처음 P2P 네트워크에 접속할 때 헬로우 메시지를 전송한다[2]. 따라서, 가장 간단한 P2P 트래픽 탐지 방법은 이러한 포트번호를 이용하는 것이다. 그러나, 포트번호를 이용하여 P2P 트래픽을 차단하는 방화벽이나 네트워크 정책은 P2P 응용들로 하여금 디폴트 포트번호가 아닌 다른 포트 번호를 동적으로 사용하는 포트 호핑(port hopping)을 사용하게 하는 부작용이 발생하고 있다. 그러므로, 포트 번호를 사용하는 P2P 트래픽 탐지 및 차단 방법은 이러한 포트 호핑 또는 다른 새로운 P2P 응용의 등장과 같은 환경에서는 효과적이지 않을 수 있다.

eDonkey, Kazza, Gnutella 및 WinMX와 같은 대표적인 P2P 응용들은 지역별 또는 시대별에 따라 이용패턴의 차이가 발생한다. 예를 들어, 영어가 아닌 한글, 일본어, 중국어 검색을 지원하는 응용들이 이들 국가에서 등장하고 있고, 대표적인 P2P 응용이 되고 있다. 또한, eDonkey와 같은 대표적인 P2P 응용들은 업로드의 대역폭 및 대기자 수 등을 제한하는 설정을 사용하고 있고, 다수의 송신자로부터 전송을 받는 정책으로 인하여 오히려 다운로드 속도가 떨어지게 되었다. 게다가, 최근 들어 많은 기관에서 유명한 P2P 응용들의 디폴트 포트번호를 차단하고 있기 때문에 파일 다운로드가 원활하게 진행되지 않게 되었다. 이러한 이유들로 인하여 새로운 P2P 응용들이 등장하기 시작했다.

포트 호핑 및 새로운 P2P 응용들의 등장으로 인하여 P2P 트래픽 탐지가 더욱 어려워지고 있는 가운데, 패킷의 특징적인 시그너쳐(signature)를 이용하는 방법들이 최근에 사용되고 있다. 시그너쳐를 이용하는 P2P 트래픽 탐지 및 차단 방법은 포트번호 기반의 탐지 방법에 비해 숨겨진 많은 P2P 트래픽을 탐지할 수 있기 때문에 효과적으로 알려져 있다[3]. 하지만, IPsec과 같이 패킷을 암호화하여 전송하는 방법들이 표준화되고 있기 때문에 시그너쳐 기반 탐지 기법도 제한적일 수 밖에 없다. 또한, 모든 새로운 P2P 응용에 대한 시그너쳐를 찾는 작업은 P2P 응용 프로토콜을 역추적해야하기 때문에 굉장히 어렵다. 최근에 들어서는 TCP/UDP 프로토콜을 동시에 사용하면서 서버와 같은 연결 패턴을 보여주는 전송계층의 정보를 이용하는 기법[4]이 제안되어, 시그너쳐 기법과 비슷한 탐지율을 보여준다는 결과가 제시된 바 있다.

일반적으로 캠퍼스, 기관, 및 기업의 네트워크는 방화벽을 두어 비정상 트래픽을 탐지하여 차단시키고 있다. 방화벽의 접근차단리스트는 포트번호와 IP 주소에 기반을 두고 있다. 저자가 소속되어있는 캠퍼스 네트워크의 방화벽도 P2P 트래픽을 차단하기 위하여 대표적인 포트번호를 이용하고 있다. 방화벽 설치 이후 대표적인 P2P 응용들의 트래픽은 사라진 것처럼 보이지만, 알려지지 않은 응용들의 트래픽은 여전히 높게 나타나고 있으며, 이러한 트래픽은 숨겨진 P2P 응용들에 의한 것으로 추정된다.

따라서, 본 논문에서는 포트기반 방화벽이 설정되어 있는 캠퍼스 네트워크에서의 P2P 트래픽 측정을 통한 분석을 시도하였다. 특히, 다음과 같은 두 가지 목표를 추구하도록 하였다.

- 새로운 P2P 응용들의 등장으로 인한 트래픽의 변화: 기존의 P2P 응용들은 방화벽으로 인하여 사용에 어려움이 있고, 한국의 사용자들이 사용하는 독자적인P2P 응용들이 속속 등장하고 있기 때문에, 기존에 알려지지 않은 세 가지 새로운 P2P 응용들의 트래픽을 측정하고 분석하도록 한다.

- 방화벽이 존재할 때 기존 P2P 응용 트래픽의 변화: 방화벽에서 접근차단리스트에 등록된 포트번호를 디폴트로 이용하는 P2P 응용들은 포트 호핑과 같은 방법을 이용하여 여전히 사용되고 있다. 대표적으로 eDonkey 응용 트래픽을 측정하고 분석하였다.

본 연구에서 수행한 측정 결과, 측정 기간 동안 국내에서 사용되고 있는 P2P 응용들 중에서 세가지 새로운 P2P응용들에 의해서 캠퍼스망에서 나가는 전체 트래픽의 30%정도, 캠퍼스 망으로 들어오는 전체 트래픽의 5.6% 정도 차지한다는 것을 관측하였다. 또한, 포트기반 방화벽아래에서도 eDonkey 트래픽이 여전히 6.7/4.0% 정도의 점유율이 각 방향 관측되었다.

본 논문의 구성은 다음과 같다. II절에서는 관련 연구결과에 대해서 설명하고, III절에서 트래픽 측정 환경에 대해서 기술한다. IV절에서는 측정결과에 대해서 토의하고, V절에서 결론을 맺도록 한다.

Ⅱ. 관련 연구

[3]에서는 잘 알려진 포트번호만을 이용하는 트래픽 측정 방법은 P2P트래픽의 정확한 통계를 구할 수 없다는 결과를 제시하였고, 시그너쳐 기반 트래픽 측정을 수행한 결과 전체적으로 P2P 트래픽은 지속적으로 증가하고 있다는 결과를 보였다. 광대역 인터넷 가입자 트래픽을 포함하는 인터넷 백본에서의 트래픽 측정 결과[5]는 80-20 규칙(80%의 트래픽이 20%의 사용자로부터 유발된다.)이 적용될 수 있음을 보였다. 특히, Kazza 응용에 대한 포트번호를 차단한 결과 다른 포트번호를 사용하여 Kazza 트래픽 차단 효과가 없다는 것을 보였다. [6]에서는 FastTrack, Gnutella, 및 DirectConnect 트래픽을 인터넷 백본에서 측정한 결과를 제시하였는데, 10%의 IP주소가 전체 트래픽의 99%를 차지한다고 보고하였고, P2P트래픽의 점유율이 지속적으로 동일하게 관측되기 때문에 트래픽 엔지니어링에 효율적으로 사용될 수 있다고 주장하였다. [7]에서는 시그너쳐 기반의 P2P 트래픽 탐지기법을 온라인으로 적용할 수 있는 방법을 제안하였고, Kazza 프로토콜에 대해서는 포트기반의 방법보다 3배 이상의 트래픽을 관측하였다. [8]에서는 일본에서 대표적으로 사용되고 있는 Winny라는 P2P 응용을 탐지하는 기법을 제안하였다. ADSL네트워크에서의 P2P 트래픽 경향은 [9]에서 분석되었는데, 60%정도의 트래픽이 P2P 응용에 의한 것으로 관측되었다.

Ⅲ. 측정 환경 및 측정 방법

그림 1과 같이 저자가 소속되어있는 캠퍼스망은 세 개의 인터넷서비스제공업자(ISP)들과 각각 OC-3, OC-3, T3링크로 외부 상용 인터넷 망과 연결되어 있다. 경계 라우터는 캠퍼스 백본 라우터와 1G 이더넷으로 연결되어 있으며, 이 지점에 방화벽이 설치되어 있다. 본 연구를 위하여 트래픽 측정 시스템을 방화벽과 경계 라우터 사이에 위치하도록 하였으며, 광 분배기를 두어 입/출력 각 방향 트래픽을 트래픽 측정 시스템에 전송하도록 하였다. 트래픽 측정 시스템은 두 개의 광 기가 이더넷 카드를 이용하여 각 방향의 트래픽을 수집하도록 하였다. 트래픽 측정 시스템은 두 개의 Xeon CPU와 1GB 메모리를 사용하고 Linux 2.4 커널을 운영체제로 사용한다. 트래픽 측정 시스템에서는 패킷 또는 플로우 단위로 트래픽을 수집할 수 있는데, 본 연구에서는 플로우 단위의 트래픽 측정과 분석을 수행하였다. 왜냐하면, 패킷 단위의 트래픽 측정방법은 장기간 트래픽 측정으로 인하여 대용량의 저장 장치 요구사항과 대용량의 트래픽 분석의 어려움이 있는데 비하여, 플로우 단위의 트래픽 측정방법은 장기간으로 트래픽 측정할 경우에도 저장 장치의 요구사항이 대폭 감소하기 때문이다. Cisco와 Juniper사의 라우터에서는 플로우 단위의 트래픽 측정이 가능하지만, 본 캠퍼스에서 사용하고 있는 Nortel사의 라우터에서는 플로우 단위 트래픽 측정이 불가능하기 때문에 트래픽 측정 시스템에서 패킷을 수집하여 플로우로 가공하여 트래픽을 측정하도록 하였다. 플로우로 가공하는 툴은 nProbe[10]를 사용하였는데, 이는 광 기가 이더넷에 도착하는 패킷들을 Cisco사의 NetFlow[11] v5 형태의 플로우로 만들어 UDP 전송을 한다. 동일한 트래픽 측정 시스템에서 플로우 정보를 담고 있는 UDP패킷을 수신하여 저장하도록 하였다. 플로우는 IP와 TCP/UDP 헤더 필드들 중 송신자 IP 주소, 수신자 IP 주소, 송신 포트, 수신 포트, 프로토콜 타입으로 정의된다. 플로우 종료 타이머를 이용하여 각 플로우가 저장되는 플로우 해쉬 테이블의 각 플로우 엔트리를 종료시키고, 내보내도록 하였다. 즉, 설정된 플로우 종료 시간 내에 새로운 패킷이 도착하지 않으면 플로우 엔트리는 자동적으로 종료되어 플로우로 내보내어진다. 측

정 시 설정된 플로우 종료 시간값은 30초이다. 한편, 너무 오랫동안 플로우 해쉬 테이블에 저장되어 있는 플로우를 내보내기 위해서 플로우 휴지 시간값을 설정하는데, 디폴트로 30분을 이용하였다.고속의 상용 라우터에서 지원되는 NetFlow는 1GE과 같은 환경에서는 1/10 정도의 패킷 샘플링 값을 이용하고 있지만, 본 연구에서는 패킷 샘플링을 이용하지 않았다. nProbe는 pcaplib[12]을 이용하여 구현되었기 때문에 작은 크기의 패킷들이 매우 빠른 속도로 도착하게 될 경우 패킷을 측정하지 못할 수도 있다. 하지만, 본 연구에서는 이러한 작은 크기의 패킷들만 존재하는 것이 아니고 다양한 크기의 패킷들이 존재하고, 각 방향 최대 345Mbps정도의 트래픽이 전송되고 있기 때문에 패킷 손실이 장기간의 P2P 트래픽 분석 결과에 큰 영향을 끼치지 않는다고 가정하였다. 생성된 NetFlow v5 패킷들은 flow-tools[13]을 이용하여 저장하고 분석하였다. 한편, 숨겨진 eDonkey 트래픽을 측정하고 분석하기 위하여 ETRI에서 개발한 TrafView[14] 도구를 사용하여 시그너쳐 기반의 트래픽 측정 및 분석을 수행하였다.

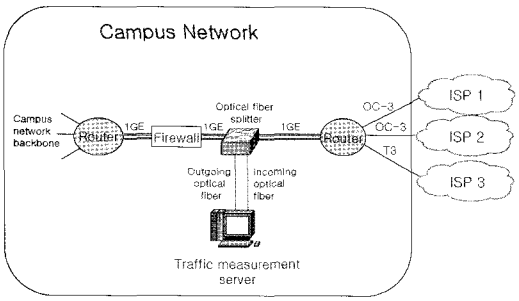


그림 1. 트래픽 측정 환경

Ⅳ. 측정 및 분석 결과

본 절에서는 2004년 10월 1일에서부터 2005년 5월 31일까지 8개월간 플로우 기반 트래픽 측정 분석 결과와 1개월간 측정한 시그너쳐 기반의 eDoneky 트래픽 분석 결과에 대해서 기술하도록 한다.

4.1 기초적인 트래픽 통계

우선 전체 트래픽에 대한 경향을 파악하기 위하여 캠퍼스 망에서 나가는(outbound) 전체 트래픽과 들어오는(inbound) 전체 트래픽 및 액티브(active) 호스트1)들에 관한 1일 통계에 대해서 분석하였다.

1) 액티브 호스트는 인터넷에 연결되어 있는 호스트로 IP 주소로 식별하였다. 본 연구에서는 하루에 적어도1MB이상의 트래픽을 전송하거나 수신한 호스트를 액티브 호스트라고 정의하였다. 최근에는 인터넷 웜에 감염되거나 다양한 스캐닝 도구로 인하여 네트워크를 스캐닝하는 트래픽이 상시적으로 관찰되고 있다. 따라서, 본 캠퍼스 망이 사용하는 B 클래스의 IP주소 프레픽스(prefix)는 총 65,534개의 IP주소가 외부로부터 트래픽을 수신하게 된다. 따라서, 이러한 스캐닝 트래픽을 줄이기 위하여 1일 평균 트래픽 양의 임계치를 두었는데, 이결과 본 캠퍼스 망에서 나가는 트래픽의 송신자 수와 비슷한 값을 생성하는 1MB값을 이용하였다.

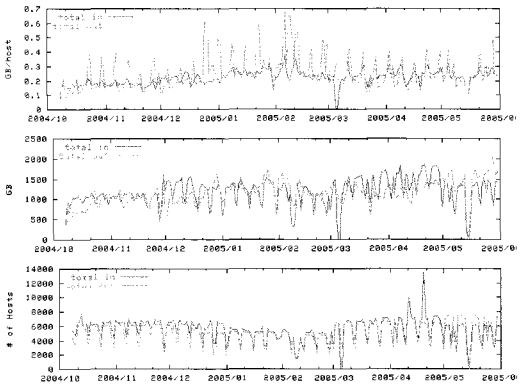


그림 2. 전체 입/출력 트래픽: 1일 바이트 총합, 1일 호스트 수, 및 호스트 당 1일 바이트 양 (2004.10.1 - 2005.5.31)

그림 2에서와 같이 인터넷에 연결되어 활동하는 호스트의 수는 1일 평균 6,212로 관측되었다. 그림 2에서와 같이 주중/주말, 학기 중/방학과 같은 주기성이 관측되고 있다. 저자가 소속되어있는 캠퍼스는 B 클래스의 주소를 사용하고 있는데, 액티브 호스트의 IP 주소 사용률은 9.5%이다. 전체적으로 트래픽의 양은 증가하고 있는 추세를 보이고 있으며, 1일 평균 1개의 호스트는 0.2GB를 전송하거나 수신하고 있다.

80포트번호를 사용하는 웹 트래픽은 그림 3과 같이 주간 및 계절별 패턴에 있어 전체 트래픽과 유사하게 나타난다. 캠퍼스로 유입되는 80 포트의 트래픽이 외부로 나가는 트래픽보다 훨씬 큰 이유는 캠퍼스에 존재하는 웹 서버가 많지 않고, 대부분의 캠퍼스 망의 사용자들이 웹 브라우징을 통하여 트래픽을 전송받기 때문이다. 1개의 호스트는 1일 평균 190MB의 웹 트래픽을 80포트를이용하여 전송받고 있다. 캠퍼스에 존재하는 80포트를 이용하는 웹 서버는 평균 747개로 관찰되는데 1일 평균 270MB를 전송하고 있다. HTTP 트래픽 이외에 잘 알려진 포트번호를 사용하는 인터넷 응용들이 존재하지만, 대부분 트래픽의 양이 많지 않다.

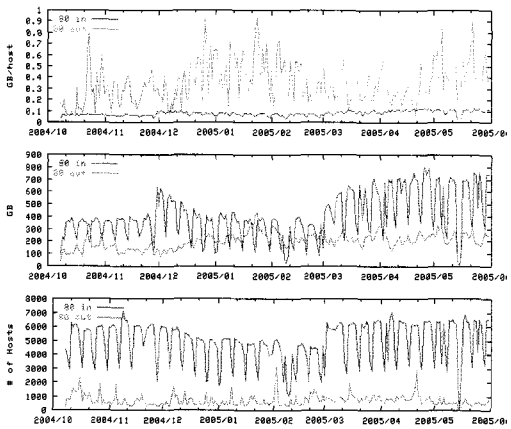


그림 3. 80포트 사용 웹 입/출력 트래픽: 1일 바이트 총합, 1일 호스트 수, 및 호스트 당 1일 바이트 양 (2004.10.1 - 2005.5.31)

표 1. 잘 알려진 포트별 1일 트래픽 통계(바이트 퍼센트)

	2004.10.6		2005.4.28	
	Out	In	Out	In
HTTP	21.3	35.2	22.3	40.4
HTTPS	0.1	1.7	0.1	0.6
FTP	3.3	1.4	2.3	1.2
NNTP	0.0	1.2	0.0	0.0
SMTP	0.2	0.1	0.2	0.2
IRC	0.3	0.1	0.0	0.0
SSH	0.1	0.0	0.1	0.2
TELNET	0.2	0.2	0.0	0.0
Real streaming	0.1	2.3	0.1	2.5
Unknown	74.2	57.8	74.7	55.0

한편, 표 1은 2004년도와 2005년도의 1일 트래픽 통계를 잘 알려진 포트 번호를 이용하여 분석한 결과를 보여주고 있다. 80 포트를 이용하는 HTTP 입력(In) 트래픽이 2005년도에 40.4%로 증가하였지만, 캠퍼스 출력(Out) 웹 트래픽은 22.3%로 2004년도와 비교해서 비슷하다. 잘 알려지지 않은 포트를 이용하는 트래픽은 캠퍼스에서 외부로 향하는 전체 트래픽 중74%를 차지하고 있다. 따라서, 본 연구에서는 알려지지 않은 포트를 사용하는 트래픽 중 트래픽 점유율이 큰 포트번호를 상시적으로 관찰하고 분류하여 많은 트래픽이 지속적으로 관찰되는 포트에 대한 분석을 시도하였다. 포트번호를 이용하여 새로운 P2P응용을 분류하고, 직접 P2P 응용을 실행시켜 포트 번호와 트래픽 특성을 비교 확인하였지만, 여전히 디폴트 번호가 아닌 다른 포트를 사용하는 동일한 P2P 응용 트래픽이 존재한다. 본 연구에서 새로운 P2P 응용들에 관한 트래픽 측정 분석에서는 디폴트 포트만을 이용하였다.

4.2 세 가지 새로운 P2P 응용 트래픽

포트기반 방화벽으로 인하여 잘 알려진 포트를 사용하는 P2P 트래픽은 차단이 되었지만, 본 연구의 측정결과 새로운 P2P 응용들에 의한 트래픽이 꾸준하게 증가하고 있는 것으로 관찰되었다. 포트별 1일 바이트 총합을 정렬한 결과, 규칙적으로 Top 10에 나타나는 세 개의 포트번호 7305, 19101 및 7878를 이용하여 본 캠퍼스에서 사용되고 있는 P2P 응용들 중의 일부분을 조사하였다. 이러한 포트번호를 이용하는 응용들은 각각 "PeePop[15]", "Clubbox[16]", 및 "Enppy[17]"로 밝혀졌다. 세 개의 P2P 응용들은 국내에서만 사용되고 있는데, 한글을 지원하고 최근에 인기를 얻고 있는 P2P 응용들로서 모두 상업화되어 서비스되고 있다. 즉, PeePop은 사용자가 가진 예산 범위 내에서 파일을 다운받을 수 있는데, 이러한 예산은 직접 구매할 수도 있고, 자신의 파일을 다른 사용자가 다운받게 되면 전송된 트래픽 양만큼 생성되는 "보상(reward)" 제도를 운영하고 있다. Clubbox와 Enppy도 유사한 상업화된 P2P 응용들이다. 이러한 응용들은 eDonkey와 같은 유사한 방식으로 동작하는데 자신의 포트번호와 IP주소를 서버 또는 수퍼피어에게 알려 다른 사용자들이 접속할 수 있게 한다. PeePop은 여러 개의 송신자로부터 다운받을 수도 있고, 하나의 송신자로부터 다운받을 수 있도록 설정할 수 있다. 이러한 보상제도는 P2P 사용자들로 하여금 자신의 호스트를 계속하여 P2P 네트워크에 참가하도록 유도하고 있다.

그림 4는 2004년 10월부터 2005년 5월까지 측정된 7305 포트 트래픽 통계를 보여주고 있다. 트래픽 측정 초기인 2004년 10월에는 7305포트 트래픽이 많이 관찰되지 않고 있지만, 2005년 5월에는 30개 이상의 호스트가 350GB를 외부로 전송하고 있다. 주목할 점은 캠퍼스 망으로 향하는 트래픽보다 외부로 나가는 트래픽의 양이 월등하게 많다는 것이다. 이러한 현상은 상업화된 P2P 응용의 보상제도로 인하여 지속적으로 P2P 응용을 지속적으로 실행시켜두고 있는데, 캠퍼스 망 호스트들이 대개 100Mbps 이더넷으로 연결되어 컨텐츠 제공 서버팜의 역할을 수행하고 있기 때문이다. 1일 평균 1개의 호스트가 외부로 전송하는 트래픽 양은 5.8GB이고, 다운받는 트래픽의 평균은 1.4GB로 나타난다. 호스트 수와 트래픽양은 측정 기간 동안 선형적으로 증가하고 있는 추세를 보여주고 있는데, 10일에 1개의 새로운 호스트가 캠퍼스 망에 나타나는 기울기를 보여주고 있다.

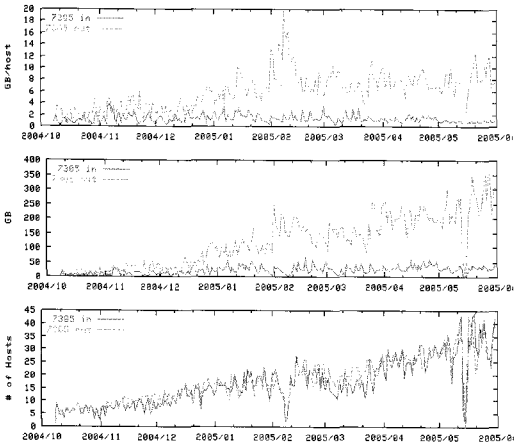


그림 4. 7305포트 사용 P2P 입/출력 트래픽: 1일 바이트 총합, 1일 호스트 수, 및 호스트 당 1일 바이트 양 (2004.10.1 - 2005.5.31)

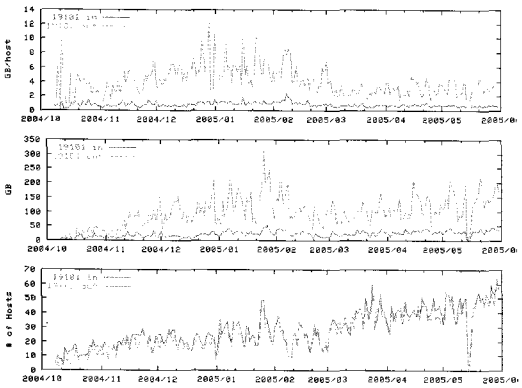


그림 5. 19101포트 사용 P2P 입/출력 트래픽: 1일 바이트 총합, 1일 호스트 수, 및 호스트 당 1일 바이트 양 (2004.10.1 - 2005.5.31)

19101포트를 사용하는 P2P 트래픽의 통계는 그림 5와 같다. 19101 포트 P2P 트래픽 역시 측정 초기 시점부터 꾸준하게 증가하고 있으며, 2005년 5월 현재 60개 정도의 호스트들이 사용하고 있음을 알 수 있다. 1개의 호스트가 1일 평균 전송받는 트래픽 양은 0.85GB이지만, 외부로 전송하는 트래픽 양은 4.2GB이다. 7305 포트 트래픽 패턴과 동일하게 외부로 전송하는 트래픽 양이 5배 정도 많이 발생한다. 전체적인 트래픽 양의 증감 추세는 7305 포트 트래픽과는 다르게 나타난다. 즉, 외부로 향하는 트래픽 양은 지속적으로 증가하지 않고, 2005년 2월을 기점으로 감소한 다음 다시 증가하고 있는 추세를 보여준다. 이는 19101포트를 사용하는 외부 P2P 사용자 수와 밀접한 관계에 있는 것으로 외부 사용자 수가 이 기간 동안 감소하였기 때문이다.

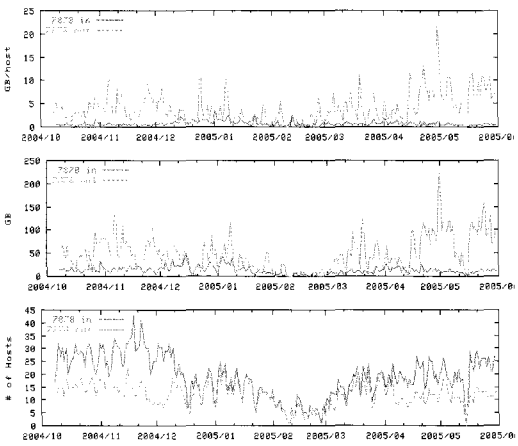


그림 6. 7878포트 사용 P2P 입/출력 트래픽: 1일 바이트 총합, 1일 호스트 수, 및 호스트 당 1일 바이트 양 (2004.10.1 - 2005.5.31)

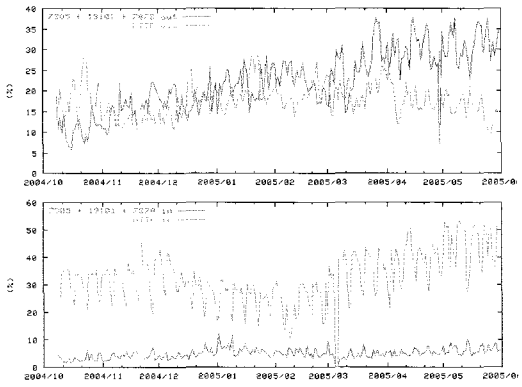


그림 7. 전체 트래픽과 비교한 7305, 19101, 및 7878 포트 사용 P2P 트래픽과 80 포트 웹 트래픽 점유율의 변화 (2004.10.1 - 2005.5.31)

세 번째로 탐지된 새로운 P2P 응용은 7878포트를 사용하는 "Enppy"라는 것으로 그림 6에서 자세한 트래픽 통계를 보여주고 있다. 1개의 호스트에 대한 1일 평균 외부로 향하는 트래픽은 3.64GB이고, 전송받는 트래픽은 0.75GB이다. 7878 포트를 사용하는 호스트의 수는 트래픽 측정 초기에는 많이 관찰되었지만, 2004년 12월 이후 감소하는 추세이다가 2005년 3월 이후 조금씩 증가하고 있으며, 2005년 5월 현재 30개의 호스트들이 관찰되고 있다.

전체 트래픽에 대해서 7305, 19101, 및 7878 포트들을 사용하는 P2P 응용 트래픽의 점유율을 그림 7에서 보여주고 있다. 80포트를 사용하는 웹 트래픽과 비교할 때, 새로운 P2P 응용에 의한 트래픽 점유율은 외부로 향하는 링크에서는 35%를 초과하고 있다. 특히, 웹 트래픽보다 2배 이상 많은 양의 트

래픽이 외부로 전송되고 있다. 하지만, 캠퍼스 망으로 전송되는 P2P 트래픽은 웹 트래픽에 비해서 상대적으로 적게 관찰된다. 이는 많은 수(6,212)의 캠퍼스 호스트들이 웹 클라이언트로 동작하여 웹 데이터를 전송받는데 비하여, 캠퍼스 망에 위치한 적은 수의 P2P 호스트들(100여 개)이 전송받는 트래픽 양이 많지 않기 때문이다.

트래픽 측정 결과 캠퍼스에서 외부로 향하는 많은 양의 트래픽이 새로운 P2P 응용에 의해서 발생하고 있다는 것을 알 수 있었다. 특히, 7305 포트를 사용하는 트래픽은 꾸준하게 늘고 있다. 이렇게 증가하는 것은 단순히 캠퍼스의 P2P 응용 사용자 수의 증가에 의한 것 보다는 외부에 존재하는 P2P 사용자들의 증가와 밀접한 상관관계를 가지고 있다. 그림 8에서는 캠퍼스 망 내부에 존재하는 P2P 서버들에 접속하여 트래픽을 전송받은 외부 사용자들의 수를 분석한 결과이다. 7305포트를 이용하는 외부 호스트들의 수는 지속적으로 증가하고 있으며 2005년 5월 현재 2,000여 호스트들이 접속하였고, 이는 충남대의 외부 트래픽의 증가를 유발시켰다. 2005년 5월 현재 3,000여 개의 호스트들이 19101포트를 이용하여 충남대 P2P 서버에 접속하였다. 외부 19101 포트 호스트 수의 증가와 충남대의 19101 포트 외부 트래픽의 증가추세는 밀접한 상관관계를 보여주고 있다. 7878 포트의 호스트 수는 1,000여 개로 관측되고 있다.

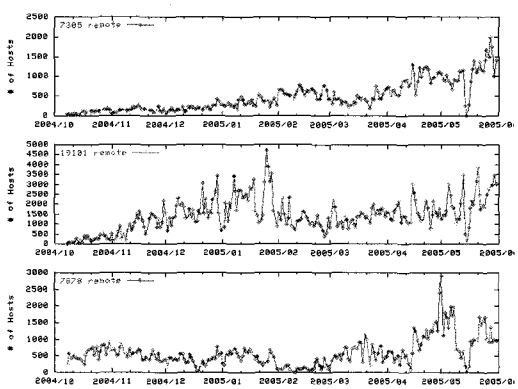


그림 8. 외부 네트워크에서의 7305, 19101, 및 7878 포트 이용 P2P 사용자 수 (2004.10.1 - 2005.5.31)

4.3 숨겨진 eDonkey 응용 트래픽

포트기반 방화벽 사용으로 인하여 대표적인 P2P 응용인 eDonkey 트래픽은 포트 4661 또는 4662에서는 거의 나타나지 않게 되었다. 하지만, 디폴트 포트 번호를 다른 것으로 바꾸어서 사용하는 숨겨진 eDonkey 트래픽은 여전히 그림 9과 같이 나타나고 있다. 패킷 페이로드의 첫 번째 바이트에 0xE3 또는 0xC5가 나오는 것이 eDonkey의 대표적인 시그너쳐라고 알려져 있다[6]. 이러한 시그너쳐 기반의 트래픽 측정을 수행하기 위하여 TrafView라는 도구를 동일한 트래픽 측정 시스템에 설치하여 2005년 5월 1달간 운영하였다. 캠퍼스 망에서 외부로 향하는 eDonkey 트래픽은 6.74% 정도이고 1일 평균 100GB가 관찰되었다. 외부에서 캠퍼스 망으로 전송되는 eDonkey 트래픽은 전체 트래픽의 3.95% 정도로 1일 평균 60GB정도이다. 포트기반 방화벽에도 불구하고, 완전히 사라지 않고 꾸준하게 사용되고 있다는 것을 보여주고 있다.

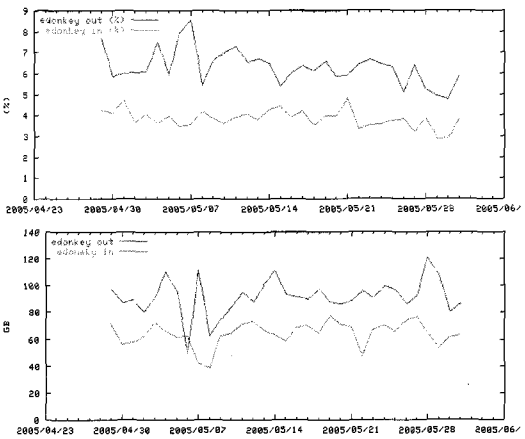


그림 9. 숨겨진 eDonkey 응용 입/출력 트래픽 (2005.4.28 - 2005.5.31)

V. 결 론

본 논문에서는 방화벽이 존재하는 고속의 캠퍼스 망에서 장기간 트래픽을 측정하고 분석하여 본 캠퍼스에서 관찰된 세 가지 새로운 P2P 응용 트래픽과 숨겨진 eDonkey 응용 트래픽의 분석 결과를 기술하였다. 포트기반 방화벽 아래에서 8개월간 측정 결과를 바탕으로, 한국에서 독특하게 사용되고 있는 세 개의 새로운 P2P응용 트래픽이 캠퍼스 망의 외부로 향하는 트래픽의 30%이상을 차지하고 있다는 것을 보였다. 또한, 잘 알려진 eDonkey응용도 포트 번호를 바꾸어서 여전히 사용되고 있음을 알 수 있고, 6% 이상의 외부로 향하는 트래픽 점유율을 보이고 있다. 고속의 캠퍼스 망에서는 일반적인 광대역 인터넷 가입자 망과는 달리 가입자의 업링크 쪽의 트래픽의 많은 양이 P2P응용들에 의해서 생성되

고 있다. 이는 P2P를 사용하는 호스트들이 고속의 캠퍼스 망에 연결되어 전체 P2P 네트워크에서 컨텐츠 제공 서버 팜과 같은 역할을 수행하고 있기 때문이다. 또한, 많은 P2P 응용들이 파일 업로드하는 만큼 다운받도록 하기 위한 제도를 도입하기 때문에 캠퍼스 망과 같은 경우 P2P 호스트들이 계속적으로 실행되고 있기 때문이다. 포트기반의 방화벽이 간단하고 사용이 편하지만, P2P응용의 다른 포트 번호 사용으로 인해 정확한 탐지가 불가능하게 된다는 것을 보였다. 시그너쳐 기반의 트래픽 측정을 통하여 여전히 사용되고 있는 P2P응용 트래픽을 관찰할 수 있었지만, 1Gbps 이상의 고속 링크에서 측정 및 IPsec과 같은 암호화 등으로 적용에 어려움이 있다. 따라서, 향후 보다 정확한 P2P 트래픽 측정을 위한 방법이 필요하다.

참 고 문 헌

[1] Packeteer, http://www.packeteer.com

[2] eDonkey, http://www.edonkey2000.com

[3] T. Karagiannis, A. Broido, N. Brownlee, K.C. Claffy, and M. Faloutos, "Is P2P Dying or just Hiding?," IEEE Globecom, 2004.

[4] T. Karagiannis, A. Broido, M.Faloutos, and K.C. Claffy, "Transport Layer Identification of P2P Traffic," ACM Internet Measurement Conference, 2004.

[5] A. Gerber, J. Houle, H. Nguyen, M. Roughan, and S. Sen, "P2P, The Gorilla in the Cable," National Cable & Telecommunications Association (NCTA) 2003 National Show, 2003.

[6] S. Sen and J. Wang, "Analyzing Peer-to-Peer Traffic Across Large Networks," IEEE/ACM Transactions on Networking, vol. 12, no. 2, pp. 219-232, April 2004.

[7] S. Sen, O. Spatscheck, and D. Wang, "Accurate, Scalable In-Network Identification of P2P Traffic Using Application Signatures," ACM WWW, 2004.

[8] S. Ohzahata, Y. Hagiwara, M. Terada, and K. Kawashima, "A Traffic Identification Method and Evaluations for a Pure P2P Application," Passive and Active Measurement Workshop, 2005.

[9] L. Plissonneau, J-L. Costeux, and P. Brown, "Analysis of Peer-to-Peer Traffic on ADSL," Passive and Active Measurement Workshop, 2005.

[10] nProbe, http://www.ntop.org/nProbe.html

[11] NetFlow, http://www.cisco.com/warp/public/cc/pd/iosw/ioft/netlct/tech/napps_wp.htm

[12] Tcpdump, http://www.tcpdump.org

[13] Flow-tools, http://www.splintered.net/sw/flow-tools/

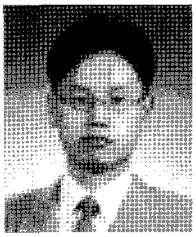
[14] T. Choi, C.Kim, S. Yoon, J. Park, B. Lee, H. Kim, H. Chung, and T. Jeong, "Content-aware Internet Application Traffic Measurement and Analysis," NOMS, 2004.

[15] PeePop, http://www.peepop.net

[16] Clubbox, http://www.clubbox.co.kr

[17] Enppy, http://enppy.entica.com

이 영 석 (Youngseok Lee) 정회원



1990년 3월~1995년 2월 서울대학교 컴퓨터공학과(학사)

1995년 3월~1997년 2월 울대학교 컴퓨터공학과(석사)

1997년 3월~2002년 8월 서울대학교 컴퓨터공학부(박사)

2002년 10월~2003년 7월 University of California, Davis 방문연구원

2003년 7월~현재 충남대학교 전기정보통신공학부 조교수

<관심분야> 차세대 인터넷, MPLS, 트래픽 엔지니어링, WDM