

Network Abnormal Behaviour Analysis System

Sunoh Choi, Yangseo Choi, Jooyoung Lee, Jonghyun Kim and Ikkyun Kim

Network Security Research Group

Electronics and Telecommunication Research Institute (ETRI)

218 Gajeong-ro Yuseong-gu Daejeon Korea

{suno, yschoi92, joolee, jhk, ikkyun21}@etri.re.kr

Abstract— As cyber attacks have increased in recent years, network forensics, which collects and analyses network packets as well as digital forensics, has been studied. However, high-speed networks such as 1 or 10 Gbps networks have many network flows. For example, a 1 Gbps network has hundreds of millions of network flows per day. Analysing network traffic in this situation is very difficult and time-consuming. In this paper, we propose a system that can analyse network abnormal behaviour quickly and easily. We first propose a system that stores the TCP flag when generating network flows. Second, we present some ways to use the TCP flag in network flows to analyse network anomalies such as persistent outbound connections.

Keywords— Network Flow, Analysis

I. INTRODUCTION

Many cyber attacks have occurred recently. Many security devices such as intrusion detection / prevention systems (IDS / IPS), firewalls, antivirus, security information and event management (SIEM) are used to prevent these attacks. Nowadays, much research is being done in network forensics. Network forensics is to analyse cyber attacks by capturing network packets. After analysing how cyber attacks occur in network packets, you can apply the network system to prevent the same cyber attacks.

However, network forensics has some difficulties. First network traffic is volatile. When a network packet is sent, it is discarded. In addition, a high-speed network such as 1 Gbps will generate many packets and flows. Network forensics require network packets to be captured and recorded. However, on a 1 Gbps network, about 10 TB of data is recorded per day, with about 200 million flows recorded and about 1 billion packets recorded. For network forensics, you need to quickly search for network flows, network packets, or patterns. In this paper, we focus on network flow analysis for network forensics.

In this paper, we propose a network abnormal behavior analysis system. First, we provide a way to create a network flow with tcp flag information. The tcp flag information later plays an important role in network abnormal behavior analysis. Second, it provides a way to quickly detect network flows using bitmap indexes. Because you have a lot of network flow every day, you need a way to quickly find your network flow.

Third, we propose a method to analyse network abnormal operation such as persistent outbound connection using network flow and bitmap index. By analysing persistent outbound connections, we can find infected hosts and command and control (C & C) servers. Finally, you can see that these modules perform reasonably well by implementing these three modules and running the experiment.

II. NETWORK ABNORMAL BEHAVIOUR ANALYSIS SYSTEM

The purpose of this paper is to propose a system for analysing network abnormal behaviour. It has two parts. The first is to introduce the function of the module in the network abnormal behaviour analysis system. The second is to provide a way to analyse network abnormal behaviour.

In Figure 1, the system consists of three modules. There are network flow generation module, network flow search module, and network abnormal behaviour analysis module.

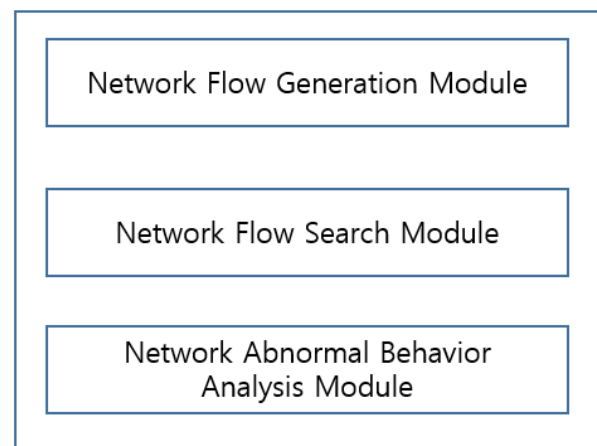


Figure 1. Network Abnormal Behavior Analysis System

A. Network Flow Generation Module

The inputs of the system are pcap files. If the pcap files are given to the system, firstly, the flow generation module generates network flow information from the pcap files. The format of network flow information is as follows.

(sourcehost, sourceport, destinationhost, destinationport, packets, bytes, protocol, service, filetype, tcpflags,

starttime, endtime)

Network flow information makes us to know how network packets have transmitted between two network devices. Sourcehost means an IP address of the network device sending network packets. Sourceport means a port number of the network device sending network packets. Destinationhost means an IP address of the network device receiving network packets and destinationport means a port number of the network device receiving network packets.

Packets means the number of the packets which the flow sends and bytes means the number of bytes which the flow sends. Protocol means the protocol name which is used in the flow and service means the service name which is used in the flow. (e.g., HTTP, FTP) Filetype means the file type transmitted in the flow if a file is transmitted. (e.g., PE, PDF) Starttime means when the first packet in the flow has transmitted and endtime means when the last packet in the flow has transmitted.

TCPflags means the accumulated value of tcp flags of packets transmitted in the flow. TCP has 6 flags. They are URG, ACK, PSH, RST, SYN and FIN. When the flow has a SYN packet and a ACK packet, SYN flag and ACK flag are set to 1. Then, the tcpflags value of the flow is 010010.

B. Network Flow Search Module

In section II-A, network flows are generated in the network flow generation module. Then, these network flows are stored in the network flow search module. In order to support fast search, we use bitmap index [1,2]. Bitmap index is used for the search of mass data like network flows in the high speed network. Traditionally, B-tree is widely used as an index in databases like MySQL. B-tree supports insertion, deletion and update of data. On the other hand, bitmap index is faster than B-tree in making indexes. But, it supports only insertion of data and it can't delete and update data. However, since bitmap index is faster than B-tree when making index of mass data like network flows, network flow search module uses bitmap index.

C. Network Abnormal Behavior Analysis Module

The third module of the system is network abnormal behavior analysis module. The module gives a pre-defined query to the network flow search module. Then, the network flow search module returns the query result to the network abnormal behaviour analysis module.

In this section, we propose a method to analyse network abnormal behaviour. First, we give a method to analyse persistent outbound connection. Persistent outbound connection is for an infected host to send SYN packets to a command and control (C&C) server as Figure 2. Then, if the port of C&C server is not open yet, the server sends a RST packet to the infected host. By analysing persistent outbound connection, we may find infected hosts and C&C servers.

In order to analyse persistent outbound connection, a query is written as follows.

```
SELECT sourcehost, sourceport,
```

```
destinationhost, COUNT(*)
```

```
FROM flowtable
```

```
WHERE tcpflags = 010100
```

```
GROUP BY sourcehost, sourceport, destinationhost
```

```
HAVING COUNT(*) >= threshold;
```

```
ORDER BY COUNT(*) DESC;
```

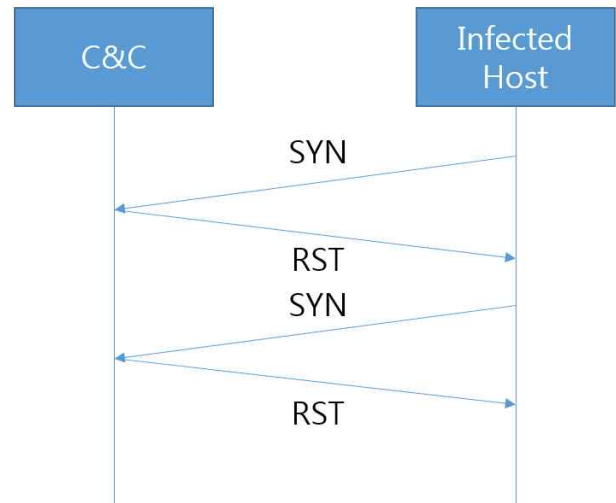


Figure 2. Persistent Outbound Connection

First, in order to analyse persistent outbound connection, we have to find flows which send a RST packet. In a RST packet, ACK flag and RST flag are set to 1. So, the value of tcp flags in the flow is 010100. Next, we should group the result set by three columns. The three columns are sourcehost, sourceport and destinationhost. Then, finally by using the aggregation function like COUNT(*), we find groups which have received many RST packets.

In the query, the destination host sends a SYN packet to the source host and the source host sends a RST packet to the destination host. For example, when the result is (A, B, C, D), the infected host C sends a SYN packet D times to the port B of the C&C server A and the C&C server A sends RST packets D times to the infected host C.

III. EXPERIMENTAL RESULTS

We implemented the network abnormal behaviour analysis system. We developed the network flow generation module using gcc version 4.8.5 on a 2.6GHz Intel Xeon CPU E5-2690 with 128GB RAM running CentOS 7. The machine is also used for gathering network packets in a high speed network. Also we developed the network flow search module by using InfiniFlux database [3] and the network abnormal behaviour analysis module using Java JDK 1.7 on a 2.6GHz Intel Xeon CPU E5-2660 with 64GB RAM running CentOS 7.

We use two data sets for the experiments. The data set A is a set of network flows for an hour on a 1Gbps enterprise network and the data set B is a set of network flows for a day on a 1Gbps enterprise network in Table 1. The number of

flows in the data set A is about 11 million and the number of flows in the data set B is about 200 million.

We executed experiments to analyse persistent outbound connection for these two data sets. The filtering time is the time to search flows whose tcpflag value is 010100. Thus, it is the time to search flows containing a RST packet. The grouping time is the time to group the flows by three columns of sourcehost, sourceport, destinationhost. Note that in the network flow search module, indexes for sourcehost, sourceport, destinationhost, destinationport, protocol, service, starttime and endtime are generated.

The filtering time for the data set A is about 2.6 seconds and the filtering time for the data set B is about 42.9 seconds. Since an index for the tcpflag column does not exist, the filtering time is proportional to the number of flows. On the other hand, the grouping time for the data set A is about 7.6 seconds and the grouping time for the data set B is about 42.9 seconds. Since indexes for sourcehost, sourceport and destinationhost exist, the grouping time is sublinear to the number of flows.

In conclusion, when there are about 200 million of flows a day, the time to analyse persistent outbound connection is totally about 90 seconds. We believe that our network abnormal behaviour analysis system shows reasonable performance to analyse persistent outbound connection.

TABLE 1. RESPONSE TIME FOR PERSISTENT OUTBOUND CONNECTION

Data Set	# of Flows	Filtering Time (ms)	Grouping Time (ms)	Total Time (ms)
A	11,711,203	2,644	7,632	10,276
B	203,582,430	42,940	48,039	90,979

IV. CONCLUSIONS

In this paper, we propose and implement network abnormal behaviour analysis system. The system has three modules: a network flow generation module, a network flow search module, and a network anomaly behaviour analysis module. The network flow generation module creates network flows in network packets and stores TCP flag information in network flows. The network flow search module, which uses bitmap indexes, supports fast retrieval of network flows. The Network Abnormal Behaviour Analysis module provides a special query for persistent outbound connections to the network flow search module and receives results for persistent outbound connections. We can analyse network abnormal behaviour by supporting other queries such as port scan and TCP SYN flooding as well as persistent outbound connections. Finally, in our experiments, we show that we get query results at a reasonable time with hundreds of millions of flows in a high-speed network.

ACKNOWLEDGMENT

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No.B0101-15-0300, The Development of Cyber Blackbox and Integrated Security Analysis Technology for Proactive and Reactive Cyber Incident Response)

REFERENCES

- [1] Bitmap Index, https://en.wikipedia.org/wiki/Bitmap_index
- [2] Sunoh Choi et.al., Performance Comparison of Relational Databases and Columnar Databases Using Bitmap Index for Fast Search of 10Gbps Network Flows, International Conference on Computer Science and its Applications (CSA), 2015
- [3] InfiniFlux, <http://infiniflux.com>



Sunoh Choi (BS'05-MS'08-Ph.D'14) is a Senior Researcher in the Network Security Research Group at Electronics and Telecommunications Research Institute (ETRI). He has received BS and MS from Korea University and Ph.D from Purdue University. He has joined ETRI, Daejeon, Rep. of Korea, in 2014, and He's working on the project 'Cyber Blackbox development for cyber incidents analysis.' His research interests include Database security, Network security, and Network forensics.



Yangseo Choi (BS'96-MS'00-Ph.D'11) is a Principal Researcher in the Network Security Research Group at Electronics and Telecommunications Research Institute (ETRI). He has joined ETRI, Daejeon, Rep. of Korea, in 2000, and he worked on the anti-cyber terror technology research team. Since 2006, he has been involved in the development of DDoS prevention system, virtual machine security system for Cloud environment. His research interests include Network security, Hacking technologies, and Network forensics



Joo-Young Lee (BS'96-MS'99) is a Principal Researcher in the Network Security Research Group at Electronics and Telecommunications Research Institute (ETRI). She has joined ETRI, Daejeon, Rep. of Korea, in 1999, and she's working on the project 'Cyber Blackbox development for cyber incidents analysis.' Her research interests include Digital forensics, Cryptography, Network security, and Network forensics



Jonghyun Kim (-) is a Principal Researcher in the Network Security Research Group at Electronics and Telecommunications Research Institute (ETRI). He has joined ETRI, Daejeon, Rep. of Korea, in 2007, and He's working on the project 'Cyber Blackbox development for cyber incidents analysis.' His research interests include Digital forensics, Network security, Network forensics, and Visualization



Ikkyun Kim (-) is a director in the Network Security Research Group at Electronics and Telecommunications Research Institute (ETRI). He has joined ETRI, Daejeon, Rep. of Korea, in 2002, and He's working on the project 'Cyber Blackbox development for cyber incidents analysis' and 'Targeted attack Identification and Traceback Technology Development.' His research interests include Digital forensics, Network security, Network forensics, and Visualization