Flow 특성을 이용한 인터넷 트래픽 분석

옥도민, 김재훈, 이영석, 최양희

Internet Traffic Measurement and Analysis by Flow characterization

Tomin Ok, JaeHoon Kim, Yongseok Lee, Yanghee Choi
Dept. of Computer Engineering, Seoul National University

요 약

인터넷 서비스의 급속한 확산과 기술 발전으로 인하여 대량의 정보를 고속으로 전송하기 위해 고속의 링크 기술을 이용하고, ATM 인터넷 기간망을 수용하고 다양한 서비스 품질을 제공하는 등 다양한 차세대 인터넷 기술의 등장과 함께 네트워크 구조에 많은 변화가 이루어 지고 있다. 또한 다양한 인터넷 응용의 등장과 함께 인터넷 이용자의 네트워크 트래픽이 급속도록 증가하고 있다. 이에 따라 인터넷의 다양한 트래픽을 특성지을 수 있는 플로우 개념을 이용하여 이용자 네트워크 사용 패턴 및 망 관리 방안등을 제시 할 수 있다. 이러한 데이터를 분석하여 인터넷의 성능 및 특성을 평가하는 것은 단기적, 중기적, 장기적인 네트워크 설계 관점에서 매우 중요한 기초 연구이다.

주요어 : 인터넷 플로우, CFLOWD[1], NetFlow[9]

1. 개요

급증하는 인터넷 사용에 대한 체계적인 분석을 위해 인터넷 트래픽을 측정하고 분석하는 일은 매우 시급하고도 기본적인 일이다. 특히 최근들어 WWW(World Wide Web), 멀티미디어 스트림 응용 및 화상회의 등의 다양한 멀티미디어 응용이 개발되어 사용되고 있고, 이들 응용들의 트래픽은 인터넷 성능에 큰 영향을 끼치고 있다. 따라서 인터넷 서비스 제공자들은 증가하는 사용자의 요구사항을 만족시키기 위하여 라우터, 스위치, 링크 등의 네트워크의 주요 요소들의 성능을 측정하여 네트워크를 관리하여 문제를 해결하거나, 인터넷 트래픽 수집 및 분석을 기반으로 네트워크의 주요 요소들의 성능을 측정하여 네트워크를 관리하여 문제를 해결하거나, 인터넷 트래픽 수집 및 분석을 기반으로 네트워크의 용량을 증설하거나 네트워크를 재 설계할 수 있도록 하고 있다. 차세대 인터넷은 다양한 서비스 품질(Quality of Service)과 고속의 전송률을 제공할 수 있도록 설계되고 있으며, 이미 Internet2, NGI, CA*Net3, TEN155, APAN 등의 시험망이 테스트되고 있다. 현재 인터넷 트래픽에 대한 측정 데이터와 분석 결과는 미래에 대한 지표가 될 수 있고, 차세대 인터넷을 구축하기 위한 주요 정보가 될 것이다.

인터넷 성능 측정 및 분석은 네트워크에서 발생하는 전체 트래픽의 흐름을 파악하여 현재 가용한 네트워크 자원들의 적절한 균형을 맞추어 네트워크 비용 절감 효과를 유도할 수 있도록 하고, 중장기적으로 네트워크를 증설하거나 구조를 재설계할 때 활용하여 불필요한 네트워크 과잉 투자를 방지할 수 있도록 한다. 또한 인터넷 트래픽과 네트워크 성능을 지속적으로 관측하여 분석함으로써 특정 네트워크 지역의 병목현상이나 발생 가능한 네트워크 장애를 발견하거나 예측하여 사전에 대비하도록 하여 네트워크 관리 비용을 줄일 수 있도록 한다. 인터넷 망에서 유통되고 있는 트래픽을 실제로 관측하여 분석함으로써 이용자의 인터넷 사용 패턴, 사용자 트래픽 모델 등 다양한 인터넷 사용현황에 대한 정보를 제공받을 수 있다. 일반적으로 트래픽 측정은 네트워크 관리자에게는 실시간 네트워크 모니터링, 네트워크 관리, 네트워크 설계 및 증축, 사용량에 따른 과금 정책 등에 사용될 수 있다.

2 장에서는 인터넷 트래픽 플로우 특성에 관련된 연구를 살펴보고, 측정 및 분석에 이용된 트래픽 플로우 모델을 3 장에서 설명한다. 그리고, 측정구조 및 저장형식을 4 장에서 보여주고 측정 결과 및 분석 결과를 5 장에서 설명하고 마지막으로 6 장에서 결론 및 향후 계획에 대해서 언급하기로 한다.

2. 관련 연구

최근 인터넷 트래픽 패턴 분석 연구에서는 많은 데이터가 송신자와 수신자의 연속적인 데이터 흐름으로 구성되는 것이 발견되었고, 이를 바탕으로 패킷 단위로 처리하여 전송하는 라우터의 성능을 개선시킨 고속 스위칭 기술들이 개발되었다. Ipsilon 사의 IP switching, Cisco 사의 tag switching 등의 MPLS(Multi-Protocol Label Switching) 기술들은 ATM 셀을 분해하고 재조립해 패킷 단위로 처리하던 라우터의 단점을 플로우(FLOW) 단위의 ATM 고속 연결을 이용한 하부 계층 스위칭 방법을 사용하여 IP 라우팅 성능을 개선 시켰다. MPLS 기술들은 인터넷 트래픽을 연속적인 패킷 흐름인 플로우를 기본으로 가정하고 있고, 많은 측정 결과들도 이러한 플로우의 특성을 반영한다. IP 플로우는 멀티미디어 스트림 응용, IPv6, RSVP(Resource ReSerVation Protocol), MPLS 등에서 사용되는 중요한 인터넷 트래픽 모델이다. 따라서 인터넷 트래픽의 특성을 파악하기 위해서는 ATM 셀이나 이더넷 프레임에서 부터 IP 패킷, 응용 플로우까지 다양한 수준의 데이터에 대해 시행되어야 하고, 특히 IP 플로우에 근간한 트래픽 측정 및 분석 방법이 개발되고 있다.

IETF 의 RTFM(Realtime Traffic Flow Measurement) WG[3]에서는 다양한 플로우를 정의하여 측정하고 실시간으로 분석할 수 있는 프레임워크를 제시하고 있다. 그리고, Cisco 사는 NetFlow 라는 통계화된 플로우 데이터를 이용하여 수집 및 분석을 할 수 있도록 하고 있다. NetFlow Collector, NetFlow Analyzer 등 Cisco 상용 측정 및 분석 응용이 있고, CAIDA[2]에서 CFLOWD 공개 측

정 응용을 제공하며 이를 이용해서 NetFlow 데이터를 수집 및 분석이 가능하다.

3. 인터넷 트래픽 플로우 모델

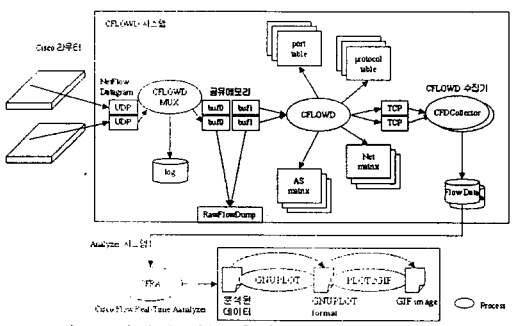
플로우는 일련의 동일한 특성을 갖는 패킷들의 집합으로 정의될 수 있다. 일련의 동일한 특성으로는 특정 송신자와 수신자 쌍으로 정의될 수 있으며, 플로우의 크기(granularity)를 결정하는 중요한 요소로 작용한다. 이러한 플로우 모델은 [4]에서 Packet Trains 모델로 제안되었다.

비연결형 IP 플로우를 고려한 타임아웃 기반 플로우 모델로 IP 파편화에 에 따른 TCP 나 UDP 응용의 정보 유실, 최근 멀티미디어 스트림 응용과 ATM 등장으로 인한 IP 플로우의 ATM 연결 매핑등을 고려한 모델[5]로 NetFlow 에서도 이 모델을 기반으로 하여 Flow Switching 을 구현하고 있다.

4. 트래픽 측정 구조 및 환경

NetFlow 는 Cisco 라우터에서 플로우 기반의 트래픽 측정을 가능하게 하며, 트래픽을 플로우 단위로 분류하거나 우선순위를 구분하여 다양한 서비스 품질을 제공할 수 있도록 한다. 네트워크 플로우는 주어진 송신자와 목적지 종단간의 일련의 단방향 패킷들인데, 플로우의 세분화는 플로우의 종단에 따라 다양하다. IP 주소, 전송계층 응용 포트번호 등으로 플로우 종단을 지칭할 수 있으며, NetFlow 는 여기에 덧붙여 IP 프로토콜 타입, 서비스 타입(Type of Service : ToS), 입력 인터페이스 식별자를 이용하여 유일한 플로우를 식별하게 된다

NetFlow 의 데이터 그램은 16 바이트(버전 5 인경우 24 바이트)의 헤더와 최대 25 개(버전 5 인경우 30 개)의 플로우 항목들을 포함하고 있다. 각 플로우 항목은 48 바이트의 고정 크기이다. 라우터는 최소한 초당 한 개 이상의 플로우 개체를 만들어 전송한다. 따라서 망의 상태에 따라 상당히 많은 양의 NetFlow 테이터 그램이 전송되므로 이를 분석하기 위한 저장형식을 새롭게 정의하고 있다. AS 메트릭스, net 메트릭스, port 메트릭스, protocol 테이블, 인터페이스 테이블, 그리고 nexthop 테이블등으로 데이터를 가공하여 저장할 수 있다.



[그림 1] 인터넷 성능 측정 시스템 구조 (CFLOWD)

본 연구에서는 OC3 링크로 구성된 국내 인터넷 백본인 KORNET 망의 트래픽을 CiscO NetFlow 와 Cflowd [그림 1]측정 구조를 이용하여 측정하고 분석하였다.

5. 측정 결과 및 분석

[표 1]은 1999 년 8 월 9 일 하루동안 KORNET 망에서 측정된 데이터들의 전체 플로우수, 패킷수, 바이트수를 보여주고 있다. 트래픽 양에 있어서는 TCP 트래픽이 90%정도 차지하고 있음을 알 수 있다. 플로우의 TCP 와 UDP 트래픽의 플로수는 거의 비슷하거나 오히려 UDP 의 플로우 수가 더 많음을 알 수 있는데, 이는 플로우 크기가 매우 작은 DNS, BattleNet, 등의 트래픽이 대부분이기 때문이다.

	flows	packets	bytes
total	29818152	277187388	1.51E+11
tcp	14026920	200319932	1.39E+11
udp	15486684	74622624	1.17E+10

[표 1] 전체 트래픽양

프로토콜별 세분화된 응용별 트래픽을 [표 2], [표 3] 에서 각각 보여주고 있다. TCP 의 경우 HTTP 트래픽이 70% 이상을 차지하고 있다. UDP 의 경우에는 DNS 트래픽이 상대적으로 많았으며, 베틀넷과 같은 네트웍 게임 트랙픽 (port 6112, 28800)도 상대적으로 많음을 알 수 있다. 응용의 특성상 플로수의 비율과 패킷수, 바이트수의 비율이 다르다는 것을 알 수 있는데, 플로우당 패킷수가 가장 큰 응용은 TELNET 이며 바이트수가 가장 큰 것은 FTP-DATA 응용이다. DNS 응용인 경우 플로우당 패킷수와 바이트수가 가장 적지만 전체 응용중에서 플로우 수는 가장 많다.

		flows	packets	bytes
tcp	http	10357748	122976384	9.97E+10
	443	168592	1415320	6.24E+08
	ftp-data	23452	3650640	4.72E+09
	telnet	17876	7145152	2.29E+09
	8080	14432	1858448	1.77E+09

[표 2] top 5 TCP 응용

		flows	packets	bytes
udp	domain	12025628	15157536	5.97E+09
	28800	1554392	8280688	2.65E+08
	6112	860508	21194212	2.15E+09
	27015	118408	3440884	8.80E+08
	2213	4592	9840	5.58E+05

[표 3] top 5 UDP 응용

TCP 트래픽의 대부분을 찾이하는 HTTP 의 플로우 크기 분포를 보면 텍스트, 이미지등 다양한 데이터를 포함하고 있기 때문에 다른 응용에 비해 폭넓은 분포를 이루고 있다. 이에 반해 UDP 트래픽 응용이나 ICMP 트래픽의 경우 2Kbyte 이하의 플로우가 95% 이상을 찾이하고 있다. 따라서 플로우를 기반으로 하는 라우터들의 경우 TCP 와 UDP 트래픽을 분리하는 것이 성능면에서 효율성을 높일 수 있다. 또한 응용별 플로우의 크기도 TCP 기반 HTTP, FTP-DATA, TELNET 트래픽과 UDP 기반 DNS, BattleNet 등의 트래픽으로 구분된다.

[그림 2]에서 플로우의 크기는 전체 트래픽 양의 70% 정도가 1Kbyte 이하이며, 90%정도가 10Kbyte 이하의 크기를 갖는 플로우이다. 플로우 크기는 플로우 스펙의 정의에 의해 조정이 가능하므로 네트웍의 성능을 결정하는 중요한 요소로 작용할 수 있다.

플로우 지속시간 분포를 보여주는 [그림 4]에서 지속시간이 3 분이상 (전체플로우의 1%)을 관리할 수 있으면 전체 트랙픽양의 40%정도의 관리가 가능해 질 수있음을 알 수 있다. [그림 5,6]에서는 프로토콜, 응용별로 구분하여 보여주고 있다. 전체 트래픽의 30 초이내의 지속기간을 갖는 플로우가 90%정도를 찾이하며, 98% 정도가 3 분이내의 지속기간을 갖는다. 플로우는 짧은 기간 플로우(shor-term flow)와 긴 기간 플로우(long-term flow)로 분리할 수 있다. 즉, HTTP, DNS, SMTP, NTP, SNMP 등의 응용은 짧은 지속시간을 갖는 플로우이고, 상대적으로 FTP-DATA, TELNET 등의 응용은 긴 지속시간을 갖는 플로우 이다. 플로우 상호도착지연(inter-arrival -time)에 대한 분포를 살펴보면, 1 초 이내에 99.5%의 플로우가 발생했고, 플로우 지속시간에 비례해서 짧은 지속시간 플로우들의 상호도착지연도 짧은 것을 볼수 있으며, 긴 지속시간을 갖는 응용들은 긴 상호도착지연을 갖는 것을 알 수 있다.

6. 향후 연구 및 결론

본 논문에서는 플로우를 기반으로하는 인터넷 트래픽의 특성을 살펴보았다. 추가적으로 네트웍별 플로우특성과, 수신자 기반 플로우 특성 및 토폴로지 기반 플로우특성등 플로우의 통합 레벨을 세분화하여 살펴보는 것 또한 네트웍관리를 위해 중요하다. 이러한 국내 인터넷 트래픽의 특성은 현재 개발 중인 MPLS 라우터등의 개발에 중요한 기초자료서 역할을 할 수 있을 것이다. 따라서 플로우 명세를 플로우 통한 레벨에 맞춰 변화시켜 인터넷 트래픽의 특성을 측정 분석하는 작업이 필요다.

7. 참고 문헌

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

[2] CAIDA, http://www.caida.org

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

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

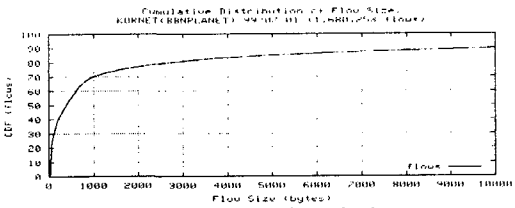
[5] K. C. Claffy, "Internet Workload Characterization," Ph.D. dissertation, Univ. CA, San Diego, June 1994

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

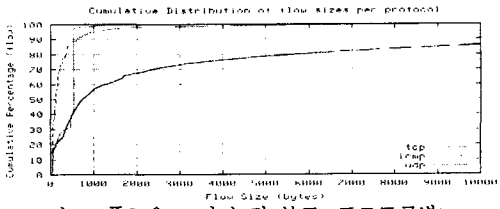
[7] IETF IPPM WG, http://www.ietf.org/html.charters/ippm-charter.html

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

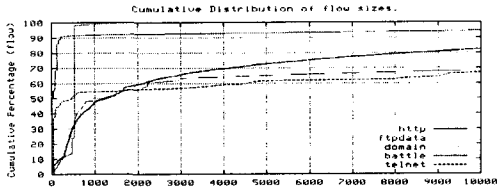
[9] Cisco Systems, "NetFlow", White Paper, 1997.



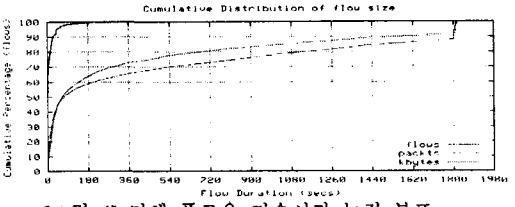
[그림 2] 플로우 크기 누적 분포



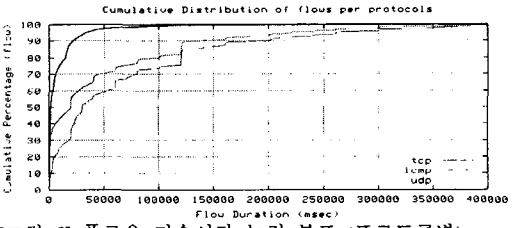
[그림 2] 플로우 크기 누적 분포 (프로토콜별)



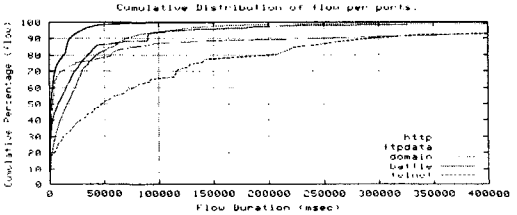
[그림 3] 플로우 크기의 누적 분포 (응용별)



[그림 4] 전체 플로우 지속시간 누적 분포 (플로우/패킷/바이트)



[그림 5] 플로우 지속시간 누적 분포 (프로토콜별)



[그림 6] 플로우 지속시간 누적 분포 (응용별)