

Introduction to a Network Forensics System for Cyber Incidents Analysis

Yangseo Choi*, Joo-Young Lee*, Sunoh Choi*, Jong-Hyun Kim*, Ikkyun Kim*

*Cyber Security Research Division, ETRI, Daejeon, South Korea

{yschoi92, joolee, suno, jhk, ikkim21}@etri.re.kr

Abstract— Recently, sophisticated attacks are increased against specific business companies, organizations and various facilities and the attackers are trying to remove attack traces such as system logs and related information on the victim systems. Therefore, it is getting more difficult to collect the information for attack analysis. In order to overcome this situations, companies and organizations have started to collect the network traffic as secondary information for attack analysis. However, most of them are focusing on gathering the network packets. But one of the most important parts is to extract the useful information for attack analysis from the collected data. In this paper, we suggest a network forensics system, Cyber Blackbox, which is focused on the traffic analysis.

Keywords—Network forensics, cyber blackbox, attack analysis, network security, information security

I. INTRODUCTION

As the IT environment is improved, various attacks are developed. Some of them were quite simple but some of them were very high skilled attacks. For example, accessing the root privileges in a UNIX system would be a simple attack and APT (Advanced Persistent Threat) [1] attacks are very sophisticated attacks. Especially, these sophisticated attacks have their own specific goals such as collecting important business information and paralyzing target systems. Also, due to their high attack techniques, the characteristics of APT attacks are quite similar to the normal user or application's various activities. For example, network traffic generation activities and connection maintain period are quite same to normal users. It means it is quite difficult to detect those attacks in real time. Usually, due to the detection difficulties, these kinds of attacks are found when the attacks are completed.

When an attack is found, analysts investigate the target system and try to find how it has been intruded by attackers. However, the attackers usually remove most of the attack traces such as system logs and related information. Also the attackers utilize anti-forensic techniques to hide their tracks further. So, it is quite difficult to find out what was happening in the victim system.

In order to overcome this situations, we need another information source which can't be removed by attackers to examine the victim system and attack itself. One of the possible candidates is network traffic. Because, we can find

useful information from the network traffic. We can find who or which system generated a network connection to the victim and what kind of data or files were transmitted to the victim system. That is the reason why companies and organizations have started to collect their network traffic and extract more information from it.

However, network traffic is volatile data. If someone wants to analyse network traffic after an intrusion, it is needed to capture all the network traffic at ordinary times with a high performance network traffic storing system. For that several network traffic storing systems have been developed. Usually, the main goal of these systems is network packet storing. However, as a network forensics system, network traffic analysis for cyber incident investigation is also very important. For that, we have developed a network forensics system called Cyber Blackbox which focuses on security event extracting and attack scenario generation. In this paper, we are going to introduce the design and implementation results for Cyber Blackbox. The Cyber Blackbox that is introduced in this paper is a partial product of whole project, the next version of Cyber Blackbox is now under implementing.

This paper is organized as follows; in chapter 2, we are going to show some related works and in chapter 3, the design of Cyber Blackbox will be shown. The implementation detailed and experimental results will be depicted in chapter 4 and this paper will be finished with conclusion in chapter 5.

II. RELATED WORKS

A. n2disk: ntopng [2]

ntop, a small size company, has developed n2disk and ntopng for traffic recording and traffic analysis respectively. n2disk can capture full-sized network packets at multi-Gigabit rate from a live network interface. n2disk has been designed to write files into disks for very long periods. Users have to specify a maximum number of distinct file that may be written during the execution, and if n2disk reaches the maximum number of files, it will start recycling the files from the oldest one. n2disk stores network traffic as a pcap [4] file.

ntopng is an open-source traffic monitoring application designed for high-speed networks. ntopng's key features are large networks real-time analytics and the ability to characterize application protocols and user traffic behavior. ntopng was validated in various monitoring environments ranging from small networks to ccTLD [19] traffic analysis.

B. FloSIS [17]

FloSIS is a software-based flow storing and indexing system. FloSIS is characterized as the following three aspects. First, it exercises full parallelism in multiple CPU cores and disks at all stages of packet processing. Second, it constructs two-stage flow-level indexes, which helps minimize expensive disk access for user queries. It also stores the packets in the same flow at a contiguous disk location, which maximizes disk read throughput. Third, it optimizes storage usage by flow-level content deduplication at real time. The evaluation in [17] shows that FloSIS on a dual octa-core CPU machine with 24 HDDs achieves 30 Gbps of zero-drop performance with real traffic, consuming only 0.25% of the space for indexing. FloSIS is totally a network traffic storing system, so the attack analysis area is not included at all.

C. OmniPeek – Savvius [18]

When OmniPeek was developed by Savvius for the first time, it was designed for network management system. At that time Savvius was known as Wild Packet. However, it has been changed their company name and become one of the biggest network forensics system development companies in the whole world. They have changed their company name when they have decided to strengthen their security related part in OmniPeek.

There are various models of OmniPeek products. They implemented various models for various network types and performance. The OmniPeek Enterprise combines all Savvius network analysis features in a single product. It's both a stand-alone network analyzer for local captures and a console for analyzing traffic collected from multiple interfaces on multiple capture engines and Omnipliances. Even though, it is developed as a network management system, it's been quite a while since the first version of OmniPeek was developed and this product is quite stable.

D. NetWitness [3]

NetWitness by RSA is composed with a Decoder, Concentrator and Broker. Decoder is a network traffic collector. Concentrator is a meta data generator about collected network traffic. And, Broker is a query processor against multiple Concentrator. In order to analyze network traffic, Investigator is used. It collects all the needed information via Broker and analyzes the collected information. It can analyze Layer 2 to 7.

III. DESIGN OF CYBER BLACKBOX

A. Summary of Cyber Blackbox

Cyber Blackbox is designed to operate as a network forensics system [5, 6] which collects network traffic to use it as evidence and generates useful information and analyzes the collected data. It extracts various information from the collected network packets for attack analysis and supports users can search specific information what they want to find. Additionally, it generates attack scenarios based on the extracted information. In order to collect the network traffic as an evidence Cyber Blackbox guarantees integrity and

confidentiality. The conceptual goals of Cyber Blackbox is shown in Figure. 1.

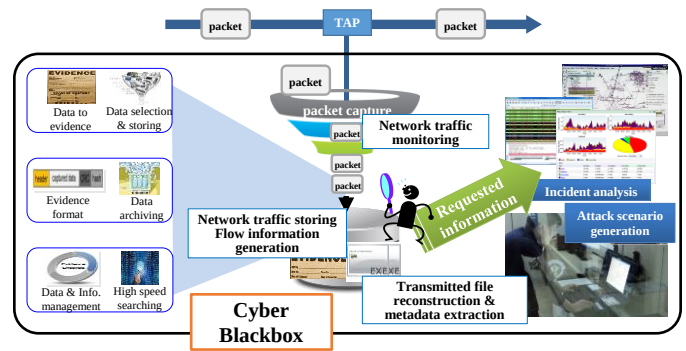


Figure 1. Conceptual goals of Cyber Blackbox

B. Construction of Cyber Blackbox

There are two subsystems in Cyber Blackbox, and they are Cyber Blackbox Sensor (CBS) and Cyber Blackbox Manager (CBM) subsystems. The relationship between CBS and CBM is shown in Figure 2.

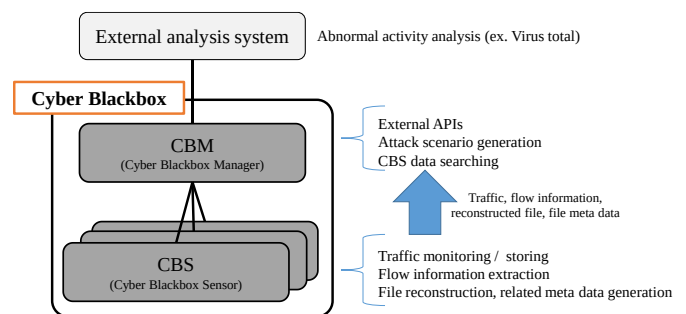


Figure 2. Relationships between subsystems of Cyber Blackbox system

The main objective of CBS is to collect network data and extract related information. All the collected network traffic, flow information, reconstructed files and related metadata are generated and stored in CBS in real-time.

CBM is designed to provide various interfaces to users for information search and attack scenario generation. CBM can collect data from CBS and analyze it. Also CBM can manage multiple CBS simultaneously. When a user wants to analyze some network traffic and related information then the user can connect to a CBM and gather the information from dedicated CBSs and proceeds the analysis. Also, CBM can cooperate with external systems to collect more information.

Main functionalities of Cyber Blackbox which is shown in Figure 3 could be listed as follows;

- Network traffic storing and flow information generation.
- Network transmitted file reconstruction and related metadata generation.
- Integrity preserving of collected data and data management.
- Cyber incidents analysis based on the collected network traffic and related information
- Incident information sharing

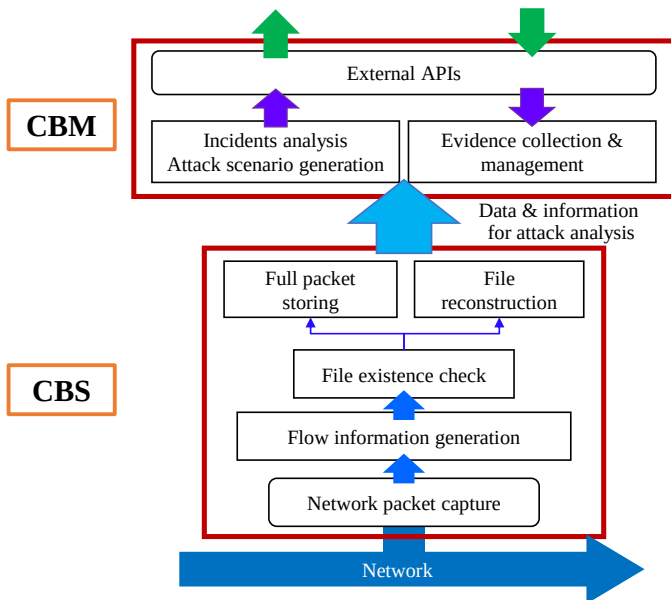


Figure 3. Construction of Cyber Blackbox system

IV. IMPLEMENTATIONS OF CYBER BLACKBOX

A. CBS subsystem

Among the main functionalities that we mentioned in Chapter 3, main roles of CBS are network traffic storing and flow information generation, network transmitted file reconstruction and related metadata generation, and integrity preserving of collected data and data management. In order to implement those three functionalities we have developed “Traffic & Flow Information Gathering Block (TFGB)” and “Transmitted File Reconstruction Block (TFRB)”

In this version of Cyber Blackbox, CBS is designed to collect network traffic with 1Gbps performance. For that, the hardware specifications are depicted in Table 1.

TABLE 1. SYSTEM SPECIFICATIONS FOR CYBER BLACKBOX

	Type	Specification
SW	OS	Linux CentOS 7.0
	Kernel version	3.10.0-123.el7.x86_64
HW	CPU	Intel 2.6GHz(IVY Bridge) 2EA
	Board	CAR-5031(caswell)
	Memory	DDR3 64GByte (8GB x 8)
	HDD	SATA 2TB (1TB x 2)
	SDD	SSD 1TB x 1
	NIC	1000BASE-LR/SR x 4 port, SFF Connector SMF/MMF(Fiber)
	Management NIC	1Gbps Ethernet x 2 port

1) Traffic & Flow Information Gathering Block (TFGB)

TFGB collects network traffic via network interface card and generates flow information from captured network traffic. When TFGB collects traffic, it utilizes DPDK (Data Plane Development Kit) [8] and the storing format for gathered

traffic is quite similar to pcap format. We just added 20 bytes additional information header to the pcap file for file management and integrity preserving. The header contents are shown in Table 2.

TABLE 2. TRAFFIC FILE HEADER

No.	Type(Size in bytes)	Contents
1	Version(byte:1)	Version of added header
2	Rsvd1(byte:1)	Reserved field 1
3	System_id(short:2)	Identification number of CBS
4	Data_count(int:4)	Number of stored packets
5	Total_length(int:4)	Total length of file
6	Timestamp(int:4)	File creation time
7	Hash_algo_id(byte:1)	Captured file hashing algorithm
8	Rsvd2(byte:1)	Reserved field 2
9	Seq_num(short:2)	Sequence number for stored file

In order to preserve the integrity of the stored pcap file TFGB calculate hash value of the pcap file and save it. The hash algorithm could be identified from the “Hash_algo_id” field (0:MD5, 1:SHA256). With this hash value, we can check the integrity of stored pcap files. In order to preserve this integrity characteristic, the hash value should not be changed. For that the hash values are supposed to be stored in a WORM (Write Once Read Many) [9] storage area. Virtual volumes in CBS are used for WORM storage. Those virtual volumes can be created and deleted, however the files in the virtual volume could be modified.

As we mentioned, when traffic is collected, flow information is generated as well. The generated information is shown in Table 3. Flow information is useful information for incident analysis. When the flow information is generated, it is stored as the sqlite [10] file format. Then a user can access this file via sqlite query statement. We adapted sqlite due to the good performance of batch process. With sqlite we can insert more than 50,000 flow information per second [11].

TABLE 3. GENERATED FLOW INFORMATION

No.	Type(Size in bytes)	Contents
1	sip(byte:4)	Source IP Address
2	dip(byte:4)	Destination IP Address
3	sport(short:2)	Source Port
4	dport(short:2)	Destination Port
5	action(byte:1)	Action Field
6	protocol(byte:1)	Protocol number
7	sys_id(short:2)	System ID
8	service_id(short:2)	Service ID
9	app_id(short:2)	App ID
10	sp_id(short:2)	Service Provider ID
11	dev_id(short:2)	Device ID
12	stime(int:4)	Start Time of Flow
13	etime(int:4)	End Time of Flow
14	inpkt(int:4)	Inbound Packets
15	inbyte(long:8)	Inbound Bytes
16	outpkt(int:4)	Outbound Packets
17	outbyte(long:8)	Outbound Bytes

Also, when a packet is captured TFGB investigate whether there is a PE [16] file or not. If a PE file is in the packet, then TFGB stores the session information (5-tuple) and sends the packet to TFRB. Then TFRB reconstructs the PE file from the packets to a real file. Afterwards, if the packets that are included in the same session, then TFGB sends the packet to TFRB automatically.

We have implemented TFGB which can collect 1Gbps network traffic in real-time. In order to verify its traffic collection performance we have tested it with N2X [12] system. CBS shows 2Gbps performance (1Gbps x 2port (in & out)). We have generated 512byte size packets as 1Gbps bandwidth with 2 ports for 30 seconds. The statistic data for the test is shown in Figure 4.

Port	Tx Test Packets	Rx Test Packets	Tx Test Octets	Rx Test Octets	Tx Test Throughput (Mb/s)	Rx Test Throughput (Mb/s)	Rx Packet Loss	Average Latency (us)
All Ports	14097746	14097746	6964286524	6964286524	1741.072	1741.072	n/a	742.36
103/3	7048873	7048873	3482143262	3482143262	870.536	870.536	n/a	743.31
103/4	7048873	7048873	3482143262	3482143262	870.536	870.536	n/a	741.40

Figure 4. Performance test results (1Gbps)

Also, the stored traffic and the number of captured packets are shown in Figure 5. From Figure 4 and 5, we can identify that all the transmitted packets are captured and stored correctly.

512byte_2G_135301	2015-02-06 오후...	Wireshark capture...	2,349,840KB
512byte_2G_135311	2015-02-06 오후...	Wireshark capture...	2,404,690KB
512byte_2G_135321	2015-02-06 오후...	Wireshark capture...	2,404,688KB
512byte_2G_135331	2015-02-06 오후...	Wireshark capture...	54,866KB
File: "C:\Users\user\Desktop\512byte_2G_135301.pcap"	2406 MB 00...	Packets: 4592052	
File: "C:\Users\user\Desktop\512byte_2G_135311.pcap"	2462 MB 00...	Packets: 4699240	
File: "C:\Users\user\Desktop\512byte_2G_135321.pcap"	2462 MB 00...	Packets: 4699236	
File: "C:\Users\user\Desktop\512byte_2G_135331.pcap"	56 MB 00:0...	Packets: 107218	

Figure 5. Number of captured and stored packets

2) Transmitted File Reconstruction Block (TFRB)

TFRB reconstructs PE files which are transmitted via network. When a packet is suspected of containing a PE file then the packet is handed over to TFRB from TFGB. Then TFRB investigates several packets that are in the same session and decides whether the packets contain a real PE file or not.

Currently, TFRB only can reconstruct PE files. However, the target file types will be extended. PE stands for portable executable and it is the executable file for Microsoft Windows operating system such as Windows XP, Windows 7, Windows 2000, etc. Due to the popularity of PE for malwares we have selected PE file for our first reconstruction target.

We have selected PE file for our reconstruction target because almost all the malwares are based on the Microsoft Windows systems such as Windows XP, Windows 2000 and Windows 7. These days the malware Linux malware and mobile malware also implemented however still the most part of malwares are PE files. When a PE file is reconstructed then the following information which is shown in Table 4 is generated at the same time.

TABLE 4. META INFORMATION FOR RECONSTRUCTED PE FILE

No.	Type(Size in bytes)	Contents
1	Hash(byte:16)	Source IP Address
2	File_name(byte:32)	Destination IP Address
3	File_size(int: 4)	Source Port
4	Stime(time_t:4)	Destination Port
5	Etime(time_t:)	Action Field
6	sip(byte:4)	Source IP Address
7	dip(byte:4)	Destination IP Address
8	sport(short:2)	Source Port
9	dport(short:2)	Destination Port
10	Protocol(byte:1)	Protocol number

B. CBM subsystem

The main functionalities of CBM are cyber incidents analysis based on the collected network traffic and related information and incident information sharing. In order to implement those two functionalities we have assigned two blocks. They are "Intrusion factor Analysis and Scenario generation Block (IASB)" and "Cyber Blackbox Interface Block (CBIB)".

1) Intrusion factor Analysis and Scenario Generation Block (IASB)

At first, for network traffic investigation, we need to create a "case" with specific time period information. Then IASB loads corresponding flow information and PE file reconstruction metadata. When the corresponding information is loaded, the investigator can explore the information to find specific clues of attacks. IASB can display flow information as shown in Figure 6 and 7.

The first scene of IASB is shown in Figure 6. It shows the full flow information of the selected period. From that scene, we can filter out or select some meaningful flows based on the information and experiences that the investigator has. Basically, the flows could be selected based on the IP address, protocol, time, PE file detection results, port number, etc. The filtered flow information could be shown as in Figure 7.

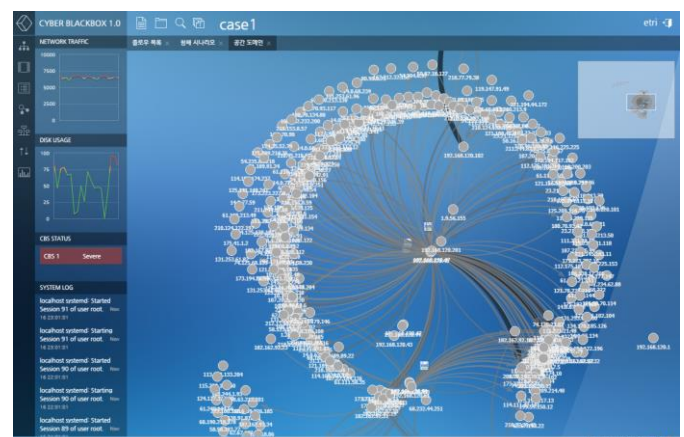


Figure 6. Space domain area for selected flow information

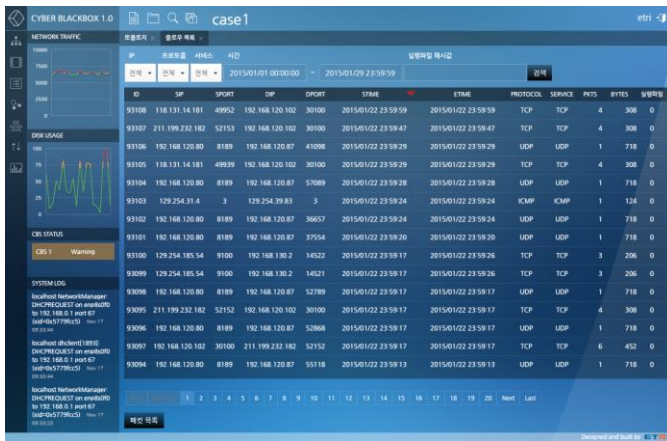


Figure 7. List of filtered flow information

As we mentioned, from the selected information a user can narrow down the target flow information and finally check the last flows in space domain area as shown in Figure 8. It means a user can directly identify suspicious flows and their connections at the same time.



Figure 8. Filtered flow information in space domain

2) Cyber Blackbox Interface Block (CBIB)

In our system, we have implemented the information sharing interface between Cyber Blackbox and external malware analysis infrastructure such as virus total [7] and malicious code analysis infrastructure from KISA (Korean Internet & Security Agency) [15]. With this interface, users can share executable files which is reconstructed by TFRB in CBS. When Cyber Blackbox submits an executable file to external malicious code analysis infrastructure, it returns their malicious information after analyzes it.

V. CONCLUSIONS AND FUTURE WORKS

In this paper, we have proposed a new network forensics system, Cyber Blackbox. The main objectives of Cyber Blackbox are to collect full network packets in real time and provide some interfaces to users for attack analysis with collected data and information. In order to make the collected traffic into an evidence, integrity preserving technology for

collected traffic has been developed. Also, to extend the attack analysis boundaries we have developed the interface APIs between Cyber Blackbox and external malicious code analysis infrastructures.

Cyber Blackbox is constructed with two subsystems, CBS (Cyber Blackbox Sensor), CBM (Cyber Blackbox Manager) and the roles for each subsystem is to collect traffic and data by CBS and analyze collected data and information by CBM.

Currently, Cyber Blackbox is implemented which can cover 1Gbps network bandwidth. However, we are developing the next version of Cyber Blackbox and its target is 20Gbps. Besides, target reconstruction file types will be increased. Now we are developing the file reconstruction technology which can reconstruct HTML, PDF, ZIP, APK, JPG, GIF, JAR, and so on.

ACKNOWLEDGMENT

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

REFERENCES

- [1] A Websense White Paper, *Advanced Persistent Threats and other advanced Attacks: Threat Analysis and defense strategies for smb, mid-size, and enterprise organizations*, Available: <https://www.websense.com/assets/white-papers/whitepaper-websense-advanced-persistent-threats-and-other-advanced-attacks-en.pdf>
- [2] Luca Deri, Maurizio Martinelli, Alfredo Cardigliano, *Realtime High-Speed Network Traffic Monitoring Using ntopng*, 28th Large Installation System Administration Conference (LISA14), Nov. 2014, pp. 78-88
- [3] NetWitness, Available: <http://www.emc.com/security/security-analytics/security-analytics.htm>
- [4] Tcpdump & Libpcap. Available: <http://www.tcpdump.org/>
- [5] WildPackets, *Real-World Security Investigations with Network Forensics*, White Paper
- [6] Wikipedia, *Network forensics*, Available: https://en.wikipedia.org/wiki/Network_forensics
- [7] Virus Total, <https://www.virustotal.com/>
- [8] Intel, *Intel DPDK: Data Plane Development Kit*, Available: <http://dpdk.org/>
- [9] WORM, *Write once read many*, Available: https://en.wikipedia.org/wiki/Write_once_read_many
- [10] SQLite, Available: <https://www.sqlite.org/>
- [11] *Improve INSERT-per-second performance of SQLite?*, Available: <http://stackoverflow.com/questions/1711631/improve-insert-per-second-performance-of-sqlite>
- [12] Agilent, *N2X Packets and Protocols Getting Started Guide*, Available: http://staff.ii.pw.edu.pl/n2x/N2X_Getting_Started_Guide.pdf
- [13] *Microsoft Portable Executable and Common Object File Format Specification*, Feb. 1999, Microsoft Corporation, Available: <http://download.microsoft.com/download/e/b/a/eba1050f-a31d-436b-9281-92cdfae4b45/pecoff.doc>
- [14] Deri L., Cardigliano A., and Fusco F., *10gbit line rate packet-to-disk using n2disk*, IEEE INFOCOM Workshop on Traffic Monitoring and Analysis(2013)
- [15] KISA, *Korea Internet & Security Agency*, <http://www.kisa.or.kr/eng/main.jsp>
- [16] Microsoft, *Microsoft Portable Executable and Common Object File Format Specification – Revision 8.3*, Microsoft, Feb. 2013

- [17] Jihyung Lee, Sungryoul Lee, Junghee Lee, Yung Yi, KyoungSoo Park, *FloSIS: A Highly Scalable Network Flow Capture System for Fast Retrieval and Storage Efficiency*, In Proceedings of the USENIX Annual Technical Conference (USENIX ATC'15), Santa Clara, CA, July 2015
- [18] Omnippeek, Available: https://www.savvius.com/products/overview/omnippeek_family/omnippeek_network_analysis
- [19] cc TLD, *Country code top-level domain*, Available: https://en.wikipedia.org/wiki/Country_code_top-level_domain



Sunoh Choi (BS'05-MS'08-Ph.D'14) is a Senior Researcher in the Software Contents Research Laboratory at Electronics and Telecommunications Research Institute (ETRI). He is 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, Attack detection and analysis, Digital forensics, and Network forensics.



Yangseo Choi (BS'96-MS'00-Ph.D'11) is a Principal Researcher in the Software Contents Research Laboratory at Electronics and Telecommunications Research Institute (ETRI). He 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.



Jonghyun Kim (-) is a Principal Researcher in the Software Contents Research Laboratory at Electronics and Telecommunications Research Institute (ETRI). He is 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.



Joo-Young Lee (BS'96-BS'99) is a Senior Researcher in the Software Contents Research Laboratory at Electronics and Telecommunications Research Institute (ETRI). She is 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.



Ikkyun Kim (-) is a Principal Researcher in the Software Contents Research Laboratory at Electronics and Telecommunications Research Institute (ETRI). He is 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.