플로우 기반 인터넷 트랙픽 측정 및 분석*

이영석, 옥도민, 최양희, 신효정*, 진영민*

서울대학교 컴퓨터공학과, 한국통신 통신망연구소*

Flow-based Internet Traffic Measurement and Analysis

Yongseok Lee, Tomin Ok, Yanghee Choi, Hyojeong Shin*, Yeongmin Jin*

Dept. of Computer Engineering, Seoul National University, and Korea Telecom Telecommunication Network Research Lab *

요 약

인터넷은 급속도로 사용자, 트래픽, 호스트, 네트워크 등이 증가하고 있다. 이러한 증가에 따라 인터넷의 효과적인 관리를 위해서는 기본적인 인터넷 사용 트래픽의 측정 및 분석이 필요하다. 최근의 OC3mon, NetFlow System 등의 트래픽 측정 도구들은 고속의 인터넷 백본에서 유통되는 트래픽들의 특성들을 파악할 수 있도록 한다. 특히 고속 인터넷 플로우 스위칭이 이용하는 IP 패킷들의 연속적인 흐름인 플로우(flow)를 바탕으로 인터넷 트래픽 특성을 파악 할 수 있도록 한다. 본 논문에서는 플로우 기반의 인터넷 트래픽 측정 도구인 Cisco 라우터의 NetFlow 인터페이스와 Cflowd 수집기를 사용하여 국내 교육 인터넷 망의 트래픽을 측정하여 분석한 국내 인터넷 트래픽 특성 결과를 제시한다.

주요어 : 트래픽 측정, 인터넷 플로우

1. 개요

급증하는 인터넷 사용에 대한 체계적인 분석을 위해 인터넷 트래픽을 측정하고 분석하는 일은 매우 시급하고도 기본적인 일이다. 특히 최근에 들어 WWW(World Wide Web), 멀티미디어 스트림 응용 및 화상회의 등의 다양한 멀티미디어 응용이 개발되어 사용되고 있고, 이들 응용들의 트래픽은 인터넷 성능에 큰 영향을 끼치게 된다. 따라서, 현재 인터넷 트래픽에 대한 측정 데이타와 분석 결과는 미래에 대한 지표가 될 수 있고, 차세대 인터넷을 구축하기 위한 주요 정보가 될 것이다.

인터넷이 제공하는 네트워크 서비스의 성능을 지연시간, 지연시간 편차, 대역폭, 손실율 등의 메트릭(metric)을 측정함으로써 인터넷의 성능을 직접적으로 파악할 수 있게 한다. 인터넷 백본에서의 트래픽 유통량이나 사용 패턴을 분석하여 네트워크의 성능이나 문제점을 개선시키는데 사용한다. 특히 주요 인터넷 백본에서 측정된 결과는 앞으로의 인터넷의 트래픽과 네트워크의 미래를 예측할 수 있는 중요한 자료가 될 것이다.

현재 CAIDA(Cooperative Association for Internet Data Analysis)[1]의 OC3mon[6], IETF RTFM(Realtime Traffic Flow Measurement) WG[2]의 NeTraMet[3], Cisco 라우터에서 지원하는 NetFlow[11]등이 IP 플로우를 기반으로 하는 인터넷 측정 프레임워크 및 도구들을 제공하고 있다.

본 논문에서는 인터넷 백본망의 트래픽을 분석하기 위해서 플로우 기반 측정 도구인 Cisco 라우터의 NetFlow 와 트래픽 수집기인 Cflowd[7] 를 이용하여 국내 교육망[12]의 백본 트래픽을 측정하였다. 그리고, 이 측정 데이터를 플로우에 기반한 인터넷 트래픽 특성을 제시하기로 한다.

2 장에서는 인터넷 트래픽 측정에 관련된 연구 그룹과 도구들을 살펴보고, 인터넷 트래픽 플로우 모델을 3 장에서 설명한다. 그리고, 인터넷 백본 트래픽 측정 도구 및 구조를 4 장에서 보여주고 측정 결과 및 분석 결과를 5 장에서 설명한다. 마지막으로 6 장에서는 결론과 향후 계획에 대해서 언급하기로 한다.

2. 관련 연구

NSF(National Science Foundation)에서 1997 년에 창설한 인터넷 데이터 분석 단체인 CAIDA 에서는 인터넷 측정을 위한 기본 구조 제시, 측정 메트릭 선정, 트래픽 시각화, 시뮬레이션, 분석 차세대 인터넷 프로토콜 및 기술 등에 관한 것을 포괄하고 있다.

최근의 인터넷 트래픽 측정 도구들은 응용별 타입(HTTP, e-mail, FTP, real-audio, and telnet), 트래픽 송신자/수신자, 패킷 크기, 플로우 지속시간 분포 등에 관한 정보를 제공해 줄 수 있다. OC3/OC12 플로우 모니터는 MCI 에서 개발되어서 vBNS 의 주요 OC3 링크에서 인터넷 플로우를 측정하고 분석하고 있다[5]. 고속의 인터넷 백본을 이루는 ATM 망에서의 인터넷 트래픽을 측정하기 위해 개발된 OC3mon 은 IP-over-ATM 트래픽이 지나가는 OC3 광 케이블을 절단기(splitter)르 측정시스템과 연결하여 ATM 셀 데이터를 측정하여 IP 패킷 및 플로우 데이터를 측정하여 분석한다.

그리고, Cisco 라우터는 NetFlow 라는 플로우 데이터를 이용하여 수집 및 분석할 수 있도록 한다. NetFlow Collector, NetFlow Analyzer 등의 Cisco 상용 측정 및 분석 응용이나 Cflowd 공개 측정 응용을 이용하여 NetFlow 데이터를 수집하여 분석할 수 있다.

IETF 의 RTFM(Realtime Traffic Flow Measurement) WG 에서는 다양한 플로우를 정의하여 측정하고 실시간으로 분석할 수 있는 프레임워크를 제시하고 있다.

인터넷 트래픽 측정을 위해서 기본적으로 필요한 것

*본 연구는 한국통신 지원으로 수행되었다

은 측정하고자 하는 메트릭을 설정하는 것이다. 이와 관련된 그룹으로 IETF IPPM(IP Performance Metrics) WG[4]에서 연결성(connectivity), 단방향 지연시간(one way delay), 전송 용량(bulk transfer capacity) 등의 주요 메트릭과 측정을 위한 기반 구조를 정의하고 있다 지연시간, 패킷 손실, 가용 대역폭 등은 기본적인 인터넷 성능 측정 메트릭이다 TCP 프로토콜의 처리율과 같은 메트릭을 측정하기 위해 treno, tcpanaly, 등의 도구들이 사용되고 있고, 종단간 지연시간 메트릭을 측정하기 위해서 ping, traceroute 등이 이용된다. 가용대역폭 메트릭을 측정 분석하는 도구로는 pathchar, bing, b-c probe 등이 있다

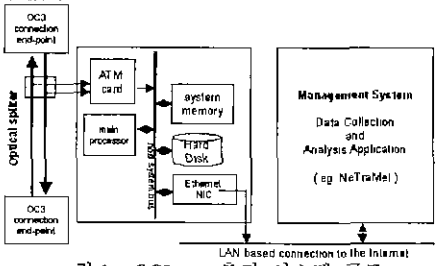

그림 1. OC3mon 측정 시스템 구조

3. 플로우 모델

플로우는 연속된 IP 패킷들로 정의될 수 있다. 즉, 특정 송신자에서 수신자로 전송되는 IP 패킷들의 흐름을 플로우로 구분하는데 이러한 플로우 모델은 [10]에서 Packet Train 모델로 처음 제안되었다. 하지만, 비연결형 IP 계층, IP 파편화에 따른 TCP 나 UDP 포트 정보 유실, 최근의 멀티미디어 스트림 응용들의 등장 및 ATM 등장으로 IP 플로우의 ATM 연결 매핑 등을 고려하여 타임아웃을 고려한 비연결형 IP 플로우 모델이 [8]에서 제안되었다 그림 2 에서 플로우의 정의에 필요한 시간 인자를 설명한다

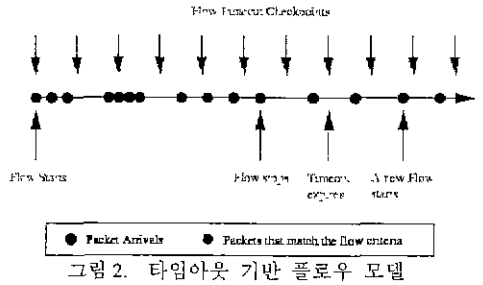

그림 2. 타임아웃 기반 플로우 모델

인터넷 플로우의 지속 시간은 응용별로 다양한 모습을 취하고 있기 때문에 플로우를 탐지하기 위한 시간에 따라 플로우의 특성이 달라질 수 있다. 예를 들어 타이머 값을 짧게 한다면 긴 플로우를 여러 개의 짧은 플로우로 구분한다. 인터넷 트래픽은 지속적인 플로우 기반의 FTP, telent, http, 멀티미디어 응용 트래픽과 짧은 지속 기간의 DNS, SMTP, NTP, POP, SNMP 등의 응용으로 구분할 수 있다

4. 트래픽 측정 환경

본 연구에서는 OC3 링크로 구성된 국내 인터넷 백본인 KORNET 과 T3 링크로 구성된 교육망의 트래픽을 측정하고 분석하고 있는데, 논문에 제시된 데이터는 교육망에서 측정된 트레픽이다. 교육망 트래픽을 측정하기 위해서 Cisco 사의 라우터의 NetFlow 패킷을 수집하는 Cflowd 를 사용하였다.

측정 환경은 그림 3 에서 나와있듯이 Cisco 라우터와 측정 시스템간의 NetFlow 데이터그램은 UDP 패킷으로 전송된다.

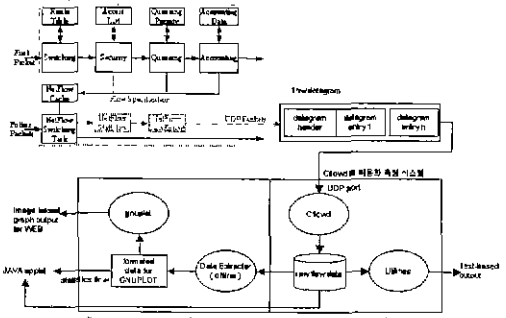

그림 3. 교육망 인터넷 트레픽 측정 환경

NetFlow 에서 정의하고 있는 플로우는 (송신자 IP 주소, 송신자 포트, 수신자 IP 주소, 수신자 포트, 프로토콜, TOS 필드) 쌍의 플로우 스펙을 이용하고 있다.

측정된 NetFlow 데이터 그램은 측정 시스템에 저장되어 오프라인으로 플로우 분석에 사용되거나, 실시간 플로우 분석을 한다. 현재 오프라인으로 저장된 플로우 파일을 분석한 자료가 5 장에서 설명될 것이고, 실시간 플로우 분석기는 Java Applet 이나 CGI 응용으로 구현할 예정이다.

5. 트래픽 측정 결과 및 분석

교육망 백본에서 1998 년 8 월 10 일에서 8 월 14 일까지 측정된 플로우 트레픽은 시간대별 플로우 수, 패킷 수, 바이트 수 등의 트래픽 유통량과 응용 포트별, 프로토콜별로 자세한 사용 현황을 파악할 수 있다

그림 4 에서는 전체 측정된 플로우 수의 변화율을 1 분 단위로 합산한 결과를 보여주는 것으로 주간 시간대에서의 플로우수가 야간 시간대에 비해 대부분을 차지한다는 것을 알 수 있다. 이것은 인터넷 사용자의 이용 시간대별로 플로우 변화율을 잘 나타낸다. 그림 5 는 IP 패킷의 프로토콜 필드로 플로우를 구분한 것을 보여준다 TCP, UDP, ICMP 의 순서대로 플로우 수의 변화율을 보여주고 있다 TCP 플로우가 절대적으로 많은 것은 인터넷 대부분의 주요 응용이 TCP 프로토콜을 사용하고 있기 때문이다. 대표적인 것이 바로 WWW 응용의 HTTP 프로토콜이다.

그림 6 에서는 플로우의 정보에서 바이트 수에 대한 변화율을 보여주고 있다 교육망은 현재 T3(45Mbps) 링크의 백본으로 구성되어서 그만큼의 데이터 전송율을 지원하고 있다. 프로토콜별, 응용 포트별 사용 패턴은 기본적으로 HTTP 트래픽의 절대 우위에 따른 모습을 보여주고 있다.

인터넷의 많은 응용들은 포트에 따라 구분을 하고 있

다 그 중에서 현재 인터넷에서 멀티미디어 스트림 응용의 대표적인 예인 RealMedia 응용 프로그램이 사용하는 6970 ~ 7170 포트의 트래픽을 분석한 것이 그림 7 에 나와있다.

응용 포트별 플로우의 지속 시간 누적 분포를 그림 8 에서 구하였다. 244 초의 플로우의 지속 시간은 90%, 458 초의 플로우 시간은 95%, 840 초일 경우 98%로 대부분의 플로우가 1000 초 이내이다. 그리고, 인터넷 플로우는 다수의 짧은 플로우와 긴 플로우의 응용으로 크게 나뉘어진다. 즉 telnet, nntp 등의 응용은 상대적으로 긴 시간의 플로우를 지속하는 반면, HTTP, ftp, DNS, SMTP, NTP, POP, SNMP 등의 응용은 짧은 시간의 플로우이다.

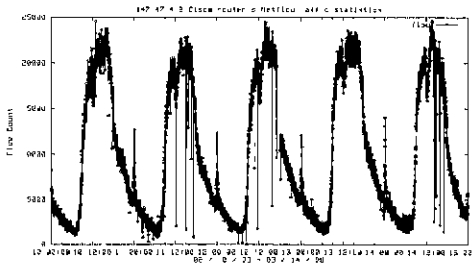

그림 4. 전체 플로우 수 변화율

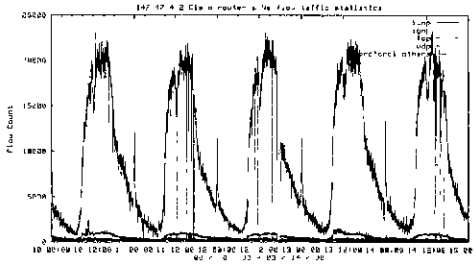

그림 5. 프로토콜별 플로우 수 변화율

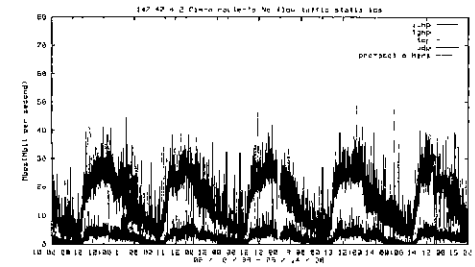

그림 6 프로토콜별 바이트 수 변화율

6. 결론 및 향후 연구

본 논문에서는 인터넷 트래픽 측정 및 트래픽 패턴과 특성을 분석하였다. 대부분의 트래픽은 TCP 연결의 데이터를 이용하는 것들이 많았고, 그 중 WWW 사용에 따른 HTTP 프로토콜 데이터가 상당수 차지하였다. 또한, 인터넷 트래픽의 플로우 특성은 대부분 짧은 시간동안 지속되는 플로우가 많다는 것이다.

이러한 국내 교육망 인터넷 트래픽의 특성들은 이미 알려진 외국 인터넷 백본에서의 트래픽 특성과 유사하며, 앞으로의 인터넷 트래픽 증가에 대한 예측을 할 수 있는 기초 자료로서의 역할을 할 수 있을 것이다. 따라서 인터넷 백본에서의 트래픽 특성을 보다 정확하게 파악하기 위해서는 OC3mon 과 같이 ATM 셀에서 패킷 정보, 플로우 정보를 측정하는 작업이 필요할 것이다 또한 측정된 트래픽의 다양한 시각화 작업[9]과 실시간 관측 시스템의 개발이 필요하다.

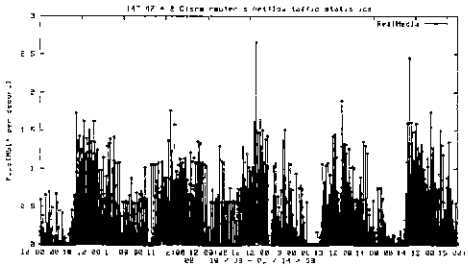

그림 7 RealMedia 바이트 수 변화율

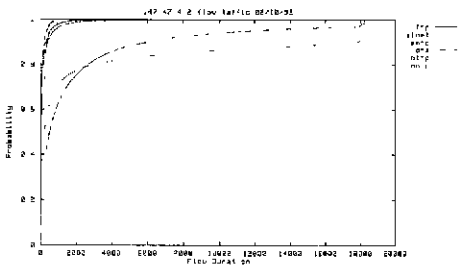

그림 8. 응용 포트별 플로우 지속시간 누적 확률 분포

7. 참고 문헌

[1] CAIDA. http://www.caida.org

[2] IETF RTFM WG, http://www.ietf.org/html charters/rtfm-charter.html

[3] N. Brownlee, C. Mills, G. Ruth, "Traffic Flow Measurement: Architecture, " RFC2063, January 1997

[4] IETF IPPM WG, http://www ietf org/html.charters/ippm-charter.html

[5] K. Thomson, G. J Miller, and R. Wilder, "Wide-Area Internet Traffic Patterns and Characteristics", IEEE Network November/December 1997

[6] J. Apisdorf et al., "OC3MON Flexible, Affordable, High-Performance Statistics Collection", Proc. INET'97, June 1997.

[7] D. McRobb and J. Hawkinson, "cflowd: A Cisco Flow-Export Collector", http://engr.ans.net/cflowd/

[8] K. C Claffy, "Internet Workload Characterization," Ph.D. dissertation, Univ. CA, San Diego, June 1994

[9] K. Claffy and T. Monk, "What's Next for Internet Data Analysis ? Status and Challenges Facing the Community," Proc. of IEEE, October 1997

[10] R. Jain and S A. Routhier, "Packet Trains – Measurements and a New Model for Computer Network Traffic", IEEE JSAC, Sep 1986

[11] Cisco Systems, "NetFlow", White Paper, 1997.

[12] http://www.kren.ne.kr