인터넷 응용 트래픽의 분석 및 동향

최진섭°, 이사야, 백현호, 정중수, 윤승현", 정태수"
안동대학교 공과대학 정보통신공학과, ETRI 인터넷 트래픽 연구실"
{lee75°, singer, baekh}@anuis.andong.ac.kr, jschung@andong.ac.kr, {hun20", jung03"}@hotmail.com

Analysis and Trends of Internet Application Traffics

Jin-Sub Choi°, Sa-Ya Lee, Joong-soo Chung, Hyun-ho Baek, Seong-Hyun Yun", Tae-Soo Chung"
Dept. of Information Communication Engineering, Andong National University

요 약

오늘날 인터넷의 백본 발달과 더불어 수많은 응용 서비스들이 사용되고 있다. 이러한 응용 서비스는 인터넷 초기 출현 시에는 웹, 파일전달, 이메일 등의 well-known 서비스가 주축을 형성하였다. 그러나 최근 인터넷의 폭발적인 사용과 다양한 컨텐츠의 요구로 unwell-known 서비스가 매우 많이 등장하였다. 또한 인터넷에서 작동하는 트래픽을 모니터링하여 (un)well-known 포트를 사용하는 패킷의 PDU 정보를 보고서 응용 서비스의 유형을 찾는 기법은 트래픽 분석자에게 매우 유용한 정보이다.

본 논문에서는 TCP와 UDP 위에서 동작하는 (un)well-known 포트를 사용하는 패킷의 PDU 정보에 의한 응용 서비스의 유형을 찾는 트래픽 분석 기법을 수행하였다. 이러한 분석을 위하여 수많은 트래픽 중 활용도가 많은 응용 서비스를 추출하기 위하여, 안동대학교 네트워크에서는 ethereal에서 제시된 netflow 및 tcpdump 기법을 활용하였다. 추출된 트래픽의 분석을 위하여 그 서비스를 PC에서 구동시켜 ethereal 트래픽 분석장치로 모니터링하여 분석하였다.

I. 서 론

오늘날 인터넷의 백본망과 라우터의 급속한 발달과 더불어 수많은 응용 서비스들이 사용되고 있으며, 향후에도 다양한 컨텐츠가 요구되고 있다. 한편 종단 사용자 관점에서 살펴보면 이더넷 LAN 카드를 장착한 사용자 수는 LAN 뿐만 아니라 공중망의 ADSL(Asynchronous Digital Subscriber Loop) 까지 확대되어 가는 추세이다.

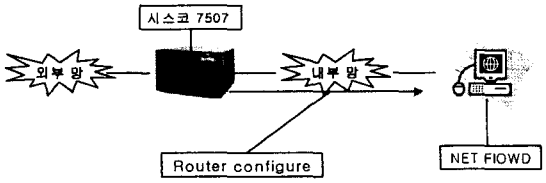
인터넷 초기에는 TCP(Transmission Control Protocol)나 UDP(User Datagram Protocol) 프로토콜을 활용하는 응용 서비스 중 웹, 파일전달, 이메일 등의 well-known 포트를 사용하는 서비스가 주축을 형성하였다. 그러나 최근 인터넷의 폭발적인 사용과 다양한 컨텐츠의 요구로 unwell-known 포트를 사용하는 서비스가 매우 많이 등장하였다. 다양한 서비스의 등장과 네트워크의 발달로 인터넷상에서 동작하는 프로토콜의 면밀한 분석을 위하여 국내외에서는 프로토콜 분석 장비가 많이 출시되고 잇다[1,2,3]. 이러한 프로토콜 분석 장비는 대다수 TCP/IP 프로토콜 슈트만 분석하고 있다. 아울러 많은 네트워크 관련 기관들은 앞 다투어 새로운 응용 서비스의 출현으로 unwell-known 서비스의 속성을 제시하고 있다[4]. 일반적으로 unwell-known 서비스는 TCP나 UDP의 well-known 포트번호 사용범위를 제외한 임의의 포트번호를 사용한다. 최근에는 방화벽 등의 공격을 피하기 위해 well-known 포트번호를 활용하여 임의의 서비스를 제공하기도 하는 실정이다.

본 논문에서는 TCP와 UDP 위에서 동작하는 (un)well-known 포트를 사용하는 패킷의 PDU 정보에 의한 응용 서비스의 유형을 찾는 트래픽 분석 기법을 수행하였다. 이러한 분석을 위하여 수많은 트래픽 서비스 중 활용도가 많은 응용 서비스를 추출하기 위하여, 안동대학교 네트워크에서 ethereal에서 제시된 netflow[5] 및 tcpdump[6] 기법을 활용하였다. 추출된 응용 트래픽의 분석을 위하여 ethereal 트래픽 분석장치[3]를 활용하였다. 분석을 위해 사용된 환경으로는 펜티엄 III 프로세서 800 MHz PC 기반의 Linux 기반 OS를 축으로 하였다.

II. 분석환경

국립 안동대학교 전산망에서 분석하였으며, 본 교의 전산망은 외부 인터넷 망과 DS3 급의 전송로를 통한 시스코 7507 라우터 장비와 접속된다. 내부망은 다시 기가비트 이더넷 망으로 서브넷과 접속 되고 있으며, 안동대학교 전산망을 통한 트래픽 분석을 위하여 netflow와 tcpdump 기법을 사용하였다. netflow 기법은 7507 라우터에서 수집한 패킷들의 정보를 내부망을 통해 분석하는 PC에게 전달해 준다. tcpdump 기법은 7507 라우터에 허브를 우선 접속한다. 이후 허브에 임의의 한포트를 할당하여 트래픽 분석용 PC를 접속시킨다. 따라서 tcpdump 기법은 netflow 기법보다 번잡하나, 패킷 트래픽의 원시 데이터까지 추출할 수 있는 장점이 있다.

1. netflow를 이용한 분석 환경 설정



netflow을 이용한 트래픽 측정방법은 라우터가 일정 시간동안 통계를 낸 트래픽을 PC에서 받는 것이다. 우선 라우터의 환경 설정을 수행한다. 라우터는 다음과 같은 명령을 내려서 global configuration 모드로 들어간다.

```
# configure
# ip cef
(config)# interface ATM1/0 (설정하고 싶은 인터페이스)
(config-if)# ip route-cache flow
(config-if)# exit
(config)# ip flow-export destination 203.255.255.248 2055
          (cflowd가 설치된 시스템의 IP주소)
(config)# ip flow-export version 5 peer-as
          (netflow 버전이 5이고 AS는 인 접한 AS에 대한 정보를 얻고자 함.
           만약 원래의 소스 정보를 얻으려면 origin으로 설정)
```

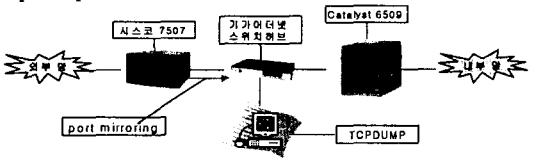
통상 어떤 인터페이스에 netflow를 활성화하면 해당 인터페이스로 입력되는 트래픽에 대해서만 netflow 패킷을 생성해 준다. 따라서 그 인터페이스로 나가는 트래픽을 같이 보려면 여타 모든 다른 인터페이스에 대해서 netflow기능을 활성화 해주어야 한다. netflow 기법을 활용하기 위해서는 우선 cflowd를 설치하여야 한다. 이를 위해서는 우선 FreeBSD를 설치한 후 http://www.caida.org의 arts++와 cflowd 다운받아 수행한다. 설치가 완료되면 통상 설치된 파일은 /usr/local/arts 아래에 존재 한다. Cflowd 설치 후에는 설정 파일을 환경에 맞도록 수정을 해 주어야 한다. Cflowd가 설치된 디렉토리로 가면 etc디렉토리가 있고 그 아래에 보면 cflowd.conf.example와 cfdcollect.conf.example가 있으며 이것을 각각 cflowd.conf와 cfdcollect.conf로 고친 후 파일 내부를 수정한다.

```
flowd.conf 설정방법
OPTIONS {
  LOGFACILITY:        local6
  TCPCOLLECTPORT:     2056
  PKTBUFSIZE:         1048576
  TABLESOCKFILE:
    /usr/local/arts/etc/cflowdtable.socket
  FLOWDIR:
    /usr/local/arts/data/cflowd/flows
  FLOWFILELEN:        1000000
  NUMFLOWFILES:       10
  MINLOGMISSED:       1000
}

                                  COLLECTOR {
                                    HOST:          211.248.0.136
                                    # IP address of central collector
                                    ADDRESSES: { 211.248.0.136 }
                                    AUTH:          none
                                  }

                                  CISCOEXPORTER {
                                    HOST:          211.248.0.254
                                    ADDRESSES: { 211.248.0.254 }
                                    CFDATAPORT: 2055
                                    SNMPCOMM: 'public'
                                    LOCALAS:   0
                                    COLLECT:
                                      { protocol, portmatrix, netmatrix, flows }
                                  }
```

2. tcpdump를 이용한 분석 환경 설정



우선 라우터의 송, 수신 트래픽을 측정하기 위해서는 라우터나 그에 접속된 이더넷상의 허브등의 장치를 통해 PC에서 라우터의 트래픽을 측정한다. 이러한 환경은 http://www.tcpdump.org에서 pcap library와 tcpdump를 다운받아 설치한다. 추후 분석을 위해 파일로 출력 하려면 통상 다음과 같이 옵션을 주어서 캡쳐한다.

1) tcpdump의 명령어 구성

tcpdump의 일반적인 명령어 구성은 다음과 같다.

```
tcpdump [-adeflnNOpqRStuvxX ][ -ccount ] [ -Cfile_size][ -Ffile ]
[ -iinterface][ -m module ] [ -rfile ][ -ssnaplen][ -Ttype ][ -wfile]
[ -Ealgo:secret][ expression ]
```

제일 마지막인 [expression]은 분석자가 원하는 파일을 생성하는데 중요한 옵션이다. 분석자는 이 옵션을 어떻게 주느냐에 따라서 생성되는 파일이 달라질 수 있다.

2) 원시 데이타의 수집 방법

통상 다음과 같은 명령으로 원시 데이터를 수집할 수 있다.

```
tcpdump -enx -s 1500 -w (file name)
```

III. 분석 방법

1. netflow를 이용한 분석 방법

netflow는 많이 사용되는 트래픽의 포트번호를 구하는 방법만 제시된다.

1 단계. artsportms를 이용해 포트별 데이터로 변환한다

```
usage: artsportms [-p] [-s srcPort] [-d dstPort] infile(s)
```

2단계. 리눅스나 유닉스의 sort 명령어를 통하여 파일을 sorting한다

```
cat ./port_1.txt | sort -k 5, 5 > sort_Result_by_Bytes
ex) port_1.txt 부분중
srcPort  dstPort  Pkts      Pkts/sec  Bytes      Bits/sec
-------  -------  --------  --------  ---------  -----------
20       1570     164412    843.138   246458565  1.01111e+07
2806     2090     11015     56.4872   15478324   635008
```

3 단계. 리눅스나 유닉스의 sort 명령어를 통하여 소스와 데스트네이션 포트별로 sorting한다

```
cat ./sort_Result_by_Bytes | sort -k 2, 2 > dport_1_byte.txt
ex) sport_1_byte.txt 부분중
0  1230  1  0.00333333   52  1.38667
0     7  1  0.00332226   284  7.54817
```

4단계 sport_1_byte.txt와 sport_2_byte.txt 그이상의 txt 파일을 모두 합친후 다시 sorting한다.

```
cat ./sport_1_byte.txt >> sport_total.txt
cat ./sort_total.txt | sort -k 1, 1 > sport_total_byte.txt
```

5 단계 stotal_byte.txt와 dtotal_byte.txt에서 크기순으로 다시 sorting한다

```
cat ./stotal_byte.txt | sort -k 3, 3 > sport.txt
cat ./dtotal_byte.txt | sort -k 3, 3 > dport.txt
```

이후 상기 결과를 엑셀에 가져와서 포트별 특정서비스의 목록을 넣는 부가 처리를 한다. 이와 같이 처리한 후에 C 프로그램을 작성하여 많이 사용되는 트래픽의 포트번호를 구한다.

2. tcpdump를 이용한 분석 방법

tcpdump는 많이 사용되는 트래픽의 포트번호를 구하는 방법과 PDU 내의 원시 데이터를 추출하여 그 용용을 찾는 방법이 있다.

1) 많이 사용되는 트래픽의 포트번호를 구하는 방법

tcpdump의 환경 설정후 현재 많이 사용되는 트래픽의 포트번호를 구해야 하는데, 그 방법의 절차는 다음과 같다.

- 1 단계: Coral reef(crl_flow)를 이용하여 tcpdump 자료를 flow자료로 변환한다.
- 2 단계: 상기 flow자료를 t2_convert를 이용하여 proto_ports_table 을 생성한다.

```
t2_convert Proto_Ports_Table < (flow file) > output
#프로토콜, ok필드, src 포트, dst 포트, 패킷수, 바이트수, 플로우수
      6      1      8080      3421      9      3020      1
```

- 3 단계: 2 단계의 결과를 이용하여 TCP, UDP, ICMP 별로 port Matrix를 생성한다. 이때 bytes, packets, flows 수로 모두 누적 시켜야 추후 응용 서비스 판단 시 유용하게 사용된다.
- 4 단계: 상기에서 얻어진 port matrix에서 src또는 dst 측이 well-known 포트번호이면 반대쪽을 ephemeral 포트로 간주하여 반대쪽 포트 번호를 알려진 포트 번호로 치환한다.
- 5 단계: 이제 상기 자료를 이용하여 src 포트 또는 dst 포트로 summary를 한다. 이후 상기 결과를 엑셀에 가져와서 포트별 특정서비스의 목록을 넣는 부가 처리를 한다. 이와 같이 처리한 후에 perl 프로그램을 작성하여 내림차순으로 많이 활용되는 패킷의 응용포트명을 구한다.

2) 원시 데이터를 바탕으로 분석 파일생성

tcpdump를 이용하며 받은 패킷들은 상기의 과정중 2 단계를 거치면서 많이 사용되는 트래픽의 포트번호를 구하게 된다. 이렇게 구한 많이 사용되는 트래픽의 포트번호를 바탕으로 각 응용별 flow를 분석하기 위해서는 각각의 scr, dst port 별 패킷을 모으는 작업이 필요하다. 분석 파일 생성은 2 단계에서 구한 많이 사용되는 트래픽의 포트번호를 바탕으로 tcpdump 내부 명령을 사용하여 분석하기 위한 포트번호에 해당하는 패킷의 원시 데이터를 추출한다. 예를 들어 src, dst port[PortNum] 같은 특별한 조건을 만족시키는 패킷들을 dump_file 로부터 축출하여, 용의한 분석을 하기 위하여 각각의 조건에 맞는 파일을 생성하는 과정은 다음과 같다. 즉, tcpdump를 이용한 port(dst,src)별 파일생성은 다음과 같은 절차를 따른다. tcpdump -enx -s 1500 -w dump_file.dat 명령을 사용하여 dump_file.dat 라는 파일을 생성했다. 그리고 분석자가 dump_file.dat 로부터 src port 1570 인 패킷만을 골라서 port_1570.dat 파일 생성하고자 원한다. 그러면 다음과 같은 명령을 사용할 수 있다.

```
tcpdump -enx -s 1500 -r dump_file.dat-w port_1570.dat src port 1570
```

이와 같은 명령을 사용하면 port_1570.dat 파일에는 src port 1570에 해당하는 모든 패킷이 입력된 순서대로 위치한다. 같은 방법으로 dst port 1570에 대한 패킷들을 원한다면 다음과 같다.

```
tcpdump -enx -s 1500 -r dump_file.dat -w port_20.dat dst port 1570
```

또한 src, dst 와 상관없이 port 1570 에 대한 모든 패킷을 원한다면 다음과 같은 명령을 사용한다..

```
tcpdump -enx -s 1500 -r dump_file.dat -w port_20.dat port 1570
```

[expression] 옵션은 은 여러 가지 조건을 비교적 자유롭게 조합할 수 있다. 이런 방법으로 생성된 파일들은 초기 dumpfile 보다 분석이 용의하여, 분석자 입장에서 매우 편리하게 이용할 수 있다. 또한 snifferPRO나 Ethereal 같은 프로그램에 로드하여 좀더 비쥬얼한 환경에서 분석 작업을 진행할 수 있다.

IV. 분석결과

인터넷 트래픽을 안동대학교 교내 망에서 상기의 방법을 활용하여 평일 낮시간에 측정한 unwell-known 트래픽에 대한 사용빈도가 많은 서비스를 대상으로 정리하였다. 분석한 결과 주로, 많이 활용되는 트래픽은 서비스 종류별, 네트워크를 활용한 전달 기능관점 살펴보았다. 서비스 종류별로 살펴보면 게임, 채팅, 메신저, 파일공유 등으로 분류된다. 네트워크를 활용한 전달 기능관점 살펴보면, p2p 서비스, p2서버 등으로 분류된다. 이러한 서비스들의 특성을 살펴보면 p2서버의 서비스로부터 p2p 서비스의 특징으로 추이되는 현상을 볼 수 있다.

많은 트래픽은 iana 기관에서 제시된 할당된 포트들을 국내에서는 임의의 서비스를 할당하여 처리하였다. 그 한 예로 iana 기관에서의 포트번호 1093, 1094는 rootd로 활용을 권장하나 국내망에서는 드림위즈 지니 서비스로 활용된다.

V. 결 론

본 논문에서는 TCP와 UDP 위에서 동작하는 (un)well-known 포트를 사용하는 패킷의 PDU 정보에 의한 응용 서비스의 유형을 찾는 트래픽 분석 기법을 수행하였다. 이러한 분석을 위하여 수많은 트래픽 중 활용도가 많은 응용 서비스를 추출하기 위하여, 안동대학교 네트워크에서 ethereal에서 제시된 netflow 및 tcpdump 기법을 활용하였다. 추출된 트래픽의 분석을 위하여 ethereal 트래픽 분석장치를 활용하였다. 추출된 트래픽 서비스의 포트 번호로 서비스명을 알기 위하여 iana에서 제시된 포트 사용 번호를 먼저 점검하였다. iana 기관에서 제시되거나 경험적으로 알고 있는 정의된 서비스의 포트명이면 응용 서비스를 PC에 다운로딩하고 인터넷 환경에서 수행한다. 아울러 트래픽 분석기를 활용하 패킷을 캡쳐하여 그 특성을 분석하였다. 또, 추출된 포트중 어느 서비스인지도 모르면 tcpdump를 수행하며, 발신지 ip를 추적하여 그 서비스 명을 파악하여야 한다. 서비스명을 찾은 다음 그 서비스를 인터넷 환경에 접속된 PC에 다운로딩하고 이의 수행과 더불어 패킷을 캡쳐하여 서비스의 특성을 분석하였다. 향후 현재의 분석 기법으로는 서비스명의 포트번호를 알고 있으면 분석이 용이하며, 서비스 회사나 기관에서 동일한 서비스의 포트명을 바꾸면 그 포트의 서비스를 자동으로 추적하는 기능이 필요하다.

[참고 문헌]

[1] "Sniffer_Pro Protocol Analyzer User manual", http://www.snifferpro.com.

[2] "PA100 Protocol Analyzer User manual", C&C 인스투루먼트, 2000, http://www.cncinst.com

[3] "EtherealProtocol Analyzer User manual", http://www.ethereal.com

[4] "PORT NUMBER", http://www.iana.com

[5] http://www.caida.org

[6] http://www.tcpdump.org