엔터프라이즈 네트워크의 트래픽 경향 변화에 대한 분석

*김 진, 박병철, 홍성철, 홍원기

*포항공과대학교 정보통신대학원
포항공과대학교 컴퓨터공학과
{*aroma, fates, pluto80, jwkhong}@postech.ac.kr

Enterprise Network Traffic Trend Analysis

*Jin Kim, Byung Chul Park, Seong-Cheol Hong, and James Won-Ki Hong

*Graduate School of Information Technology, POSTECH
Dept. of Computer Science and Engineering, POSTECH.

요 약

본 연구는 포항공대 네트워크를 대상으로 수집한 3 년간의 트래픽 중에서 하루씩을 선정하고, 특정한 의미가 있는 시간대를 비교, 분석한 내용이다. 연구에 사용한 데이터는 포항공대 백본 네트워크에서 수집되었다. 연구 방법으로는 flow 기반의 분석 방법론을 바탕으로, flow 의 지속시간과 바이트 크기에 대한 분석, 사용된 IP 에 대한 분석, 대역폭의 변화, 포트의 사용에 대한 분석을 수행하였다. 이 결과를 바탕으로 엔터프라이즈 네트워크에 대한 3 년간의 트래픽 경향을 분석하였다.

I. 서론

오늘날 인터넷 사용자들이 늘어남에 따라 네트워크를 사용하는 어플리케이션의 수도 늘어나게 되었다. 이러한 현상은 다양한 종류들의 트래픽을 만들어 내게 되고, 이들은 결국 네트워크 안에서 복잡하게 나타날 수 밖에 없다. 또한, 이러한 현상과 더불어 사용자의 패턴이나 네트워크 트래픽의 특성도 변화되어 왔다. 때문에 해당 네트워크의 트래픽을 분석하는 연구[1, 2]나 이상현상에 대한 분석을 하는 연구[3]들이 진행되어 왔다. 본 연구에서는 포항공대 네트워크에 대한 특징 분석과 사용자의 사용행동 등에 대한 변화를 분석하기 위해 지난 3 년간 수집한 트래픽을 바탕으로 flow 기반의 분석방법을 사용하였다.

II. 본론

1. 데이터 수집, 선정 및 분석방법

본 연구의 사용한 트래픽 데이터 수집은 포항공대 학내 네트워크와 인터넷 사이에 오가는 패킷을 대상으로 하였고 사용한 파일은 하루 중 3 개의 의미 있는 시간대로 나누어 사용하였다. 각 시간대는 오후, 저녁 시간대 1 시간씩과 새벽 1 시간대를 포함하여 3 개의 시간대 (총 181.2Gbytes, 9 개 파일)를 선정, 분석을 하였다. 각 시간대는 byte 단위사용량이 가장 낮은 새벽 5~6 시와 사용량이 가장 높은 오후 1~2 시, 저녁 8~9 시로 선정을 하여 연구를 수행하였다. 분석에 사용한 데이터 패킷은 전체 byte 를 기준으로 99% 이상을 차지하는 TCP 와 UDP 패킷을 대상으로 하였다.

분석 방법으로는 전체적인 대역폭의 변화를 조사하였고, flow 에 대한 정보를 바탕으로 한 지속시간(sec)과 그 크기(byte)에 대해서 분석하였으며, 사용된 IP 와 포트에 대한 분석을 하였다.

2. 사용한 대역폭 변화

그림 1 은 대역폭 소비량 변화를 Mbps 단위로 나타내었다. 전체적으로 연간 소비량은 증가하는 추세를 보이며, 그 추세는 특히 2008 년과 2009 년의 오후와 저녁 시간대에 급격하게 올라감을 확인할 수 있다. 13 시에서부터 20 시까지는 volume 이 가장 큰 peak 시간대임을 확인하였다. 이후 그 수치가 점점 감소하다가 새벽 5 시 이후 다시 증가하는 일반적인 사용현상을 보인다. 2009 년 현재는 peak 시간대 평균 550~580Mbps 정도로 측정이 되었음을 알 수 있다. 표 1 에서는 대역폭의 변화를 평균 대역폭 수치와 전년도 대비 증감비율을 나타내었다.

표 1. 평균 대역폭의 변화 (Mbytes)

	2007	2008	2009
오전 5~6 시	272	136(-50%)	405(+197.79%)
오후 1~2 시	366	427(+16.67%)	554(+29.74%)
저녁 8~9 시	361	428(+18.56%)	584(+36.45%)

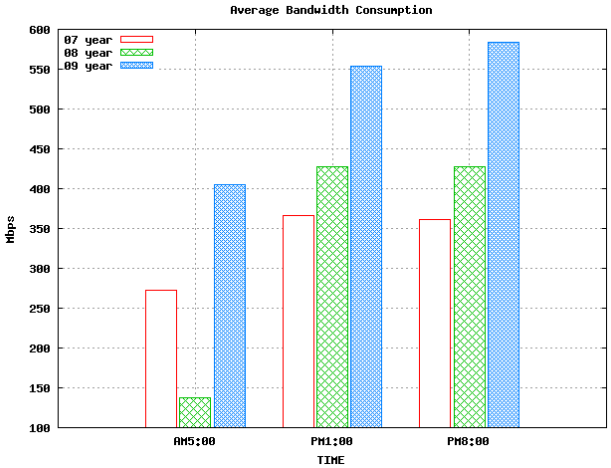

그림 1. 대역폭 변화

"본 연구는 지식경제부 및 정보통신연구진흥원의 대학 IT 연구센터 지원사업의 연구결과로 수행되었음" (IITA-2009-C1090-0902-0045)

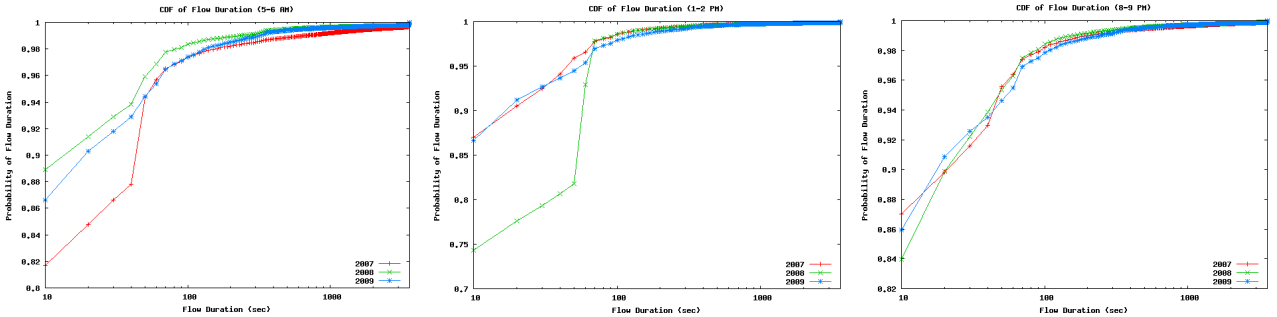

그림 2. Flow 지속 시간에 대한 CDF (초)

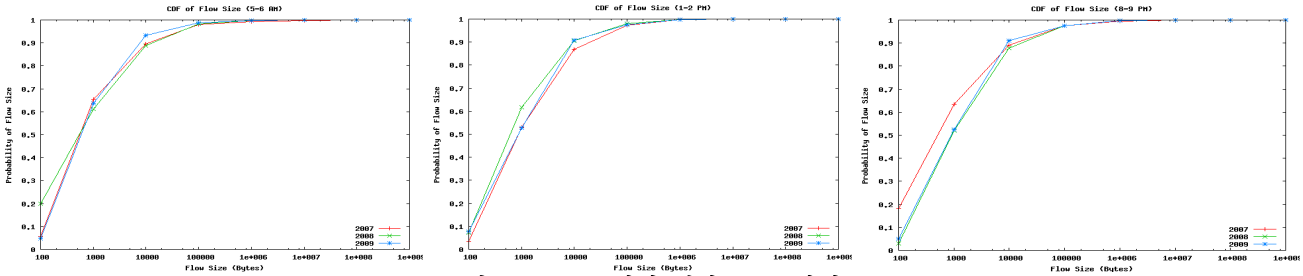

그림 3. Flow 크기에 대한 CDF (바이트)

3. Flow 별 분석

그림 2 와 3 은 flow 에 대한 분석 결과이다. Flow 지속시간에 대한 CDF(그림 2) 를 보면 전체적으로 약 95% 이상의 flow 들이 100sec 이하의 지속시간을 갖는다. Flow 크기에 대한 CDF(그림 3) 를 보면 각 시간대의 flow 들의 50%가 1,000 bytes 이하의 크기를 가지는 것을 볼 수 있다. 새벽 5~6 시 사이에는 1,000 bytes 이상의 flow 들이 50% 정도를 차지하는 것으로 측정 되었다. 또한 지속시간 CDF 와 비교하였을 때, flow 의 지속시간이 짧은 flow 들의(100sec 이하) 비율이 높음에도 불구하고, flow 크기가 큰 것의(10,000 bytes 이상) 비율이 작다는 것은 크기가 비교적 적은 flow 들의 개수가 많지만, 대부분의 대역폭은 소수의 크기가 큰 flow 에 의해 점유된다고 여겨진다.

4. 사용된 IP 에 대한 분석

표 2 는 전체 사용된 IP 개수를 나타낸 표이다. 2008 년 새벽 5 시에 사용된 주소의 수를 보면 오후나 저녁시간대의 주소의 수와의 차이가 적다는 것이다. 일반적으로는 새벽 시간에 사용된 IP 의 개수는 오후 시간 때의 개수보다 낮게 측정되지만, 2008 년에는 높게 측정 되었다. 따라서 새벽 5 시의 source 와 destination 을 구분한 IP 조사를 수행하였다. Source 만을 대상으로 조사하였을 때의 개수가 destination 만을 대상으로 조사한 것의 개수보다 현저히 낮음을 알 수 있다. 또한 이때의 패킷 중 destination 이 포항공대를 대상으로 하는 IP 주소인 패킷 대부분이 크기가 0 인 것이 많았고, UDP 보다는 TCP 를 많이 사용한 것으로 조사되었다.

표 2. 사용된 IP 의 개수

	새벽 5 시	오후 1 시	오후 8 시
2007	8,471	10,924	9,920
2008	32,647	32,738	31,256
2009	26,255	32,754	28,005

5. 80 번 포트 사용에 대한 연구

일반적으로 80 번 포트는 HTTP 프로토콜을 사용하기 위한 포트로 사용되지만, 최근에 많은 프로그램들이 동적인 포트 번호 할당으로 Well-Known Port 의 의미가 희미해지고 있다. 또한, 다른 연구들에서 웹 트래픽의 증가로 80 번 포트에 대한 바이트 소비량이 집중적으로 많은 것으로 측정 되었지만[2], 지역적 특성 (캠퍼스) 상 전체적인 포트에 대한 바이트 소비량이 증가함을 확인하였다.

III. 결론

본 연구 결과, 전체 네트워크의 소비 대역폭 변화는 전체적으로 상승하는 추세를 보이고 있다. Flow 에 대한 분석결과, 크기가 작고 지속시간이 짧은 flow 들이 많이 관찰되었고, 지속시간이 긴 flow 들이 많은 양의 volume 을 사용하는 것으로 보인다. 사용된 IP 에 대한 분석결과, 네트워크 스캐닝으로 의심되는 행위가 관찰 되었다. 포트에 대한 분석결과로 전체적인 사용량(byte)이 늘어난 것이 확인되었다. 추후 연구로는, 본 연구에서는 비정상적인 트래픽에 대한 분류를 하지 않았지만, 이를 분리하여 일반적인 트래픽들과 비정상적인 트래픽들에 대한 연구를 수행해야 할 것이다. 또한, flow 정보를 이용한 다양한 시각의 분석 방법들을 적용해서 다른 의미 있는 결론들이 나와야 할 것이며, 연구를 통해서 알게 된 비정상적인 현상들에 대한 정의를 내리고 이에 대한 해결책을 제시 할 수 있는 연구를 진행되어야 한다.

참 고 문 헌

[1] Myung-Sup Kim, Young J. Won, and James W. Hong, “Characteristic analysis of internet traffic from perspective flows.”, Computer Communications, Volume 29, Issue 10, June 19, 2006, pp 1639-1652.

[2] Pierre Borgnat, Guillaume Dewaele, Kensuke Fukuda, Patrice Abry, and Kenjiro Cho, “Seven Years and One Day: Sketching the Evolution of Internet Traffic.”, INFOCOM2009, Rio de Janeiro, Brazil, April 2009.

[3] David Moore, Geoffery M. Voelker, and Stefan Savage, “Inferring Internet Denial-of-Service Activity.” 10th USENIX Security Symposium, Washington, D.C., USA, August 13-17 2001, pp. 9-22.