

A Packet Processing of Handling Large-capacity Traffic over 20Gbps Method Using Multi Core and Huge Page Memory Approache

Young-Sun Kwon*, Byeong-Chan Park*, Hoon Chang*

*Student, Dept. of Computer Science and Engineering, Soongsil University, Seoul, Korea

*Student, Dept. of Computer Science and Engineering, Soongsil University, Seoul, Korea

*Professor, Dept. of Computer Science and Engineering, Soongsil University, Seoul, Korea

[Abstract]

In this paper, we propose a packet processing method capable of handling large-capacity traffic over 20Gbps using multi-core and huge page memory approaches. As ICT technology advances, the global average monthly traffic is expected to reach 396 exabytes by 2022. With the increase in network traffic, cyber threats are also increasing, increasing the importance of traffic analysis. Traffic analyzed as an existing high-cost foreign product simply stores statistical data and visually shows it. Network administrators introduce and analyze many traffic analysis systems to analyze traffic in various sections, but they cannot check the aggregated traffic of the entire network. In addition, since most of the existing equipment is of the 10Gbps class, it cannot handle the increasing traffic every year at a fast speed. In this paper, as a method of processing large-capacity traffic over 20Gbps, the process of processing raw packets without copying from single-core and basic SMA memory approaches to high-performance packet reception, packet detection, and statistics using multi-core and NUMA memory approaches suggest When using the proposed method, it was confirmed that more than 50% of the traffic was processed compared to the existing equipment.

▶ **Key words:** Network, Traffic Analysis, Huge Page, NUMA, L4 and L7 Metadata

[요 약]

본 논문에서는 멀티 코어 및 Huge Page 메모리 접근법을 이용한 20Gbps 이상의 대용량 트래픽 처리 가능한 패킷 처리 방법을 제안한다. ICT 기술이 발전함에 따라 전 세계 월 평균 트래픽은 2022년 396억사바이트에 이를 것으로 예측된다. 이러한 네트워크 트래픽의 증가와 동시에 사이버위협 또한 증가하고 있어 트래픽 분석에 대한 중요도가 높아지고 있다. 기존 고비용의 외산 제품으로 분석되고있는 트래픽은 단순히 통계 데이터를 저장함과 동시에 가시적으로 보여주는 것에 불과하다. 네트워크 관리자들은 다양한 구간에서 트래픽을 분석하기 위해 많은 트래픽 분석 시스템을 도입하여 분석하고 있으나, 망 전체의 통합된 트래픽을 확인할 수 없다. 또한, 기존 장비는 10Gbps급이 대부분이기 때문에 매년 증가되고 있는 트래픽을 빠른속도로 처리할 수 없다. 본 논문에서는 20Gbps 이상 대용량 트래픽 처리를 하기 방법으로 단일코어와 기본 SMA 메모리 접근법을 이용한 방법에서 멀티코어와 NUMA 메모리 접근법을 이용하여 고성능으로 패킷수신, 패킷검출, 통계까지 raw 패킷을 copy 없이 처리하는 과정을 제안한다. 제안한 방법을 이용하였때, 기존 장비보다 50%이상 트래픽이 처리되는 것을 확인할 수 있었다.

▶ **주제어:** 네트워크, 트래픽 분석, Huge Page, NUMA, L4 및 L7 메타데이터

- First Author: Young-Sun Kwon, Corresponding Author: Hoon Chang
- Young-Sun Kwon (dolphini0727@naver.com), Dept. of Computer Science and Engineering, Soongsil University
- Byeong-Chan Park (pbc866@gmail.com), Dept. of Computer Science and Engineering, Soongsil University
- Hoon Chang (hoon@ssu.ac.kr), Dept. of Computer Science and Engineering, Soongsil University
- Received: 2021. 06. 10, Revised: 2021. 06. 28, Accepted: 2021. 06. 28.

I. Introduction

ICT 기술이 발전함에 따라 네트워크 트래픽이 급격하게 증가하고 있다. Cisco가 발표한 보고서에 따르면 2017년부터 2022년도까지 개인이 갖는 사물인터넷 기기는 인당 2.4개에서 3.6개로 증가할 것이고, 122.4 엑사바이트였던 전 세계 월 평균 IP 트래픽이 2022년 3배 이상 증가된 396 엑사바이트에 이를 것으로 예측하였다[1]. 네트워크 트래픽이 증가하면서 네트워크를 통한 사이버위협 또한 증가하고 있다[2]. 최근에는 네트워크 트래픽이 암호화되고 있으며 이는 네트워크 관리자가 패킷 내의 페이로드를 식별하거나 분석하기 어렵게 만들고 있다[3].

기존 트래픽 분석 시스템의 경우, 공격 의심 트래픽이 발생하면 사후에 해당 공격의 특성을 분석하여 시그니처를 생성하고, 이를 기반으로 이후 동일한 공격에 대해 탐지하는 수동 대응을 주로 한다. 이런 탐지 결과를 각 로컬 도메인의 입구에 존재하는 방화벽 시스템과 결합하여 각 로컬 도메인으로 유입되는 공격 트래픽을 차단하는 수동 대응 방식이 이루어지고 있다[4]. 이 경우 탐지의 범위 및 성능이 해당 공격의 시그니처에 기반하므로, 신·변종 트래픽을 유발하는 새로운 공격방식을 사용할 경우 적절한 대응이 어렵다. 네트워크상의 트래픽 정보를 수집하고 이를 처리하는 것은 망 관리입장에서 오래전부터 이루어졌다[5]. 하지만 이는 망 엔지니어링이나 망 관리 측면에서 네트워크상에 처리되는 트래픽에 대한 통계 데이터를 제공하고자 하는 것이 주된 목적이며, 트래픽 특성을 분석하여 그 결과로서 공격에 의해 발생하는 이상 트래픽을 탐지하기 위한 목적으로 이용하기에는 부적절하다[6, 7].

본 논문에서는 20Gbps 급 대용량 트래픽을 처리할 수 있는 방법을 제안한다. 이러한 메타데이터를 기반으로 트래픽의 상세 분석보다는 의심스러운 트래픽의 빠른 분석을 통해 우선적으로 이상 트래픽 가능성 정도를 판단할 수 있도록 한다.

본 논문의 구성은 다음과 같다. 2장은 관련 연구로서 다양한 트래픽 탐지 방법을 기술한다. 3장에서는 20Gbps 급의 대용량 트래픽 처리를 위한 방법을 기술하며, 4장에서 실험 및 결과를 보고 5장에서 결론으로 마무리한다.

II. Related Works

1. A Signature-based Traffic Analysis Method

시그니처 기반의 트래픽 분석 방법[8, 9]은 특정 어플리케이션에서 발생시킨 트래픽을 분석하여 다른 어플리케이션과 구별하기 위한 방법으로, 특정 어플리케이션을 분석하여 특징을 추출하고, 특징을 이용하여 어플리케이션을 구분하는 방식이다. 이렇게 구분된 어플리케이션은 높은 정확도를 보이지만, 특징을 찾는 과정에서 사람의 수작업으로 이루어져 새로 생성되는 어플리케이션에 대해 바로 대처할 수 없다는 점과 어플리케이션의 시그니처를 확인할 수 없다면 분류할 수 없다는 단점이 있다. 예를 들어 HTTP 및 SMTP와 같은 일반적인 패킷은 특징을 찾을 수 있으며 추출할 수 있는 HTTP의 Header 정보의 예는 Table 1과 같다.

Table 1. HTTP Response Header

Element	Explanation
Response Line	HTTP/1.1 200 OK
Response Header	Server: nginx Date: Sun, 10 Apr 2016 05:08:46 GMT Content-Type: image/gif Content-Length: 43 Connection: close Pragma: no-cache Expires: Tue, 01 Jan 1980 09:00:00 GMT

그러나 암호화된 패킷인 HTTPS 및 SMTPS 등과 같은 프로토콜은 특징을 찾을 수가 없어 분류할 수가 없다.

2. A Correlation-based Analysis Method Traffic

트래픽 상관관계 분석 방법[10]은 패킷의 주소체계를 이용하여 분석하는 방법으로, 주소체계는 IP주소, 포트 번호, 프로토콜 등과 같으며, Table 2와 같다.

이를 활용하여 트래픽 발생 지점, 형태등을 분석하여 네트워크 망의 연관성을 가중치화하여 임계값을 적용시켜 트래픽을 분류할 수 있는 방법이다. 트래픽 분류에 있어 어플리케이션들이 가지고 있는 특징을 활용하면 높은 분석률을 확인할 수 있다. 그러나 각 어플리케이션에 대한 고유한 특징을 각각 적용할 수가 없으므로 분류에 많은 오류가 있으며, 실제 트래픽에 적용하였을 경우 신뢰성이 부족하다는 단점이 있다.

Table 2. Correlation Method

Element	Explanation
Server-client based	address, port, number, transport layer, protocol(TCP/UTP)
based on time of occurrence	address, start time
host-to-host based	source address, destination address
based on statistics	address, number of packet, size of byte

3. A Machine Learning-based Traffic Analysis Method

머신러닝 기반 트래픽 분석 방법[11]은 어플리케이션 별로 특징이 될 수 있는 여러 항목을 각종 머신러닝에 학습시켜 트래픽을 분류할 수 있는 방법이 있으며, Table 3과 같다.

Table 3. Typical Machine Learning-based Traffic Classification

Approach	Features Used	Purpose Protocol
Naïve Bayes	Duration, TCP Port, Interval, Size etc.	FTP, SSH, SMTP, WWW, DNS etc.
K-means	Packet length mean, duration etc.	Web, P2P, FTP etc.
Decision Tree	Protocol, Duration, Byte volume etc.	FTP, Telnet, SMTP, DNS, HTTP

특징이 되는 항목은 Port Number, Flow Duration, Inter-arrival Time, Packet Size 등을 Classification 및 Clustering의 머신러닝을 이용하여 학습시킨다. 이러한 방법은 분류에 있어 높은 분석률을 제공할 수 있다. 그러나, 학습이 되지 않는 특정 데이터나 사람이 분류했을 때 같은 어플리케이션으로 판단되나 기존에 학습된 항목과 다른 데이터를 가진 패킷은 같은 어플리케이션으로 분류할 수 없어 제한적이기 때문에 모든 인터넷에 적용하였을 경우 정확성이 떨어진다.

III. A Packet Processing Method for Capable of Handling Large-capacity Traffic over 20Gbps

1. Architecture

본 논문에서 제안하는 20Gbps 이상 트래픽 처리 가능한 패킷 처리 방법은 패킷을 검출하는 H/W를 중심으로 검출한 패킷을 빠르게 분석할 수 있는 부분인 Kernel, Core Processing, 어플리케이션으로 나뉘며, Fig. 1과 같다.

이러한 구조에서 다음과 같은 과정을 통해 대용량의 패킷을 빠른 속도로 분석할 수 있어야 한다.

- Packet Zero Copy 구현하여 대용량 트래픽 처리
- 별도의 모듈을 통해 분석 및 통계를 종합적으로 처리

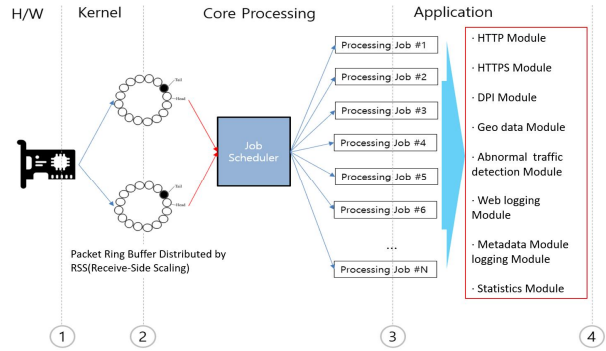


Fig. 1. H/W Structure for Handling Large-Capacity Traffic

2. An Analysis Method for Handling Large-capacity Traffic

기가급 이상 트래픽 처리를 위한 과정으로 패킷수신, 패킷 검출, 통계까지 raw 패킷을 copy 없이 처리하는 과정이며, Fig. 1에서 ① ~ ③의 부분이며, 두가지 과정으로 나뉜다.

첫 번째, ① ~ ②에 대한 과정이며 듀얼 랜 포트를 사용하여, 패킷검출 속도를 10G 이상으로 검출할 수 있도록 Intel 시스템에서 호환가능한 NIC(Network Interface Card) Driver를 커스터마이징 한다. 커스터마이징 과정은 RSS(수신측 배율) 값을 활용한 Buffering 튜닝과정이다. RSS 값에 따라 검출되는 패킷 수를 확인할 수 있으며, Fig. 2 및 Fig. 3과 같다.

```

eth7-0    eth7-1    eth8-0    eth8-1
1.60K    3.93K    4.09K    4.09K

tcp pps  udp pps  icmp pps  priv tcp  priv udp  priv icmp
529.22K  0.00    0.00

[root@nsm net]# ls -al /proc/net/ |grep etrc
-rw-rw-rw- 1 root root 0 Feb 12 13:28 etrc_rx_001b21bcbf9e
-rw-rw-rw- 1 root root 0 Feb 12 13:28 etrc_rx_001b21bcbf9e.1
-rw-rw-rw- 1 root root 0 Feb 12 13:28 etrc_rx_001b21bcbf9f
-rw-rw-rw- 1 root root 0 Feb 12 13:28 etrc_rx_001b21bcbf9f.1
-rw-rw-rw- 1 root root 0 Feb 12 15:20 etrc_tx_001b21bcbf9e
-rw-rw-rw- 1 root root 0 Feb 12 15:20 etrc_tx_001b21bcbf9e.1
-rw-rw-rw- 1 root root 0 Feb 12 15:20 etrc_tx_001b21bcbf9f
-rw-rw-rw- 1 root root 0 Feb 12 15:20 etrc_tx_001b21bcbf9f.1
    
```

Fig. 2. RSS = 2 < NIC Port 2

```

eth7-0   eth7-1   eth7-2   eth7-3   eth8-0   eth8-1
 3.56K   19.85K   20.76K   19.45K   4.09K   4.09K
eth8-2   eth8-3
 4.09K   4.09K

tcp pps  udp pps  icmp pps  prv tcp  prv udp  prv icmp
601.60K  0.00   0.00

[root@nsm net]# ls -al /proc/net/ |grep etrc
rW-rw-rw- 1 root root 0 Feb 12 15:25 etrc_rx_001b21bcbf9e
rW-rw-rw- 1 root root 0 Feb 12 15:25 etrc_rx_001b21bcbf9e.1
rW-rw-rw- 1 root root 0 Feb 12 15:25 etrc_rx_001b21bcbf9e.2
rW-rw-rw- 1 root root 0 Feb 12 15:25 etrc_rx_001b21bcbf9e.3
rW-rw-rw- 1 root root 0 Feb 12 15:25 etrc_rx_001b21bcbf9f
rW-rw-rw- 1 root root 0 Feb 12 15:25 etrc_rx_001b21bcbf9f.1
rW-rw-rw- 1 root root 0 Feb 12 15:25 etrc_rx_001b21bcbf9f.2
rW-rw-rw- 1 root root 0 Feb 12 15:25 etrc_rx_001b21bcbf9f.3
    
```

Fig. 3. RSS = 4 < NIC Port 2

RSS 값을 4로 설정했을 때 듀얼 포트에서 패킷이 약 3배 이상 검출되는 것을 확인할 수 있었다.

두 번째, ② ~ ③의 부분이며, Hash Map을 이용한 Processing Tread #1 ~ N개를 이용하여 시스템 자원을 최대한 활용할 수 있도록 튜닝하는 과정이다. 메모리 처리 과정에서 NUMA[12]를 적용하여 대용량 트래픽 발생시 메모리를 병렬처리 최적화를 진행하며 Fig. 4와 같다.

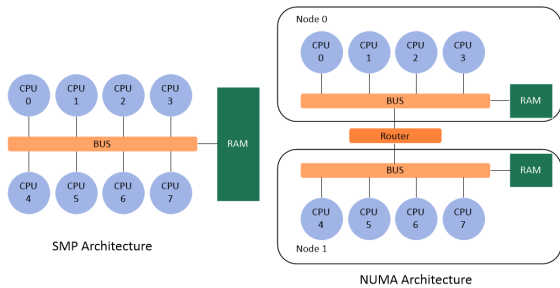


Fig. 4. NUMA Architecture

기본 SMP 구조에서 NUMA 구조를 활용하여 BUS의 병목 현상 및 지연 현상을 줄일 수 있으며, node 0은 짝수 Core, node 1은 홀수 Core로 나뉜 것으로 확인할 수 있으며, Fig. 5와 같다.

```

[root@np-234 ~]# numactl -s
policy: default
preferred node: current
physcpubind: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
cpubind: 0 1
nodebind: 0 1
membind: 0 1
[root@np-234 ~]# numactl -H
available: 2 nodes (0-1)
node 0 cpus: 0 2 4 6 8 10 12 14 16 18 20 22
node 0 size: 16338 MB
node 0 free: 14905 MB
node 1 cpus: 1 3 5 7 9 11 13 15 17 19 21 23
node 1 size: 16384 MB
node 1 free: 15208 MB
node distances:
node 0 1
0: 10 20
1: 20 10
    
```

Fig. 5. NUMA Environment Verified with Linux Command

또한, Huge Page 기법[13]을 활용하여 메모리 블록 사이즈 접근성을 높이며, Fig. 6과 같다.

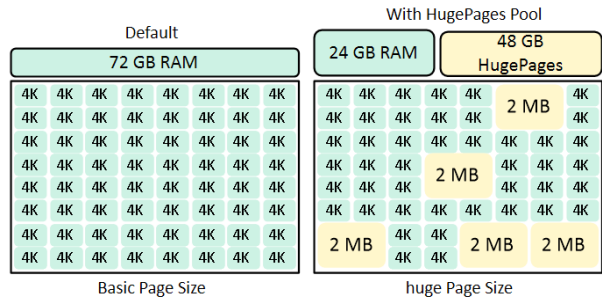


Fig. 6. Apply Huge Page Pool

메모리 블록 사이즈를 4Kbytes(OS기본)에서 2Mbytes로 변경하여, 메모리 사용시 Page Fault가 날 수 있는 확률을 줄여서 접근 속도를 향상시킨다. 접근 실패 확률을 줄일 수 있다. 이를 활용했을 때 처리되는 메모리 크기를 확인할 수 있으며, Fig 7과 같다.

```

[root@np-234 ~]# grep Huge /proc/meminfo
AnonHugePages: 47104 kB
HugePages_Total: 256
HugePages_Free: 256
HugePages_Rsvd: 0
HugePages_Surp: 0
Hugepagesize: 2048 kB

[root@np-234 ~]# grep Huge /proc/meminfo
AnonHugePages: 575488 kB
HugePages_Total: 256
HugePages_Free: 12
HugePages_Rsvd: 12
HugePages_Surp: 0
Hugepagesize: 2048 kB
    
```

Fig. 7. Before and After Applying Huge Page Pool

Huge Page를 적용했을 때의 처리되는 메모리 크기는 약 8% 향상된 것을 확인할 수 있다.

3. L4 and L7 Metadata

검출된 패킷은 대용량으로 TCP/UDP 세션을 처리할 수 있도록 나누어지고 Thread 별로 처리되며, Fig. 1의 ③ ~ ④의 과정이다. 각 모듈인 HTTP, GEO Data, 파일 추출 모듈 등 별도의 Thread 처리해야 하는데, 모듈별로 처리하기 위해서는 L4와 L7의 정보를 바탕으로 처리된다.

20Gbps 급 대용량 트래픽 처리시 L4의 메타데이터를 초당 20K / Logs 추출 및 저장 및 L7의 메타데이터를 초당 15K / Logs 추출 및 저장 기능을 제공하여야 한다. 저장되는 L4 및 L7의 메타데이터의 포맷은 Table. 4와 같다.

L4 and L7 메타데이터 추출 및 저장을 위해 포맷, 용량, 파일 개수, 로그 포맷, 파일 개수 등을 설정할 수 있어야 한다.

Table 4. L4 and L7 Metadata Format

Element	Explanation
Type	L4 or L7
Enable_Log	Whether to save log files
Filename	Log file name
Path	Storage path
Rolling_size	Log file size
Rolling_sec	Log file recording time
Rolling_count	Number of log files
Log_buffer_size	Log file maximum size
format_filemane	Format file name for L4 or L7

IV. Experiments and Results

본 논문에서 제안하는 20Gbps 급 처리가 가능한 패킷 처리 방법을 구현하기 위해 Table. 5와 같은 장비를 구비하여 실험하였다.

Table 5. Packet Analyzer Specifications

Element	Explanation
CPU	Inter(R) Xeon(R) Sliver 4114 CPU @ 2.20GHz
Memory	64GB
OS	Redhat CentOS 7.5.1804
Kernal Version	3.10.0-862.11.6.el7.x86_64
I/O Interface	1000Base-T x 4 Ports 10G Fiber X 4 Ports

기가급 이상 트래픽 처리를 위해 패킷수신, 패킷검출, 통계까지 raw 패킷을 복사없이 일괄적으로 처리할 수 있도록 구현하였으며, 구조는 Fig. 8과 같다.

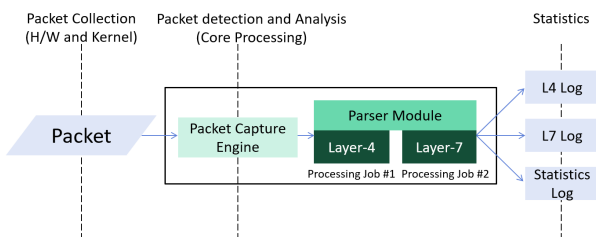


Fig. 8. High-performance Traffic Processing Architecture

실험은 패킷 검출기가 패킷을 검출하는 과정에서 패킷 분석기가 패킷 검출 및 분석하여 L4 모듈을 통한 L4 데이터 분석 및 L7 모듈을 통한 L7 분석 과정을 통해 최종적으로 20Gbps 급의 패킷 처리를 할 수 있는지를 확인하였다. 실험은 크게 세가지로 나뉘며, 첫 번째, 패킷 분석기가 실제 트래픽을 20Gbps 급으로 트래픽을 처리하는지 확인하였으며, 두 번째, 패킷 분석기가 L4 메타데이터를 초당

20K 이상으로 Log를 저장하는지 확인하였고 마지막으로 패킷 분석기가 L7 메타데이터를 초당 15K 이상으로 Log를 저장하는지 확인하였다.

실험을 위해 Table. 6과 같은 시험환경으로 구성하여 시험하였다.

Table 6. Experiment Environment

Element	Explanation
Traffic Pattern	10 GE x 2 Ports 1-to-1 (1 Pair)
Direction	Bi-directional
Clinet / Server Numver	200/50
Protocol	HTTP, HTTPS, FTP, SMTP
Input Load (Sessions/sec)	HTTP: 20K/s HTTPS: 1K/s FTP: 3K/s SMPT: 1K/s
Aggregated Data Throughput	20Gbps
Duration	Ramp Up: Full Open+Data+Full Close (00:00:01) Steady-state: Open and Close Sessions (00:10:00) Ramp Down: Full Close (00:00:35)
Contents Size	100Kbytes

첫 번째, 패킷 분석기가 20Gbps 이상의 트래픽을 처리하는지 확인하였다. 시험절차는 Table. 3에서 설정한 값을 패킷 검출기로부터 패킷 분석기로 인가하고 처리결과를 확인하였으며, Fig. 9 및 Table. 7과 같다.

```

sessions +sessions -sessions tuples mem
254.65K 24.54K 24.18K 509.29K 1.97G

tuple total tcp udp icmp etc error
in: 254488 254488 0 0 0 0
out: 254486 254486 0 0 0 0

rx bps pps bytes packets dropped
eth8 0.00 0.00 0.00 0.00 0.00
eth3 10.00G 992.46K 27.08G 21.61M 0.00
eth4 10.00G 992.32K 27.93G 22.13M 0.00

lgn total sys_drop no_mem ln_shrt net_shrt net_unk
in: 0.00 0.00 0.00 0.00 0.00 0.00
out: 0.00 0.00 0.00 0.00 0.00 0.00

eth3-0 eth3-1 eth3-2 eth4-0 eth4-1
255.00 4.38K 4.16K 4.33K 4.21K

tcp pps udp pps icmp pps prv tcp prv udp prv icmp
48.88K 0.00 0.00
    
```

Fig. 9. Check Handling Traffic over 20 Gbps

Table 7. Traffic Processing Test Result

	Detector Transmission	Total	analyzer Processing	Traffic	Drop Count
Count	1,192,247,370	1,192,247,370	1,192,247,370		0
Byte	1,472,982,833,928	1,472,982,833,928	1,472,982,833,928		-

패킷 검출기 및 분석기의 패킷 수, 바이트 수, 패킷 드랍 카운트를 비교하였을 때, 누락 없이 모두 정상처리 된 것을 확인할 수 있었다.

두 번째, 분석기가 대용량 트래픽을 처리 시 L4 메타데이터를 초당 20K 이상으로 Log를 저장하는지 실험하였으며, 결과는 Fig. 10 및 Table 8과 같다.

```

[rt_flow.log_20200203_144940] 2020/02/03-14:49:51 24483
[rt_flow.log_20200203_144940] 2020/02/03-14:49:52 24562
[rt_flow.log_20200203_144940] 2020/02/03-14:49:53 24511
[rt_flow.log_20200203_144940] 2020/02/03-14:49:54 24573
[rt_flow.log_20200203_144940] 2020/02/03-14:49:55 24463
[rt_flow.log_20200203_144940] 2020/02/03-14:49:56 24530
[rt_flow.log_20200203_144940] 2020/02/03-14:49:57 24577
[rt_flow.log_20200203_144940] 2020/02/03-14:49:58 24483
[rt_flow.log_20200203_144940] 2020/02/03-14:49:59 24573
[rt_flow.log_20200203_144940] 2020/02/03-14:50:00 24586
[rt_flow.log_20200203_144940] 2020/02/03-14:50:01 24482
[rt_flow.log_20200203_144940] 2020/02/03-14:50:02 24532
[rt_flow.log_20200203_144940] 2020/02/03-14:50:03 25041
[rt_flow.log_20200203_144940] 2020/02/03-14:50:04 24041
[rt_flow.log_20200203_144940] 2020/02/03-14:50:05 24502
[rt_flow.log_20200203_145005] 2020/02/03-14:50:06 24560
[rt_flow.log_20200203_145005] 2020/02/03-14:50:07 24489
[rt_flow.log_20200203_145005] 2020/02/03-14:50:08 24541
[rt_flow.log_20200203_145005] 2020/02/03-14:50:09 24517
[rt_flow.log_20200203_145005] 2020/02/03-14:50:10 24577
[rt_flow.log_20200203_145005] 2020/02/03-14:50:11 24527
[rt_flow.log_20200203_145005] 2020/02/03-14:50:12 24480
[rt_flow.log_20200203_145005] 2020/02/03-14:50:13 24514
[rt_flow.log_20200203_145005] 2020/02/03-14:50:14 24567
[rt_flow.log_20200203_145005] 2020/02/03-14:50:15 24491
[rt_flow.log_20200203_145005] 2020/02/03-14:50:16 24590
[rt_flow.log_20200203_145005] 2020/02/03-14:50:17 24538
[rt_flow.log_20200203_145005] 2020/02/03-14:50:18 24462
[rt_flow.log_20200203_145005] 2020/02/03-14:50:19 24549
[rt_flow.log_20200203_145005] 2020/02/03-14:50:20 24453
[rt_flow.log_20200203_145005] 2020/02/03-14:50:21 24589
[rt_flow.log_20200203_145005] 2020/02/03-14:50:22 24469
[rt_flow.log_20200203_145005] 2020/02/03-14:50:23 24598
[rt_flow.log_20200203_145005] 2020/02/03-14:50:24 24489
[rt_flow.log_20200203_145005] 2020/02/03-14:50:25 25137
[rt_flow.log_20200203_145005] 2020/02/03-14:50:26 23951
[rt_flow.log_20200203_145005] 2020/02/03-14:50:27 24521
[rt_flow.log_20200203_145005] 2020/02/03-14:50:28 24559
[rt_flow.log_20200203_145005] 2020/02/03-14:50:29 24499
[rt_flow.log_20200203_145005] 2020/02/03-14:50:30 24539
[rt_flow.log_20200203_145040] 2020/02/03-14:50:31 24542
[rt_flow.log_20200203_145040] 2020/02/03-14:50:32 24481
[rt_flow.log_20200203_145040] 2020/02/03-14:50:33 24536
    
```

Fig. 10. L4 log Saves per Second

Table 8. L4 Metadata Processing Test Results

	Detector Sessions	Cumulative	Total number of L4 logs stored
Count	14,735,991		14,735,991
Average per Second	24,540		24,524

패킷 검출기에서는 초당 평균 24.5K의 세션을 발생시켰으며, 분석기는 초당 24.5K의 세션을 처리하였다. 초 단위로 저장된 메타데이터 개수를 확인해보면 평균 24K Logs/sec 이상의 로그를 저장한 것을 확인할 수 있었다.

마지막으로 분석기가 초당 15K 이상 L7를 저장하는지 실험하였으며, 결과는 Fig. 11 및 Table. 9와 같다.

패킷 검출기에서는 초당 평균 19.5K의 트랜잭션을 발생시켰으며, 분석기는 초당 19.5K의 트랜잭션을 처리하였다. 초 단위로 저장된 메타데이터 개수를 확인해보면 평균 19.5K Logs/sec 이상의 로그를 저장한 것을 확인할 수 있었다.

또한, 기존 10Gbps 급 트래픽 처리와 비교 분석하여 성능을 확인하였으며, 결과는 Table. 10과 같다.

```

[rt_web.log_20200203_145139] 2020/02/03-14:51:11 19521
[rt_web.log_20200203_145139] 2020/02/03-14:51:12 19570
[rt_web.log_20200203_145139] 2020/02/03-14:51:13 19524
[rt_web.log_20200203_145139] 2020/02/03-14:51:14 19544
[rt_web.log_20200203_145139] 2020/02/03-14:51:15 19529
[rt_web.log_20200203_145139] 2020/02/03-14:51:16 19517
[rt_web.log_20200203_145139] 2020/02/03-14:51:17 19535
[rt_web.log_20200203_145139] 2020/02/03-14:51:18 19535
[rt_web.log_20200203_145139] 2020/02/03-14:51:19 19573
[rt_web.log_20200203_145139] 2020/02/03-14:51:20 19512
[rt_web.log_20200203_145139] 2020/02/03-14:51:21 19556
[rt_web.log_20200203_145139] 2020/02/03-14:51:22 19539
[rt_web.log_20200203_145139] 2020/02/03-14:51:23 19538
[rt_web.log_20200203_145139] 2020/02/03-14:51:24 19484
[rt_web.log_20200203_145139] 2020/02/03-14:51:25 19574
[rt_web.log_20200203_145139] 2020/02/03-14:51:26 19489
[rt_web.log_20200203_145139] 2020/02/03-14:51:27 19563
[rt_web.log_20200203_145139] 2020/02/03-14:51:28 19718
[rt_web.log_20200203_145139] 2020/02/03-14:51:29 19271
[rt_web.log_20200203_145139] 2020/02/03-14:51:30 19579
[rt_web.log_20200203_145139] 2020/02/03-14:51:31 19551
[rt_web.log_20200203_145139] 2020/02/03-14:51:32 19477
[rt_web.log_20200203_145139] 2020/02/03-14:51:33 19558
[rt_web.log_20200203_145139] 2020/02/03-14:51:34 19535
[rt_web.log_20200203_145139] 2020/02/03-14:51:35 19491
[rt_web.log_20200203_145139] 2020/02/03-14:51:36 19564
[rt_web.log_20200203_145139] 2020/02/03-14:51:37 19553
[rt_web.log_20200203_145139] 2020/02/03-14:51:38 20060
[rt_web.log_20200203_145139] 2020/02/03-14:51:39 18979
[rt_web.log_20200203_145140] 2020/02/03-14:51:40 19502
[rt_web.log_20200203_145140] 2020/02/03-14:51:41 19659
[rt_web.log_20200203_145140] 2020/02/03-14:51:42 19492
[rt_web.log_20200203_145140] 2020/02/03-14:51:43 19454
[rt_web.log_20200203_145140] 2020/02/03-14:51:44 19584
[rt_web.log_20200203_145140] 2020/02/03-14:51:45 19537
[rt_web.log_20200203_145140] 2020/02/03-14:51:46 19471
[rt_web.log_20200203_145140] 2020/02/03-14:51:47 19547
[rt_web.log_20200203_145140] 2020/02/03-14:51:48 19547
[rt_web.log_20200203_145140] 2020/02/03-14:51:49 19494
[rt_web.log_20200203_145140] 2020/02/03-14:51:50 19548
[rt_web.log_20200203_145140] 2020/02/03-14:51:51 19547
[rt_web.log_20200203_145140] 2020/02/03-14:51:52 19508
[rt_web.log_20200203_145140] 2020/02/03-14:51:53 19575
[rt_web.log_20200203_145140] 2020/02/03-14:51:54 19436
[rt_web.log_20200203_145140] 2020/02/03-14:51:55 19576
    
```

Fig. 11. L7 Log Saves per Second

Table 9. L7 Metadata Processing Test Result

	Detector Sessions	Cumulative	Total number of L7 logs stored
Count	11,730,278		11,730,278
Average per Second	19,256		19,255

Table 10. Comparative Analysis

	10Gbps	20Gbps
Analyzer Traffic Processing Count	734,521,687,462	1,472,982,833,928
Number of L4 Log Saves	7,215,113	14,735,991
Number of L7 Log Saves	5,664,210	11,730,278

실험결과를 종합적으로 확인했을 때, 기존 10Gbps급보다 본 논문에서 제안한 20Gbps급의 대용량 트래픽 처리 방법이 약 50% 향상된 트래픽 처리 및 로그 저장 기능을 확인할 수 있다.

V. Conclusions

본 논문에서는 기존 고비용의 외산 제품으로 분석되고 있는 네트워크상의 트래픽에 대해 단순히 통계 데이터를 저장 및 가시적으로 보여주는 것과 네트워크 관리자들이 다양한 구간에서 트래픽을 분석하기 위해 많은 트래픽 분석 시스템을 도입하여 분석하고 있으나, 망 전체의 통합된 트래픽을 확인하기가 어려운 점을 대체하기 위한 방법으로 20Gbps 이상의 대용량 트래픽 처리 가능한 패킷 처리 방법을 제안하였다. 제안한 방법은 L4 및 L7에 대한 메타

데이터를 초당 20K/Logs와 15K/Logs의 저장 성능을 확보하여, 망 전체뿐만 아니라 세부적으로 트래픽 관리가 가능함을 확인할 수 있었다.

본 논문에서 제안한 방법을 사용하였을 경우, 추출된 메타데이터를 빅데이터화 하여 공격 의심 트래픽으로 분석하고 이상 트래픽을 탐지할 수 있는 raw 데이터로 활용할 수 있으며, 프로토콜, 어플리케이션 정보를 빅데이터화 하여 각종 통계자료로도 활용할 수 있다.

추후 연구로써, 각 프로토콜 및 어플리케이션에 대하여 본 논문에서 제안한 방법으로 메타데이터를 추출 및 빅데이터화 해야하며, 빅데이터 기반으로 이상 트래픽 탐지할 수 있는 연구가 필요하다.

ACKNOWLEDGEMENT

This research project supported by Ministry of SMEs and Startups(MSS) and Korea Technology & Information Promotion Agency for SMEs(TIPA) in 2021(S2909329)

REFERENCES

- [1] Cisco, "Cisco Visual Networking Index: Forecast and Trends, 2017-2022," 2018. 11
- [2] D. I. Oh, "In the post-corona era, cyber attacks will intensify," Electronic Newspaper, 2020. 05, <https://www.etnews.com/20200512000181>
- [3] K. H. Jung, B. H. Lee and D. Yang, "Performance Analysis of Detection Algorithms for the Specific Pattern in Packet Payloads" Journal of the Korea Institute of Information and Communication Engineering, Vol. 22, No. 5, pp. 794-840, 2018. 02. DOI: <https://doi.org/10.6109/jkiice.2018.22.4.794>
- [4] S. H. Lee, J. C. Na and S. W. Son, "Traffic Analysis Technology Trends in Terms of Security," <https://www.itfind.or.kr/WZIN/jugidong/1117/111701.htm>
- [5] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets. In Advances in Neural Information Processing Systems," NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems, Vol. 2, pp. 2672-2680, 2014. 03.
- [6] Schlegl Thomas, Philipp Seeböck, Sebastian M. Waldstein, Ursula Schmidt-Erfurth and Georg Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," International Conference on Information Processing in Medical Imaging. Springer, Cham, 2017. 03.
- [7] H. J. Kim, H. S. Kim, D. M. Shin, "Design and Implementation of Tor Traffic Collection System Using Multiple Virtual Machines," Journal of Software Assessment and Valuation, Vol. 15, No. 1, pp. 1-9, 2019. 06. DOI: <https://10.29056/jsav.2019.06.01>
- [8] S. Kim and S. Lee, "Automatic Malware Detection Rule Generation and Verification System", Journal of Internet Computing and Services(JICS), Vol. 20, No. 2, pp. 9-19, 2019. 09. DOI: <https://doi.org/10.7472/jksii.2019.20.2.9>
- [9] M. Thottan and C. Ji, "Anomaly Detection in IP Networks," IEEE Transactions on Signal Processing, vol. 51, no. 8, pp.2191-2204. 2003. 05. DOI: <https://doi.org/10.3745/KTCCS.2020.9.5.113>
- [10] J. K. Lee, S. J. Kim and T. Hong, "Analysis of Traffic and Attack Frequency in the NURION Supercomputing Service Network," KIPS Transactions on Computer and Communication Systems, Vol. 9, No. 5, pp.113-120, 2020. 01. DOI: <https://doi.org/10.3745/KTCCS.2020.9.5.113>
- [11] H. H. Lim, D. H. Kim, K. T. Kim and H. Y. Youn, "Traffic classification using machine learning in SDN," Winter Conference of the Korean Society of Computer and Information Technology, Vol. 26, No. 1, 2018. 01.
- [12] Vijayan, Jaikumar, McCann and Stefanie, "NUMA," Computerworld, Vol. 32, No. 23, 1998. 06.
- [13] S. Ahn, D. Kang and Y. Eom, "Analysis on the Characteristics and Performance Effects of Linux Huge Page," Journal of the Korean Society of Information Sciences, Vol. 2017, No. 06, pp. 73-75, 2017. 06.

Authors



Young-Sun Kwon received the B.S. degrees in Multimedia from Soongsil University, Korea, in 2020. She is currently a M.S. student in the Department of Computer Science and Engineering, Soongsil University.

She is interested in Internet traffic analysis, network management and Internet security.



Byeong-Chan Park received the B.S., M.S., degree in Computer Science and Engineering from Soongsil University, Korea, in 2015 and 2018, respectively. He is currently a PH.D student in the Department of Computer

Science and Engineering, Soongsil University. He is interested include DRM(Digital Right Management), Deep Learning, and Blockchain.



Hoon Chang received the B.S., M.S., degree in Electrical Engineering from Seoul National University, Korea, in 1987 and 1989, respectively. He is received PH.D degree in ECE from University of Texas at Austin, in

1993. He is currently a professor in the Department of Computer Science and Engineering, Soongsil Universtiy. He is interested in Computer System(Embedded System), VLSI/SoC, Design Automation