애플리케이션 계층에서 트래픽 분석을 위해 부하 차단기를 적용한 네트워크 트래픽 모니터링 시스템

Network Traffic Monitoring System Applied Load Shedder to Analyze Traffic at the Application Layer

손 세 일* 김 흥 준** 이 진 영***
Sei-Il Son Heung-Jun Kim Jin-Young Lee

요 약

인터넷을 통해 전송되는 트래픽의 양이 지속적으로 증가하고 있기 때문에 네트워크 트래픽 모니터링 시스템이 모든 패킷을 실시간으로 분석하기는 어렵다. 피어-투-피어(P2P), 스트리밍 미디어, 메신저 등과 같이 동적으로 포트 번호를 할당받는 애플리케이션의 사용이 늘어나면서, 사용자들은 이들이 유발하는 트래픽을 분석하기를 원하고 있다. 이 같은 고수준의 분석을 위해서는 각 패킷마다 많은 처리 시간이 필요로 한다. 본 논문에서 부하 차단기를 이용하여 패킷의 수를 제한할 것을 제안한다. 선택된 패킷은 어떤 애플리케이션이 생성한 것인지 식별된 후, 정의된 애플리케이션 계층의 프로토콜에 따라 분석된다.

Abstract

As it has been continuously increased the volume of traffic over Internet, it is hard for a network traffic monitoring system to analysis every packet in a real-time manner. While it is increased usage of applications which are dynamically allocated port number such as peer-to-peer(P2P), steaming media, messengers, users want to analyze traffic data generated from them. This high level analysis of each packet needs more processing time. This paper proposes to introduce load shedder for limiting the number of packets. After it determines what application generates a selected packet, the packet is analyzed with a defined application protocol.

☞ Keyword : Network traffic monitoring, identification of application layer traffic, load shedding, stratified random sampling

1. 서 론

최근 전화망, 사설 전용선망, 위성방송망, 케이블망 및 무선 통신망 등 다양한 형태로 복잡하게 연결되었던 네트워크들이 광대역 통합 네트워크(BcN : Broadband convergence Network)로 통합이 진행되고 있다. 이 같은 네트워크를 기반으로 동작하는 애플리케이션들이 늘어나면서 네트워크를 하나의 자원으로 관리하기 위해 네트워크 트래픽 모니터링의 필요성이 증가하고 있다.

네트워크 트래픽 모니터링과 관련된 주요 문제를 살펴보면 다음과 같다[1]. 첫째, 초고속 네트워크를 통해 전송되는 막대한 양의 패킷들을 실시간으로 어떻게 모니터링하고 관리할 것인가에 대한 것이다[1,2,3]. 지속적으로 네트워크 트래픽이 증가하고 있는 현실을 고려할 때, 단순히 별도의 하드웨어 장비를 개발하고 운영하는 것으로는 해결하기 어렵다.

둘째, P2P, 스트리밍 미디어, 메신저 등과 같이 동적으로 포트를 할당받는 애플리케이션들의 트래픽을 어떻게 식별하고 지능적으로 분석할 것인가 하는 것이다[1,4,5]. 대표적 인터넷 트래

* 정 회 원 : 단국대학교 대학원 전산통계학과
seiilson@paran.com
** 정 회 원 : 진주산업대학교 컴퓨터공학부 부교수
thinkthe@jinju.ac.kr
*** 정 회 원 : 강남대학교 교양학부 조교수
goodman3@kangnam.ac.kr
[2005/11/28 투고 - 2005/12/02 심사 - 2005/12/05 심사완료]

픽인 HTTP, FTP, TELNET 등은 잘 알려진 포트(well-known port)를 통해 패킷들이 전송되기 때문에 이들의 트래픽을 식별하고 분석하기가 용이했으나 새로운 네트워크 애플리케이션들은 포트 번호가 동적으로 할당되기 때문에 이들로부터 발생되는 트래픽을 식별하고 분석하기가 어렵다. 최근 트래픽에 관한 연구는 P2P, 스트리밍 미디어, 메신저와 같은 애플리케이션들이 전체 네트워크 트래픽의 50%이상을 차지하고 있고, 그 비율이 지속적으로 증가하고 있음을 보여준다[6].

기존의 네트워크 트래픽 모니터링 시스템은 일정 시간 동안의 패킷 처리량과 오류 등에 관련된 데이터를 스위치 장비 내에 기록하고, 이를 주기적으로 접근하여 관리자가 필요한 정보를 제공하고 있다. 대부분의 라우팅 및 스위치 장비는 IP 계층까지의 정보를 제공하고 있으며, 그 이상 계층의 정보를 관리자에게 제공하지 못하고 있다. 하지만 Web, FTP와 같은 전통적 인터넷 애플리케이션뿐만 아니라 P2P, 스트리밍 미디어, 메신저 등과 같은 애플리케이션들이 발생하는 트래픽이 빠르게 증가함에 따라 네트워크 관리자들의 고수준 정보에 대한 요구 역시 증가하고 있다.

이 같은 요구를 만족시키기 위해 네트워크 트래픽 모니터링 시스템은 애플리케이션 계층의 프로토콜을 정의하고, 이들 트래픽을 식별할 수 있어야한다. 애플리케이션 트래픽의 식별은 패킷 헤더뿐만 아니라 payload에 대한 처리가 필요하기 때문에 패킷마다 많은 처리 시간이 필요하다. 실시간으로 모든 패킷들을 처리를 할 수 없기 때문에 부하 차단기가 필요하다.

본 논문에서는 이 같은 요구사항을 만족시키기 위해 데이터 스트림 관리 시스템(Data Stream Mangement System : DSMS)의 구조를 갖는 네트워크 트래픽 모니터링 시스템의 프로토타입을 개발하였다. 개발된 시스템은 트래픽 분석을 위한 별도의 하드웨어를 필요로 하지 않으며, 시스템의 처리 능력에 따라 입력되는 패킷들로부터 부하 차단기를 통해 처리할 패킷을 선택하기 때문에 과부하시 이루어지는 단순한 패킷 제거 방식보다 정확한 통계정보를 제공한다. 또한 동적 포트를 이용하는 애플리케이션들의 트래픽에 대한 정보를 제공할 수 있는 장점이 있다.

본 논문의 구성은 다음과 같다. 2장에서는 네트워크 트래픽 분석과 부하 차단에 관한 기존 연구를 살펴보고, 3장에서는 개발된 네트워크 모니터링 시스템의 모델과 부하 차단 그리고 애플리케이션 트래픽 식별 방법을 알아본다. 4장에서는 부하 차단을 위해 사용한 층화 랜덤 추출법의 정확성을 알아보고, 5장에서는 결론과 향후 연구과제에 대해 기술한다.

2. 관련 연구

2.1 네트워크 트래픽 모니터링

NTop[7]은 웹 기반의 트래픽 모니터링 및 분석 애플리케이션으로써 호스트, 프로토콜, 애플리케이션에 따라 트래픽을 실시간으로 분석한다. NTop은 패킷 단위 분석을 하며, 호스트 정보를 저장하기 위해 해시 테이블을 사용한다. 해시 테이블 관리를 통해 해시 테이블의 항목 수와 비례하여 메모리 사용량 및 패킷 처리 시간 증가를 방지하고 있다. 또한 NTop은 트래픽 분석 시, 시간을 소요하는 반복 작업을 최소화하기 위해 2단계 캐싱을 사용한다. NTop은 장기간에 걸친 트래픽 분석에는 적합하지 못하다.

FlowScan[8]은 NetFlow[9] 포맷의 flow 데이터를 분석하고 보고하는 소프트웨어 패키지이다. NetFlow 포맷은 효용성이 우수하여 많은 장비 제조업체가 지원하고 있는 flow 데이터 포맷이다. FlowScan은 cflowd, flowscan, RRDtool 세 부분으로 구성되는데 cflowd는 패킷을 저장하고 샘플링 주기마다 파일에 타임스탬프를 부여한다.

flowscan은 관리자가 선택한 프레젠테이션 모듈을 실행하는 perl 스크립트이고, RRDtool은 주기적으로 데이터를 저장하고, 요약된 트래픽 데이터를 생성한다. FlowScan은 다양한 기간의 트래픽 분석이 가능하지만 트래픽양이 폭주하는 경우 정상적으로 동작하지 못하는 단점이 있다.

2.2 부하 차단

데이터 스트림 관리 시스템에서 입력 스트림을 통해 처리량 이상의 데이터가 수신되는 과부하 문제를 해결하기 위한 부하 차단과 관련된 연구는 난수 기반 방법과 의미 기반 방법으로 구분된다. 난수 기반 방법은 입력 스트림을 통해 전달되는 데이터를 무작위로 버리는 것으로 STREAM(the STanford stREam datA Manager)[10]을 예로 들 수 있다. STREAM은 데이터 스트림 상에서 연산자와 자원을 공유하는 다수의 집합 질의들을 효과적으로 처리하기 위한 난수 기반 부하 차단기 위치 선정에 대해 연구하였다.

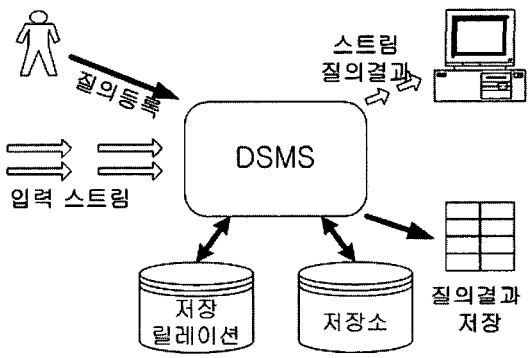
의미 기반 부하 차단 방법은 데이터 내용의 중요도를 기반으로 과부하시 중요도가 낮은 데이터를 버리는 것으로 Aurora[11]를 예로 들 수 있다. Aurora의 QoS 명세를 이용하며, 이 QoS 명세는 질의들 간의 우선순위가 아니라 질의가 접근하는 데이터들 간의 유용도를 기반으로 우선순위를 부여하고 낮은 우선순위의 데이터부터 버린다.

Das[12]에서는 다수의 데이터 스트림들 간의 조인 질의에 대한 의미 부하 제한 알고리즘을 제안하였다. 제한적인 자원과 질의 결과에 대한 오차 측정 방법이 주어질 때 과부하 발생시 조인 질의의 평균 오차율이 감소된다.

3. 기본 모델

이 절에서는 개발된 네트워크 트래픽 모니터링 시스템의 기본 구조인 DSMS에 대해 소개하고, 부하 차단기의 설치 위치에 대해 알아본다. 부하 차단기란 본질적으로 입력되는 패킷들로부터 표본을 추출하는 것이기 때문에 표본 추출 방법에 따라 산출되는 통계치의 정확도가 달라진다. 본 논문에서 이용한 층화 랜덤 추출법에 대해 설명하고, 애플리케이션 트래픽의 식별 방법에 대해 기술한다.

3.1 DSMS와 부하 차단기

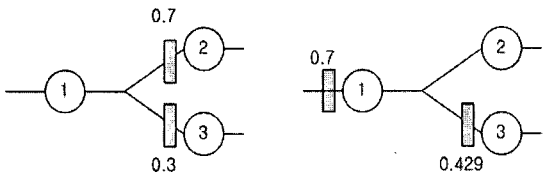


〈그림 1〉 DSMS 개요

그림 1은 DSMS의 구성 및 동작에 대한 개요를 보여준다. 사용자가 질의를 DSMS에 등록하면, DSMS는 입력되는 데이터 스트림과 저장된 릴레이션을 이용하여 질의 결과를 출력한다. DSMS에서 사용자가 등록한 질의는 실시간 연속 질의로 처리되며, 이것은 질의 처리가 한 번의 수행으로 종료되는 것이 아니라 입력 스트림을 통해 지속적으로 전달되는 데이터를 이용하여 질의가 연속적으로 처리되는 것을 의미한다. 빠른 속도로 입력되는 데이터를 실시간으로 처리하기 위해 DSMS는 많은 자원을 필요로 한다[10,11]. DSMS의 처리 능력이 입력되는 데이터를 실시간으로 처리하기에 충분하지 않거나 또는 순간적으로 예상하지 못한 대량의 데이터가 입력 스트림을 통해 전달되면, DSMS는 등록된 질의들의 결과 출력을 위해 아직 처리하지 않은 일부 데이터를 버리는데, 이것을 부하 차단(load

shedding)이라 한다.

그림 2는 부하 차단기의 설치 위치에 따라 처리 시간이 감소될 수 있음을 보여준다. 그림에서 원은 입력된 패킷 처리 모듈을 나타내며, 직사각형은 부하 차단기를 나타낸다. 그림 2의 a는 패킷의 처리를 위해 여러 모듈을 통과하는 다단계로 진행될 경우 각 단계별로 정의된 비율에 따라 부하가 차단되는 모습을 보여준다. 예를 들어, 모듈 1의 입력으로 100개 데이터가 도착하고, 모듈 1이 결과로 90%의 데이터를 출력한다면, 모듈 2로 63개, 모듈 3으로 27개의 데이터가 입력되어 세 모듈이 처리하는 데이터는 총 190개가 된다. 그림 2의 b와 같이 부하 차단을 먼저 실행하면, 모듈 1은 70개, 모듈 2는 63개, 모듈 3은 27개로 160개의 데이터를 처리한다. 그림 2의 b는 모듈에 입력되는 데이터가 부하 차단기를 통해 줄어들어 다음 단계의 입력되는 데이터의 총 수가 감소되는 것을 보여준다. 따라서 부하 차단은 패킷에 대한 처리 이전에 실행되는 것이 시스템의 성능 향상에 도움이 된다.



(a) 부하 차단기가 뒤에 위치 (b) 부하 차단기가 앞에 위치

〈그림 2〉 부하 차단기의 위치

3.2 층화 랜덤 추출법

과부하시 입력되는 패킷을 단순 랜덤 추출법 이용하여 표본을 추출할 경우 서로 이질적인 프로토콜들을 대상으로 표본을 추출하기 때문에 추정량의 분산이 커지는 경향이 있다. 이 같은 단점을 해결하기 위해 본 논문에서는 빠른 속도로 표본을 추출하면서, 추정량의 정확도를 높일 수 있는 층화 랜덤 추출법(stratified random sampling)을 사용하였다. 층화 랜덤 추출법의 장점은 각 층별로 추정값을 계산할 수 있으며, 또한 표본이 모든 층에 분포되어 있으므로 모집단을 대표하는 표본을 추출할 가능성이 높다.

추출 단위 N개인 모집단이 h개의 층으로 나누어 졌을 때, k번째 층의 크기를 N_k, 평균과 분산을 각각 μ_k, σ_k^2 라 하자. 또한, 각 층의 크기가 n_k인 표본을 단순 랜덤 추출법으로 추출하면 모평균에 대한 추정량과 분산은 다음과 같다[13].

$$\text{평균: } \overline{X_{st}} = \frac{1}{N}(N_1\overline{X_1} + \ldots + N_h\overline{X_h}) \quad (1)$$

$$\text{분산: } Var(\overline{X_{st}}) = \frac{1}{N^2}\sum_{k=1}^{h} N_k^2\left(\frac{N_k - n_k}{N_k}\right)\frac{\sigma_k^2}{n_k} \quad (2)$$

$$\text{표준오차: } \widehat{SE}(\overline{X_{st}}) = \sqrt{\frac{1}{N^2}\sum_{k=1}^{h} N_k^2\left(\frac{N_k - n_k}{N_k}\right)\frac{S_k^2}{n_k}} \quad (3)$$

($\overline{X_k}$: k번째 층에서 추출한 표본 평균, S_k^2: k번째 층에서 추출한 표본 분산)

층화 랜덤 추출법을 이용하는 경우, 각 층으로부터 얼마만큼의 표본을 추출할 것인지를 결정하는 것은 추정량의 정확도와 밀접한 관계가 있다. 본 논문에서 과부하 발생시 부하 차단 방법으로 층화 랜덤 추출법을 이용하는 주된 목적은 추정량의 분산을 최소화함으로써 추정량의 정확도를 높이는데 있다. 그러므로 각 층의 분산과 크기를 알고 있다면 분산과 층의 크기가 큰 경우에 많은 표본을 추출하고, 분산과 층의 크기가 작은 경우 적은 수의 표본을 추출하는 것이 추정량의 정확도를 증가시킨다.

층화 랜덤 추출에서 각 층에 몇 개씩의 표본을 배분할 것인가는 추정량의 정도를 결정하는 중요한 요소이다. 네트워크 트래픽 모니터링 시스템이 적정한 처리 능력을 갖는 컴퓨터에서 실행되고 있다면, 과부하 상태는 일시적이며 대부분의 경우 모든 패킷을 처리 가능한 상태 즉,

정상 부하 상태이다. 네트워크 트래픽 모니터링 시스템이 정상 부하 상태일 때, 포트 번호별로 전송되는 패킷의 수와 같은 정보를 저장하다가 과부하 상태가 되면 저장된 정보를 바탕으로 포트마다 적절한 비율(서로 다른 비율)로 패킷 표본을 추출하고 처리한다.

정상 부하 상태에서 입력된 전체 패킷의 수가 N이고, 추출할 표본 패킷의 수를 n이고,목적지 포트 번호 k로 입력된 패킷의 수를 N_k라 하면, 과부하시 목적지 포트 번호 k를 갖는 패킷이 선택될 확률 p_k는 $p_k = n\left(\frac{N_k}{N}\right)$이 된다.

3.3 애플리케이션 트래픽의 식별

기존의 클라이언트/서버 방식 애플리케이션은 고정된 포트 번호로 하나로 TCP 또는 UDP 세션을 사용하여 통신을 해왔다. 하지만 P2P, 스트리밍 미디어와 같이 동적으로 포트를 할당받는 애플리케이션들은 다수의 세션들을 사용하여 통신하기 때문에 이들의 트래픽을 식별하는 것은 매우 어렵다.

네트워크 트래픽 모니터링과 분석을 위해서는 이들 애플리케이션의 트래픽을 식별하고, 프로토콜을 인식함으로써 애플리케이션의 동작을 모니터링 할 수 있는 방법이 필요하다. 애플리케이션의 트래픽을 식별하기 위해 이들 애플리케이션이 주로 사용하는 포트 번호를 정의하였다. 동적으로 포트를 할당받는 애플리케이션의 주요 포트에 대한 정의가 가능한 것은 이들이 서로 간의 최초 연결을 위해 고정된 포트 번호를 통해 접속하고 관련된 메시지를 전송하는 형태로 동작하기 때문이다. 애플리케이션별로 사전에 정의된 포트로 메시지를 전송하는 컴퓨터의 IP 주소를 추출하여 지정된 시간 내에 1024 이상의 포트 번호를 가지면서, 사전에 정의되지 않은 포트를 이용해 전달되는 패킷은 이들 애플리케이션이 발생한 것이라 가정했다.

먼저 각 애플리케이션들이 사용하는 주요 포트들을 구분하기 위해, PC에 이들 프로그램을 구동시키고, 이들이 사용하는 포트 번호를 관찰하는 동시에 PC가 접속한 네트워크를 통해 전송되는 모든 패킷들을 캡쳐하고 분석하였다. 특히 관심을 갖은 것은 파일 공유를 목적으로 하는 P2P 애플리케이션들로써 이들은 네트워크에 많은 트래픽을 유발하고 있다. 표 1는 인터넷 애플리케이션들이 사용하고 있는 주요 포트들을 정리한 것이다.

〈표 1〉 애플리케이션들이 사용하는 주요 포트 번호

애플리케이션	주요 포트 번호
WWW	80, 8080
FTP	21
eDonkey	4661, 4662, 6667
소리바다	3938, 7675, 7676, 7677, 22321, 22322
고부기	5325
MSN Messenger	1863, 6981-6990, 14594

4. 평가

이 절에서는 부하 차단을 위한 방법으로써 층화 랜덤 추출법과 단순 랜덤 추출법의 정확성을 비교하였다. 평가를 위해 실험 데이터 전체를 모집단으로 하여 각 포트별로 전송된 패킷 크기의 평균과 표준 편차를 구하고, 층화 랜덤 추출법과 단순 랜덤 추출법으로 1%, 5%, 10%, 50%의 비율로 표본을 뽑아 패킷 크기의 평균과 표준 편차를 구하여 이를 모집단으로부터 계산된 값과 비교하였다.

실험에 사용된 데이터는 2003년 8월 14일 오후 9시 14분부터 2003년 8월 15일 오전 12시 51분 사이에 IP 주소의 프리픽스가 "203.237.225"을 갖는 네트워크로 전송된 패킷들 중 목적지 포트 번호가 0, 3938, 5325, 7674 인 패킷들

로 데이터의 수는 1,000,000개이다.

먼저 1,000,000개의 자료로부터 포트별로 전송된 패킷의 크기에 대해 분석한 내용은 표 2과 같다. 0과 3938 포트로 전송된 패킷의 비율은 1%로 이하로 적으며, 대부분의 패킷은 5325 포트로 전송되었음을 보여준다.

〈표 2〉 실험 데이터의 포트별 패킷 전송 비율

포트번호	패킷 수	비율(%)
0	5,636	0.5636
3938	9,755	0.9755
5325	941,942	94.1942
7674	42,667	4.2667
총계	1,000,000	100.0000

〈표 3〉 모평균과의 차

포트 번호	1%		5%		10%		50%	
	단순	층화	단순	층화	단순	층화	단순	층화
0	-2.4090	-2.4376	6.3812	1.9537	1.7010	0.1470	-0.3314	-0.0931
3938	-26.0057	-0.5756	-5.1309	-2.0724	-1.4863	3.7545	1.0856	-0.2640
5325	0.0176	-0.0416	-0.2368	0.0622	0.0592	0.0972	0.0649	-0.0922
7674	0.3102	0.1473	0.1116	0.1344	-0.0159	-0.0404	-0.0300	0.0193

〈표 4〉 모표준편차와의 차

포트 번호	1%		5%		10%		50%	
	단순	층화	단순	층화	단순	층화	단순	층화
0	-3.8353	-5.5299	8.7223	2.5432	1.8908	0.3934	-0.4553	-0.1113
3938	-50.4973	-23.4234	-11.7715	-3.1092	-1.7432	6.4333	2.1930	-1.1155
5325	-0.6643	0.4146	0.2339	0.4289	0.1794	0.6847	0.1298	-0.3945
7674	0.1668	0.0874	0.1456	0.0541	-0.0057	-0.0330	-0.0036	0.0166

표 3는 포트별로 전송된 패킷의 크기를 알기 위해 1%, 5%, 10%, 50%의 다른 비율로 단순 랜덤 추출을 이용해 계산된 결과와 층화 랜덤 추출을 이용해 계산한 결과의 정확도를 비교하기 위해 모평균과의 차를 구한 것이다. 표 2과 표 3를 보면 패킷이 0, 3938, 7674 포트와 같이 상대적으로 적은 비율로 전송될 때, 층화 랜덤 추출을 이용해 구한 패킷의 크기가 단순 랜덤 추출을 이용해 구한 것보다 정확함을 알 수 있다. 하지만 5325 포트와 같이 대부분의 패킷이 특정 포트로 전송되는 경우 단순 랜덤 추출을 통해서도 모집단의 통계량과 근사한 값을 구할 수 있으므로 높은 비율로 표본 패킷을 추출하는 경우 두 추출 방법 사이에 정확도의 차이가 없다.

표 4은 단순 랜덤 추출을 통해 구한 패킷 크기의 표준 편차와 층화 랜덤 추출을 통해 계산된 패킷 크기의 표준 편차를 모표준편차와 비교한 것이다. 이 표를 보면 상대적으로 적은 비율로 전송되는 포트들에 대해 층화 랜덤 추출이 보다 정확한 결과를 산출함을 알 수 있다. 또한 표본 추출 비율이 적은 경우, 예를 들어 5%와 10%로 비율로 표본 추출한 경우를 보면, 층화 랜덤 추출이 단순 랜덤 추출에 비해 모표준편차와의 차가 적기 때문에 표본 추출 비율이 적은 경우 정확도를 높일 수 있는 방법임을 알 수 있다.

5. 결론 및 향후 연구과제

인터넷의 초고속화로 인해 단위 시간 당 전송되는 패킷의 양이 급속히 증가하고 있다. 이 같은 증가의 상당 부분은 P2P, 스트리밍 미디어, 메신저와 같은 애플리케이션의 사용이 늘어났기 때문이다. 이들은 동적 포트를 이용하기 HTTP, FTP와 같은 기존의 트래픽과 비교하여 트래픽 분석에 어려움이 많다.

본 논문에서는 네트워크 트래픽을 분석하기 위해 특별한 하드웨어를 도입하지 않고, 부하 차단기를 이용해 입력되는 패킷들로부터 표본을 선택하여 분석하는 프로토타입 시스템을 구현하였다. 구현된 시스템은 보다 정확한 정보를 산출하기 위해 층화 랜덤 추출법을 사용하였다. 이것은 애플리케이션들마다 발생시키는 네트워크 트래픽의 양이 다르기 때문에 많은 패킷이 전송되는 애플리케이션부터는 적은 비율의 표본을 추출하여 필요한 정보를 구하고, 적은 양의 패킷을 전송하는 애플리케이션으로부터는 상대적으로

많은 수의 표본을 추출하여 정보를 구함으로써 사용자에게 보다 정확한 정보를 제공할 수 있었다.

P2P, 스트리밍 미디어, 메신저와 같이 동적으로 포트를 할당받는 애플리케이션들의 트래픽을 식별하기 위해 최초 연결을 위해 이용하는 포트를 선호 포트로 등록하고, 이를 감시함으로써 이들 애플리케이션이 유발하는 트래픽을 식별하였다. 그 결과 단말 노드에서도 해당 네트워크에서 발생되는 다양한 트래픽들을 분류하고 분석할 수 있었다.

또한 단순히 포트 번호만으로 애플리케이션 트래픽을 식별하고 분석하는 것이 아니라 P2P, 스트리밍 미디어, 메신저들이 사용하는 애플리케이션 계층의 프로토콜을 정의하고, 이를 이용해 접속, 질의, 전송 요청 등과 같은 고수준의 분석도 가능하다. 이 같은 분석을 위해서는 패킷의 payload를 처리하기 위해 많은 처리 시간이 필요하기 때문에 부하 차단기가 있어야만 한다.

본 논문에서는 애플리케이션 계층의 트래픽 분석을 위해 부하 차단기를 적용한 네트워크 트래픽 모니터링 프로토타입 시스템을 소개하였다. 사용자들이 요구하는 고수준의 트래픽 분석 정보를 제공하기 위해 필요한 정보를 추출할 수 있는 질의를 등록하고 지속적으로 정보를 제공할 수 있도록 질의 처리 부분의 개선이 필요하다. 특히, 애플리케이션 계층의 프로토콜의 처리 시간을 줄이기 위한 추가 연구가 필요하다.

참 고 문 헌

[1] Myung-Sup Kim, Young J.Won, James Won-Ki Hong, "Application-Level Traffic Monitoring and an Analysis on IP Networks", ETRI Journal Vol. 27, No.1, February 2005.

[2] A. Moore, J. Hall, C. Kreibich, E. Harris, and I. Pratt, "Architecture of a network monitor", Proc. of the Passive and Active Measurement Workshop (PAM), 2003.

[3] C. Fraleigh, S. Moon, C. Diot, B. Lyles, and F. Tobagi, "Architecture of a passive monitoring system for backbone IP networks", Technical Report TR00-ATL-101-801, Sprint Advanced Technology Laboratories, October 2000.

[4] K. Sripanidkulchai, B. Maggs, H. Zhang, "An analysis of live streaming workloads on the internet", Proc. of the 4th ACM SIGCOMM, pp.41-54, 2004.

[5] T. Karagiannis, A. Broido, M. Faloutsos, K. claffy, "Transport Layer Identification of P2P Traffic", Proc. of the 4th ACM SIGCOMM, pp.121-134, 2004.

[6] S. Sen, J. Wang, "Analyzing peer-to-peer traffic across large networks", Internet Measurement Conference(IMC), Proc. of the 2nd ACM SIGCOMM Workshop on Internet measurement, pp 137-150, 2002.

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

[8] D. Plonka, FlowScan, http://net.doit.wisc.edu/~plonka/FlowScan/

[9] D. W. McRobb, "cflowd design," CAIDA, September 1998.

[10] B. Babcock, M. Datar, R. Motwani, "Load Shedding for Aggregation Queries over Data Streams", Proc. of the 20th ICDE, pp. 350-361, 2004.

[11] N. Tatbul, U. Çetintemel, S. Zdonik, M. Cherniack, M. Stonebraker, "Load Shedding in a Data Stream Manager", Proc. of the 29th VLDB, pp. 309-320, 2003.

[12] A. Das, J. Gehrke, M. Riedewald, "Approximate join processing over data streams", Proc. of ACM SIGMOD, pp.40-51, 2003.

[13] 김우철, 김재주, 박병욱, 박성현 외 , "통계학개론", p.323, 영지문화사, 1992.

◐ 저 자 소 개 ◑

손 세 일 (Sei-Il Son)
1993년 유한전문대 전자계산과 졸업
1997년 방송통신대학교 전자계산학과 졸업(학사)
1999년 단국대학교 대학원 전산통계학과 졸업(석사)
2002년 단국대학교 대학원 전산통계학과 수료(박사)
1993~1996 상지전산(주) 개발부
2002~2006 단국대학교 정보컴퓨터학부 강의전임강사
관심분야 : P2P, IPTV, 스트림 데이터, 트래픽 분석, 데이터베이스, 전자상거래, etc.
E-mail : seiilson@paran.com

김 흥 준 (Heung-Jun Kim)
1989년 단국대학교 전자계산학과 졸업(학사)
1993년 단국대학교 대학원 전산통계학과 졸업(석사)
1999년 단국대학교 대학원 전산통계학과 졸업(박사)
1999~현재 진주산업대학교 컴퓨터공학부 부교수
관심분야 : 컴퓨터 구성, 모바일 네트워킹, 전자상거래, etc.
E-mail : thinkthe@jinju.ac.kr

이 진 영 (Jin-Young Lee)
1995년 단국대학교 대학원 전산통계학과 졸업(석사)
1998년 단국대학교 대학원 전산통계학과 수료(박사)
1999~현재 강남대학교 교양학부 조교수
관심분야 : 이미지 프로세스, 생체인식, 침입탐지 시스템, etc.
E-mail : goodman3@kangnam.ac.kr