Flow 기반 실시간 트래픽 수집 및 분석 시스템*

*박상훈, 박진완, 김명섭
고려대학교 컴퓨터정보학과
e-mail:{*2002270130, pakjw84, tmskim}@korea.ac.kr

Flow-based Real-time Traffic Monitoring and Analysis System

*Sang-Hoon Park, Jin-Wan Park, and Myung-Sup Kim
*Dept of Computer Information Science, Korea University

요 약

네트워크의 효율적인 관리를 위해서는 네트워크의 각 호스트에서 발생하는 트래픽을 실시간으로 모니터링 할 수 있는 시스템이 필요하다. 이러한 모니터링의 효율적인 방법 중 하나가 네트워크 장비에서 제공하는 flow 정보를 이용하는 방법이다. 하지만 이는 네트워크 장비의 과부하 발생, 운용비용 상승, 유연성 및 확장성 부족의 단점을 가진다. 이를 극복하기위하여 본 논문에서는 Enterprise 네트워크에 적합한 Flow 기반 실시간 트래픽 모니터링 시스템의 구조를 제안하고, 검증을 위해 구현한 내용을 기술한다. 본 시스템은 패킷을 수집하여 실시간으로 flow 정보를 생성하고 저장하는 Flow Generator 시스템, 저장된 flow 정보를 Analysis 시스템으로 전송하는 Flow Exporter 시스템, Traffic Analysis 시스템, 그리고 분석된 내용을 보여주는 Traffic Reporter 시스템으로 구성된다. 본 시스템은 다양한 분석 목적에 맞게 Flow 정보를 조절할 수 있는 유연성과 다양한 분석시스템을 구축할 수 있는 확장성을 가진다. 본 논문에서 기술한 시스템은 학교 Campus 네트워크를 대상으로 구축되었다.

1. 서론

오늘날의 네트워크는 그 중요성에 의해 많은 기술 발전과 더불어 네트워크를 이용하는 다양한 서비스의 등장으로 트래픽이 복잡해지고 있다. 이러한 상황 속에서 DoS/DDoS와 같은 유해 트래픽이나 비정상 트래픽 발생은 네트워크 전체에 큰 영향을 미친다. 이와 같은 문제의 해결을 위해 네트워크 실시간 모니터링의 필요성은 점점 증가하고 있다.

현재 실시간 네트워크 트래픽 모니터링은 대용량 트래픽의 실시간 처리를 위해 flow 기반 분석[1, 2]과 cluster 형태의 분석 시스템 구조[3, 4, 5]를 가진다. flow 기반 분석을 위하여 CISCO에서는 Netflow 모니터링 솔루션[1]을 제공한다. 그러나 이와 같이 네트워크 장비에서 제공하는 flow 정보를 이용하는 것은 분석 시스템이 제공되는 flow 정보에 제한을 받기 때문에 유연성이나 확장성이 떨어진다. 그리고 장비의 과부하, 운용비용의 상승 등의 여러 가지 단점을 가지고 있다. 그래서 이런 단점을 없애기 위해 IETF IPFIX WG에서는 유연한 구조의 flow 구조[6]를 제안하였고, 이를 통하여 flow 생성 시스템의 유연성을 보장하고 있다. 실시간 트래픽 모니터링을 위한 트래픽 분석시스템의 구조는 RTFM[3] 구조에서 시작하여 NG-MON[5]에 이르기까지 다양한 형태가 제안되고 있다. 이들은 flow 기반 실시간 모니터링 구조를 packet 수집에서 분석에 이르는 일련의 작업을 기능별로 구분하고 각 시스템에 할당하고 각 시스템의 상호협력을 통하여 실시간 분석을 하도록 구축된다.

본 논문에서는 flow 정보의 유연성을 확보하고, Enterprise 네트워크에 적합한 실시간 트래픽 모니터링 시스템의 구조를 제시하고 검증을 위해 구현한 내용을 기술한다. 본 논문에서 기술한 시스템은 학교 Campus 네트워크를 대상으로 구축되었다.

본 논문은 다음과 같은 순서로 기술한다. 2장은 시스템 구조 및 설계를 기술하고, 3장은 시스템 구현내용을 기술한다. 4장에서는 결론을 맺고 향후 연구에 대해 언급한다.

2. 시스템 구조

2.1 Overall System Architecture

Packet 단위의 분석은 정확하고 상세한 분석이 가능하지만, 오늘날의 많은 트래픽이 발생하는 고속 네트워크에서는 분석시스템 및 네트워크의 과부하로 적합하지 않다. 따라서 본 시스템은 RTFM[3]과 IPFIX[6]에서 제안한 Flow기반 분석 시스템 구조를 기반으로 하여 실시간 분석 시스템을 설계하였다. 또한, NG-MON[5]에서 제안한 5단계의 실시간 시스템의 단점인 복잡함과 분석시간 지연을 최대한 줄일 수 있는 구조로 Enterprise 네트워크에 적합한 실시간 트래픽 분석 시스템 구조를 설계 하였다.

그림 1은 시스템의 전체적인 구조를 표현한 것이다. 기본적으로 본 시스템은 Flow Generator, Flow Exporter, Traffic Analyzer, Traffic Reporter로 구성되고 Traffic

* 이 논문은 2007년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임.(KRF-2007-331-D00387)

Analyzer와 Traffic Reporter 사이엔 Database가 존재한다. Flow Generator는 트래픽 수집 링크로부터 raw 패킷을 수집하여 IPFIX 기반의 flow정보를 생성하여 local 디스크에 저장하는 역할을 수행한다. Flow Exporter는 local 디스크에 저장된 flow 정보를 Analyzer의 요청에 의해 전달하는 기능을 수행한다. Analyzer는 다양한 분석 목적에 맞게 수집된 flow 정보를 분석하여 그 결과를 DB에 저장하는 역할을 수행하고, Reporter는 사용자의 요청에 의해 DB에 저장된 정보를 바탕으로 Traffic Report를 생성한다. 사용자의 요청은 웹을 통해 이루어지므로 언제 어디서든 네트워크를 통해 모니터링을 할 수 있다.

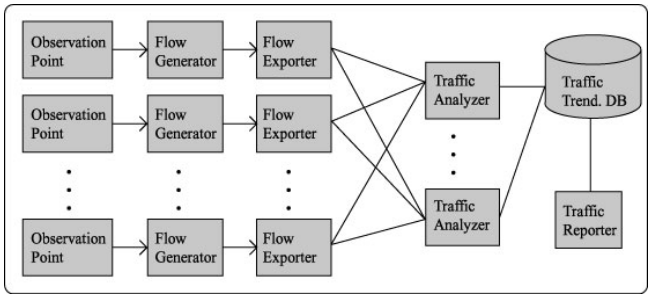

그림 1. 실시간 Traffic Analysis System 전체 구조

제안된 시스템 구조는 RTFM 구조를 기반으로 실제 네트워크에 적용 가능한 구조로 실현하였고, NG-MON의 복잡한 구조를 개선하여 저속에서 고속에 이르기까지 다양한 네트워크에 적용할 수 있는 장점이 있다.

2.2 Flow Generator

모니터링에 많이 사용되는 SNMP를 이용한 Traffic분석은 얻을 수 있는 정보가 SNMP agent가 제공하는 MIB 정보로 제한되기 때문에 모니터링에 필요한 모든 정보를 얻기는 힘들다. Flow Generator는 raw packet을 직접 수집함으로써 분석 시스템에 필요한 정보를 선택적으로 가져올 수 있는 유연함을 가진다.

IPFIX에서는 Flow Generator가 flow를 정의할 수 있고, 그에 따른 flow format을 flow template의 형태로 정의할 수 있는 유연함을 제공한다. 본 논문에서는 IPFIX의 방법에 따라 flow의 정의를 하였는데, 일반적으로 사용되는 5-tuple (src ip, dst ip, src port, dst port, protocol number) 정보가 같은 packet들을 집합[5]으로 정의한다. Flow Generator에서는 이를 기반으로 flow 정보를 생성하여 local 디스크에 저장한다. 그림 2는 본 논문에서 정의한 flow 정의에 따른 IPFIX flow template 내용이다.

Flow Generator에서는 flow정보를 1분 단위로 생성하여 저장하는데, 이는 Analyzer의 분석에 있어 1분 단위로 flow data를 전송받아 처리하는 것이 가장 효율적이기 때문이다. 예들 들어 Cisco NetFlow에서 사용하는 flow exporting scheme을 사용하는 경우 flow가 끝나는 시점에서 flow정보를 export 하는데, flow duration이 1분이 넘어가게 되는 경우 매분 bandwidth 측정이 어렵게 된다.

0	15	16	31
FlowSet ID = 0		Length = 52	
Template ID = 260		Field Count = 11	
0	sourceIPv4Address = 8	4	
0	destinationIPv4Address = 12	4	
0	sourceTransportPort = 7	2	
0	destinationTransportPort = 7	2	
0	octetDeltaCount = 1	4	
0	packetDeltaCount = 2	4	
0	flowStartSeconds = 150	4	
0	flowEndSeconds = 151	4	
0	protocolIdentifier = 4	1	
0	padding1 = 100	1	
0	padding2 = 101	2	

그림 2. 제안된 IPFIX Flow Template 세부 내용

독립된 Flow Generator는 모니터링 할 대상 링크의 개수와 트래픽 상황에 따라 시스템의 성능, 개수를 결정할 수 있는 유연함을 제공한다. 예를 들어 현재 본 논문에서 구현하여 설치한 Flow Generator는 최대 bandwidth가 100Mbps인 두 개의 링크를 모니터링하기위하여 1 GB memory, P-4 2 GHz CPU를 가진 머신 두 대를 이용하여 각각의 링크를 담당하게 하고 있다. 이 경우 두 대는 손실 없는 패킷의 수집이 가능하고, 평균 CPU 부하는 13%를 유지하였다.

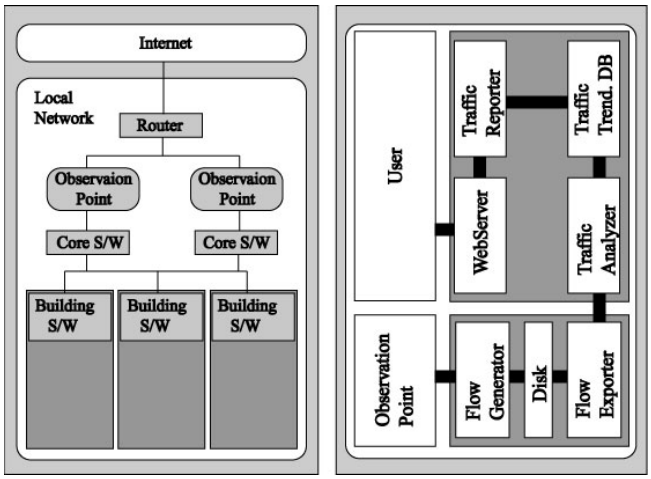

그림 3. 실시간 Traffic Analysis System 세부구조

본 논문에서 구축된 시스템은 Campus 네트워크에서 인터넷으로 오고가는 트래픽을 분석하는 목적으로 구축된 것으로 Flow Generator의 설치 지점을 선택함에 있어서 최상위 라우터가 아니라 그림 3과 같이 백본 스위치 2대와 상위 라우터와의 두 링크에서 수집하였다. 이유는 현재 Campus 네트워크에 구축된 라우터의 성능과 Flow Generator를 위해 가용한 시스템의 성능이 인터넷으로 나가는 모든 트래픽을 하나의 Flow Generator로 구축하기에는 미흡하였기 때문에 두 대의 Flow Generator 형태로 구축하였다. 이와 같이 Flow Generator를 독립된 시스템의 형태로 설계하는 것은 다양한 환경에 효율적으로 적용할

수 있는 장점이 있다.

2.3 Flow Exporter

Flow Exporter는 생성된 Flow 파일을 Traffic Analyzer의 요청에 의해 Analyzer로 전송하는 모듈이다. Ntop[7]의 경우 하나의 시스템에서 패킷의 수집에서 분석까지의 모든 작업을 수행하기 때문에 시스템 부하가 커 대용량 트래픽의 분석에 부적합하다. 그리고 하나의 시스템에서 동작하는 분석시스템은 여러 링크를 모니터링 해야 할 경우 유연하게 적용하기 힘든 단점도 가지고 있다. 또한, Header 정보나 Flow 정보를 각기 다른 사람이 각기 다른 목적으로 분석을 할 경우에도 한 대의 Machine에서 분석 시스템을 작동하면 많은 부하가 생긴다. 이와 같은 자원의 이용률 및 확장성을 고려하여 본 시스템은 Flow 정보의 수집과 분석을 하는 시스템을 따로 두었다.

2.4 Traffic Analyzer

Traffic Analyzer는 Flow Exporter에서 제공되는 flow 정보를 전송받아 목적에 맞게 분석하는 시스템이다. 분석 결과는 DB로 저장하는데, 이는 Reporter에서 분석결과를 다양한 형태로 나타내는데 가장 효율적인 방법이기 때문이다. 분석 요구에 따라 Analyzer는 다양한 형태로 그리고 동시에 여러 개의 Analyzer가 구축될 수 있다. Flow Exporter와 Analyzer로의 통신은 TCP를 사용함으로써 flow 정보의 안전한 전송을 보장한다.

2.5 Traffic Reporter

Traffic Reporter는 DB의 분석된 정보를 가져와 사용자에게 보여주는 모듈로서 웹 서버를 통해 사용자에게 모니터링 정보를 제공하는 역할을 수행한다. 웹으로 모니터링 정보를 제공함으로써 네트워크 관리자에게 편리함을 제공해 준다.

3. 시스템 구현

2장의 실시간 분석 시스템 구조를 바탕으로 Campus 네트워크의 모니터링을 할 수 있는 시스템을 개발하였다. Campus 네트워크 인터넷 링크에서 발생하는 트래픽의 사용자를 host별, subnet별로 사용량을 조사하여 어떤 host/subnet에서 download/upload를 가장 많이 하였는지를 실시간으로 분석해 주는 시스템이다. 그림 4는 이 시스템의 Web UI이다.

3.1 Flow Generator

Flow Generator에서 raw packet을 수집하기 위하여 pcap_library[8]를 사용하였다. Flow 정보를 효율적으로 생성하고 저장하기 위하여 Hash table을 사용하여 packet 정보로부터 flow 정보 생성시간을 효과적으로 하였다. Hash function은 shift folding function을 사용하였고, hash table 구조는 같은 hash key를 가진 flow들을 linked list로 연결하는 open hash table을 사용하였다.

3.2 Traffic Analyzer

Traffic Analyzer는 리눅스에서 제공하는 cron 데몬을 이용하여 1분 주기로 동작한다. Flow Generator가 Flow 파일을 생성하는 시간을 최대 2초 이내로 보고, 매분 2초에 Flow Exporter를 통해 Flow 파일을 가져온다. 분석은 Flow 정보 중 Source IP, Destination IP, Protocol, Bytes, Packets를 이용한다. Source IP와 Destination IP를 이용하여 로컬 네트워크의 IP와 외부 네트워크의 IP로 구분지어 준다. 로컬 IP를 기준으로 외부 네트워크로 송수신되는 Byte 수, Packet 수, Flow 수를 Hash Table을 이용하여 메모리에 저장한다. 메모리에 저장된 분석 정보를 Text 파일로 저장 후, Query 문을 통해 Traffic Trend. DB에 분 단위의 Table을 생성하고 저장한다. 모니터링 정보는 시간이 지나갈수록 정보의 가치가 떨어진다는 것과 Traffic Trend. DB의 효율적인 관리를 위하여 3시간 이전까지의 Table은 유지하고, 그 이전 시간의 Table은 삭제한다.

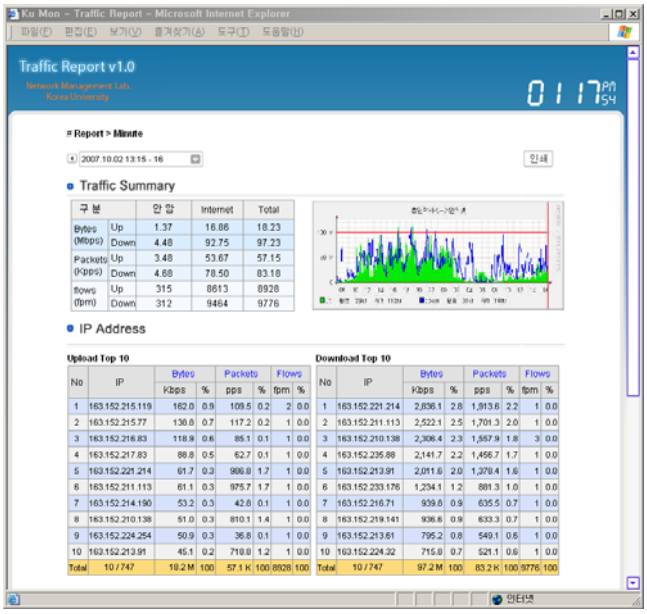

그림 4. Traffic Reporter Web User Interface

Traffic Analyzer는 1시간이 지나면 1분 단위 분석 데이터 60개를 모아 1시간 단위 데이터를 만들고, 하루가 지나면 1시간 단위 데이터 24개를 모아 1일 단위 분석결과를 만든다. 이러한 방식으로 1주 단위, 1달 단위 분석결과를 모아 DB에 저장한다. 저장된 데이터는 일정한 시간이 지나면 삭제함으로써 분석시스템의 디스크 용량을 일정한 범위 이하로 유지한다.

3.3 Traffic Reporter

Traffic Trend. DB의 분석된 정보를 가져와 웹으로 모

니터링 정보를 제공함으로써 네트워크 관리자에게 편리함을 제공해 준다.

일정주기 동안의 Traffic 통계 수치와 그래프, 다운로드와 업로드별로 Traffic이 가장 높은 10개의 Host와 Subnet을 보여준다. Byte 수, Packet 수, Flow 수 별로 순위를 정하여 보여줄 수 있으며, 백분율을 포함하여 전체 트래픽에서 각 Host가 얼마만큼의 비중을 차지하는지도 나타낸다. 또한, 보고서로 인쇄가 편리하도록 인쇄 웹 페이지를 만들어 우측 상단에 링크를 걸어두었다.

또한, 특정 IP를 관리하는 기능을 추가하여, 특정 IP에 대한 자세한 트래픽 추이를 제공할 수 있게 하였다. 관리자를 위한 특정 IP 입력이 가능하고, 특정 IP로 등록이 되면 특정 IP에 대한 일정시간 동안의 트래픽 추이가 수치와 그래프로 관리자에게 제공되어 진다.

3.4 System Specification

본 논문에서 개발한 Flow-based Real-time Traffic Monitoring and Analysis System은 학교 Campus 네트워크를 대상으로 표 1과 같은 개발 환경으로 개발되었다.

Hardware	P-4D-4.3GHz, 512MB, 80G HDD
OS	Linux Fedora 6
Web Server	Apache 2.2.3
Tool	RRD-Tool, Cron
Language	C, PHP 5
DB	MySQL

표 1. 개발 환경

Packet을 Capture하는 Machine은 Linux Fedora 6 환경을 다른 Machine은 Linux Fedora 7 환경에서 시스템이 구현되었다. Flow Generator, Flow Exporter, Traffic Analyzer가 C로 구현되었고, Traffic Reporter는 PHP를 사용하여 구현되었다. Traffic Analyzer는 리눅스의 Cron 데몬을 이용하여 매 1분마다 실행된다. DB는 MySQL을 사용하였으며, 웹 서버는 Apache 웹 서버를 사용하였다. 그리고 시간추이 그래프를 생성하기 위하여 RRD-Tool을 사용하였다. 구축된 시스템은 현재 Campus 네트워크의 모니터링을 위해 운영되고 있다.

4. 결 론

실시간 트래픽 모니터링은 네트워크 사용량이 크게 증가하고 인터넷을 통한 개인정보유출, 공격, 침해사고 등으로 인한 피해가 지속적으로 발생이 되고 있는 시점에서 효율적인 네트워크관리를 위해 반드시 필요하다. 본 논문에서는 기존의 모니터링 시스템의 단점을 개선하여 Enterprise 네트워크에 적합한 실시간 트래픽 분석 시스템의 구조를 제시하였고, 검증으로 Campus 네트워크를 대상으로 실시간 분석시스템을 설계하고 구현한 내용을 기술하였다. 제안된 시스템은 IPFIX flow 규정에 따라 flow format을 정의하였고, RTFM 구조를 기반으로 구성하여 확장성과 유연성을 향상시켰다. 본 연구에 관련된 몇 가지 향후 연구로는 다양한 분석 목적에 유연하게 적용될 수 있는 flow에 대한 정의, DoS/DDoS와 인터넷 웜과 같은 비정상 트래픽 탐지 그리고 분석 결과를 바탕으로 한 네트워크 트래픽의 적절한 제어를 통한 인터넷 품질 보장 등 다양한 분야에 연계시켜 시스템을 더욱 더 발전시킬 수 있을 것이다.

참고문헌

[1] Cisco Systems, White Papers, "Introduction to Cisco IOS NetFlow," May. 2007.

[2] Myung-Sup Kim, Hun-Jeong Kang, Seong-Cheol Hong, Seung-Hwa Chung, James W. Hong, "A Flow-based Method for Abnormal Network Traffic Detection", Accepted to appear in the Proc. of NOMS 2004, Seoul, Korea, April 2004.

[3] N. Brownlee, C. Mills and G. Ruth, "Traffic Flow Measurement: Architecture", IETF RFC 2722, October 1999.

[4] N. Brownlee, "Traffic Flow Measurement: Experiences with NeTraMet", IETF RFC2123, March 1997.

[5] Se-Hee Han, Myung-Sup Kim, Hong-Taek Ju and James W. Hong, "The Architecture of NG-MON: A Passive Network Monitoring System", LNCS 2506, DSOM 2002, October 2002, Montreal Canada, pp. 16-27.

[6] IETF Working Group IPFIX (IP Flow Information Export), http://www.ietf.org/html.charters/ipfix-charter.html.

[7] Luca Deri, Ntop, http://www.ntop.org.

[8] PCAP, http://www.tcpdump.org/pcap.htm.