ISSN:2321-1156
International Journal of Innovative Research in Technology & Science(IJIRTS)

DESIGN OF A CYBER THREAT INTELLIGENCE FRAMEWORK

NakHyun Kim, Korea Internet & Security Agency; Byung-ik Kim, Korea Internet & Security Agency; Seulgi Lee, Korea Internet & Security Agency; Hyeisun Cho, Korea Internet & Security Agency; Jun-hyung Park, Korea Internet & Security Agency;

Abstract

This paper reviews the trends of cyber threat intelligence (CTI) technology that enables preemptive detection of cyber attacks and threats, which become intelligent and advanced, and responds to them effectively; and analyzes CIT-related products, standards, and frameworks. Based on the review and analysis, this paper provides comprehensive information on the components and structure of CTI technology to enterprises and researchers that want to introduce and develop CTI technology by proposing a cyber threat intelligence framework that can express various types of CTI structures. Cyber threat intelligence is a technology that creates intelligence to respond to cyber attacks and threats that occur now, will occur, or can occur (potential), based on large amounts and heterogeneous data related to cyber incidents and threats. Cyber threat intelligence is emerging as a technology that can effectively respond to cyber incidents that are on the rise in terms of quality and quantity.

I. Introduction

We need to understand the difference between data, information, and intelligence in order to understand cyber threat intelligence. A security consulting firm named Security Architect Partners defines data as an individual item that has atomicity, and information as processed data, that is, data with given meaning. Intelligence is defined as "information about how to detect and defend against cyber incidents and threats by adding the analysis and evaluation information of experts to the data with given meaning." [23]

This definition can be explained with illustration from the perspective of cyber incident analysis in such a way that "data is the IP, domain, URL, or e-mail that can be collected from the web or network." In addition, information can be described as the URL exploited for phishing, domain that spreads malicious code, and IP that establishes C&C communication with malicious code. Cyber threat intelligence is the comprehensive analysis result which reports that "An attack group Y mainly attacked financial companies, and the malicious code A recently was found to be a variant of the malicious code A when the recently used malicious code A is analyzed. Therefore, actions are required to block the IP address of the C&C server frequently used by the malicious code A." Gartner, an information technology research and consulting firm in the U.S., defined CTI as "evidence-based knowledge, including context, mechanisms, indicators, implications and actionable advice, which can be used to respond to menace or hazard to assets." [21] In addition, Cyber Squared defined it as "a new information security area that explains how and why an intelligent cyber attacker becomes a threat." [22]

A. Trends of CTI

As cyber attacks become more sophisticated, it becomes more difficult to detect and analyze an cyber incident, and cyber threat intelligence technology draws attention as an alternative. Global cyber security companies adopt cyber security intelligence technology in their products and release services that provide cyber threat intelligence. FireEye acquired a CTI start-up called iSIGHT Partners, [1] and Symantec released an enterprise CTI service called DeepSight Intelligence. [2] Check Point released cloud-based CTI information purchase service THREATCLOUD Intelligence, [3] and Facebook opened the ThreatExchange project for CTI sharing. [4] In addition, global enterprises like IBM applied the CTI technology to their security equipment and products (e.g., X-Force Threat Intelligence [5]). Likewise, the Office of the Director of National Intelligence (ODNI) in the U.S. established the Cyber Threat Intelligence Integration Center (CTIIC) in February 2015 to analyze national cyber threats, [6] and the Department of Homeland Security (DHS) makes efforts to respond to cyber attacks using cyber threat intelligence technology by establishing the National Cybersecurity and Communications Integration Center (NCCIC) to analyze cyber threats and share information. [7] The Korea Internet & Security Agency in Korea also continues to struggle to respond to cyber threats using CTI by launching a global cyber threat intelligence network. [8] As explained, cyber threat intelligence is regarded as a new security technology to respond to cyber threats effectively.

B. CTI-related standards

This chapter reviews the STIX standard of OASIS (standardization organization) among CTI-related standards as the information description standard to share cyber threat information.

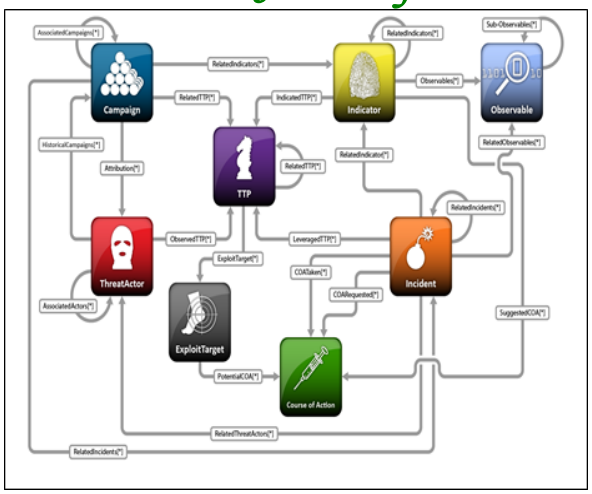
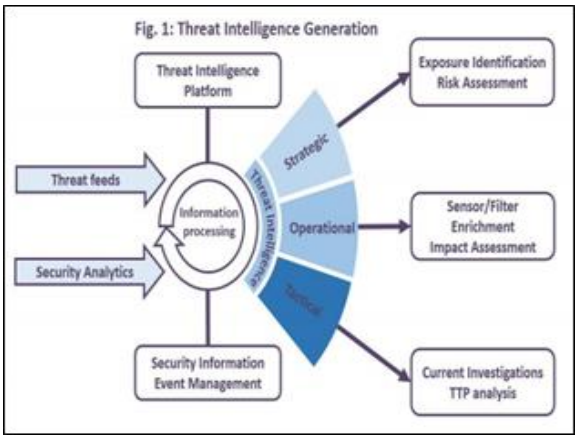


Figure 1. Major data models of Structured Threat Information eXpression (STIX)

STIX is a description system to share cyber threat information and is composed of eight major data models. [9] The major data model of STIX can be briefly described as follows: Observable (observed attack event), Indicator (information corresponding to a threat among Observable), Incident (Indicator that was founded to be a cyber attack), TTP (attack technique, tactics, and procedure of Incident), ThreatActor (TTP execution subject), Campaign (attack composed of multiple incidents and TTP), ExploitTarget (vulnerability exploited by the attacker to execute TTP), CoA (response to vulnerabilities and cyber incidents), etc. [Figure 1] shows the relationship among key STIX data models. In addition, the CTI information described using STIX can be shared using the information transmission system called Trusted Automated eXchage of Indicator Information (TAXI).

[Figure 2] shows the CTI platform of Splunk, a big data analysis company, [10] and it shows the composition of the CTI framework, such as data source and collection management, data classification, correlation analysis, and search.

The 2015 CTI-related report of the cyber incident response center in the U.S. [11] describes the CGI generation procedure as shown in [Figure 3]. Threat Feeds and Security Analytics are entered and information is processed and analyzed with the security intelligence platform and SIEM to generate threat intelligence. Generated threat intelligence can be used at the strategical, operational, and tactical levels. The strategic utilization method of threat intelligence refers to risk exposure identification and risk assessment, whereas the operational utilization method refers to the security rule tightening of security and network equipment, and risk impact assessment. As an example of the tactical utilization method, the attack technique of the cyber incident under investigation was presented.



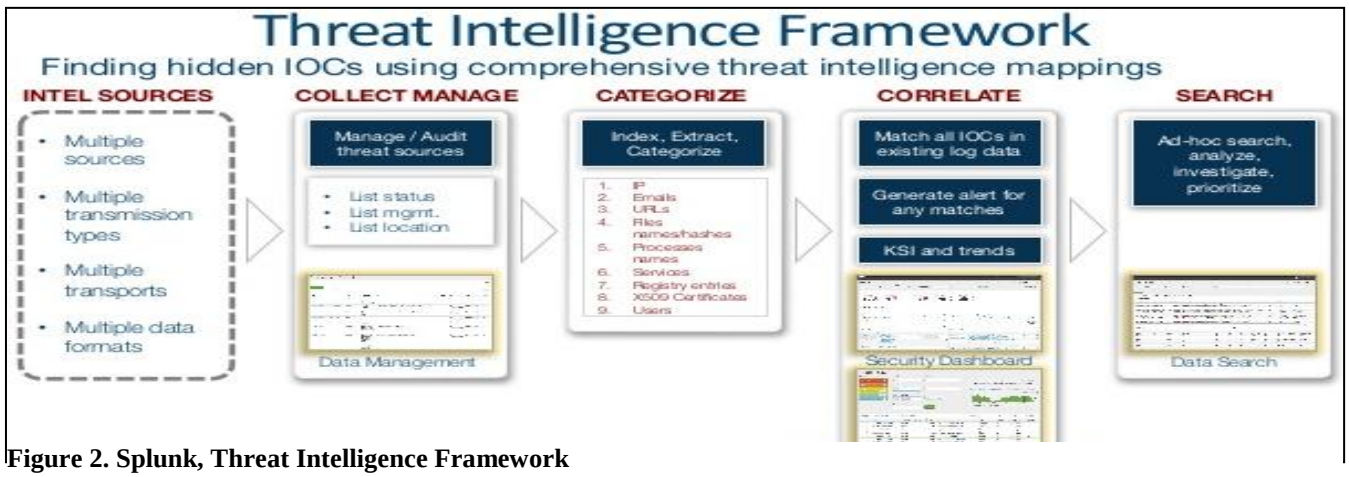


Figure 2. Splunk, Threat Intelligence Framework

C. CTI Framework Reference

Figure 3. CTI Generation Procedure, CERT UK

II. Proposed CTI Framework

[Figure 4] shows the cyber threat intelligence framework proposed by this paper. The framework is broadly composed of collection, analysis, and result utilization and sharing. The below section describes the details of each component.

A. CTI Information Collection

The type and method of collecting the CTI information is described in the "Collection" part of [Figure 4]. The collection part is composed of the internal/external collection channel and collected data processing. Various collection channels are classified and grouped by information type and source, and the framework is composed in such way that the log of the internal IT infrastructure and security equipment and setting related data as well as the data related to the vulnerability of installed S/W and H/W can be collected. As the quality of data that can be collected through analysis is determined by the quantity of collected data and quality of collected/processed data, the introduction of architecture such as big data storage and graph databases should be considered in order to process large amounts and heterogeneous data. The data collected, stored, and processed in this way is classified by the type of infrastructure-related information that is exploited for cyber incidents (e.g., IP, domain, hash), and is transferred to the analysis process.

B. CTI Analysis

The analysis part of the CTI framework is the core process to generate CTI, and various analysis techniques can be used according to the information related to the cyber incident (e.g., IP, domain, malicious code) based on the data-mining technique. The correlation of information, which was not found with the unit information related to cyber incidents up to now, can be identified by the analysis procedure and provide insights into cyber threats at present or in the future. The analysis process includes both manual and automatic analysis, and can reduce the time to analyze threats and respond to cyber incidents by replacing existing manual analysis and by integrating AI-related technologies, such as machine learning and deep learning, with calculation theories, such as case-based inference and evidence-based inference. However, the automatic analysis technology has limits which are based on the detection of a particular attack pattern as the technique of cyber attacks becomes more sophisticated. Therefore, active analysis support functions should be provided, such as presenting the analysis target of security experts and providing notification about high-risk threats that should be analyzed first.

C. Utilization and Sharing

Threats to internal assets are identified and the response procedure is determined in the result utilization and sharing part based on CTI generated by the CTI analysis part. New threats can be preemptively responded to using CTI and used as the information to support decision-making for preemptive response. CTI can also be used to advance rules about an organization's network and security equipment and update security features like vulnerability patches. Based on cyber attack information in the past, current or future cyber threats can be understood, and visibility about the entire cyber threat (attack) can be secured by sharing CTI. In addition, contributions can be made to improve the capability of responding to cyber threats at the national level by sharing CTI information generated by the Computer Emergency Response Team (CERT) using the sharing platform.

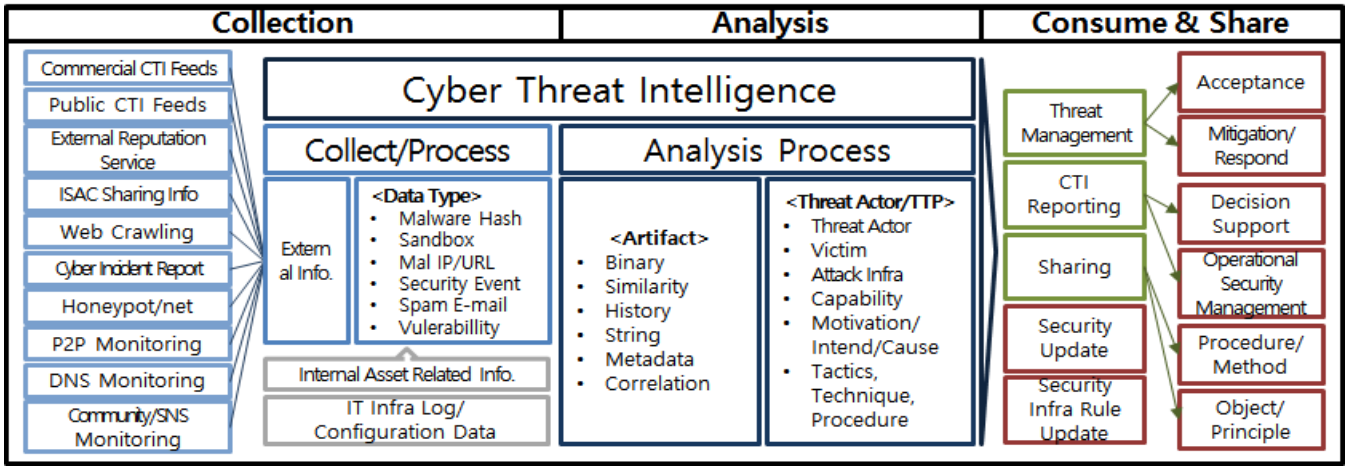


Figure 4. Proposed CTI Framework

D. Architecture Design

Large-scale data needs to be collected and analyzed to generate CTI, such as the network data inside the organization, detection results of security equipment, log data of server equipment, and Open Source Intelligence (OSINT) data that can be collected from the outside. This section proposes the architecture design of the CTI framework based on Hadoop, which is a big data platform to collect and analyze large-scale data, and the graph database, which is suitable for analyzing the correlation of cyber incidents.

Table 1. Test Data (Resource)

RDB Table Name	GDB Label Name	DB Object Count(Node)
tb_resource_id	:RESOURCE	7,803,897
tb_attribute_id	:ATTRIBUTE	880,321
tb_dic_ip	:DIC:IP	17,498,555
tb_dic_hash	:DIC:HASH	8,020,710
tb_dic_domain	:DIC:DOMAIN	5,586,714
tb_dic_aid	:DICA	12,973,518
Total		53,763,715

Table 2. Test Data (Relation)

RDB Table Name	GDB Label Name	DB Object Count(Node)
tb_resource_relationship	:HAS_REL	7,880,508
tb_attribute_relationship	:HAS_ATTRIBUTE	12,540,758
tb_dic_ip	:MAPS_TO_DIC	18,498,555
tb_dic_hash	:MAPS_TO_DIC	8,020,710
tb_dic_domain	:MAPS_TO_DIC	5,586,714
tb_dic_aid	:MAPS_TO_DICA	12,973,518
Total		65,500,763

III. Conclusion

A. Performance comparison of the relational DB and graph DB

The performance of the database that actually stores and manages data is an important factor in the CTI system, which should collect and analyze large-scale data. The performance of the graph database used by the architecture, which is proposed by this paper, and the existing relational database was compared using the data to analyze actual cyber incidents and by executing queries used for the actual analysis function. The performance of the proposed

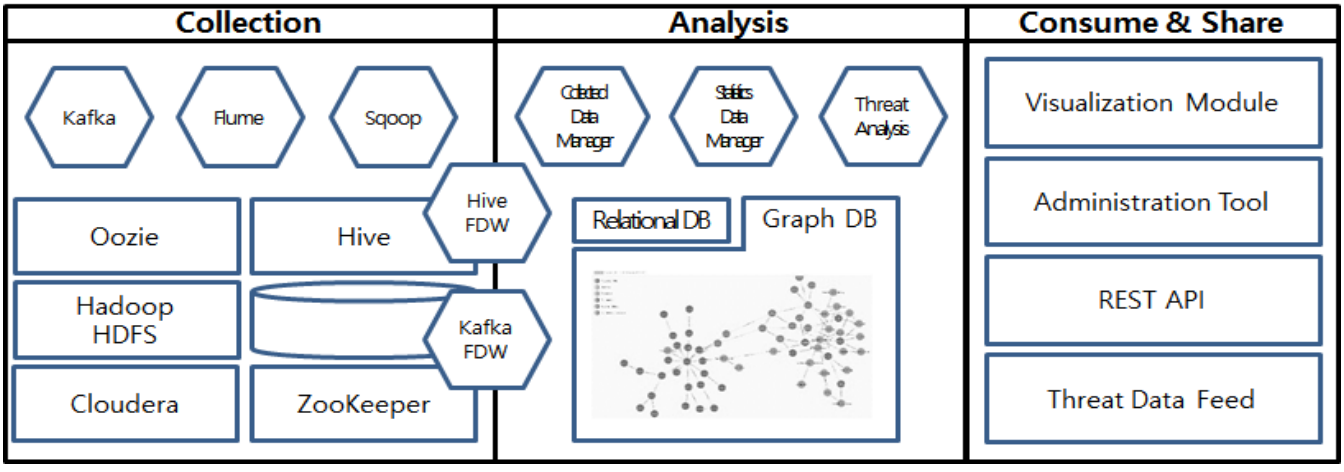


Figure 5. CTI System Architecture

architecture was partially measured by comparing the performance of the relational database and graph database and this can be used to decide the usefulness of the proposal. The graph database has data structure that is suitable for describing the correlation of various attack resources exploited for cyber incidents. The graph database retrieves the data using the cypher query. It was found that the graph database has performance superiority over the relational database when performance is compared with the relational database based on the same query using the correlational cyber incident data.

B. Conclusion and Future Study

This paper proposes a new CTI framework and a system architecture by analyzing the trends of CTI technology at home and abroad. This draws attention as it is a technology to cope with cyber incidents, related standards and frameworks, and CTI-generation procedures. The CTI framework is composed of the data collection and processing part, data analysis part for CTI generation, and generated CTI sharing and utilization part. To check the usefulness of the proposed architecture, the performance of the graphic database included in the architecture was compared with the relational database, and the performance of the proposed architecture in a big-data environment was estimated. Improvement methods will be figured out and reflected by comparing and analyzing CTI-related solutions at home and abroad to see whether the proposed framework can accommodate various CTI technologies. In addition, more studies will follow regarding the technology of processing large amounts of heterogeneous data that compose the frameworks, data mining analysis techniques, correlation analysis among data, improving the reliability of generated CTI information, and the method of estimating the risk level of the cyber incident related information exploited for cyber attacks.

Type	SQL(Structured Query Language)	CQL(Cypher Query Language)	Response time	
			RDB	GDB
Query#1	SELECT relationship kind, COUNT(DISTINCT aid) cnt FROM tb_attribute_relationship WHERE rid = 68347 AND time >= 20160101 AND time <20160901 GROUP BY relationship;	MATCH (a:RESOURCE)<-[r:HAS_ATTRIBUTE]-(b:ATTRIBUTE) WHERE a.rid = 68347 AND 20160101 <= r.time <20160901 RETURN r.relationship AS kind, count(distinct b.aid);	0.13	0.06
Query#2	SELECT a.kind, COUNT(DISTINCT a.rid) cnt, a.r_time FROM (SELECT relationship kind, r_to rid, r_time FROM tb_resource_relationship WHERE r_from = 68347 UNION SELECT relationship kind, r_from rid, r_time FROM tb_resource_relationship WHERE r_to = 68347) a WHERE a.r_time >= 20160101 AND a.r_time <20160901 GROUP BY a.kind;	MATCH (a:RESOURCE)-[r:HAS_REL]-(b:RESOURCE) WHERE a.rid = 68347 WITH r.relationship AS kind, r.r_to AS rid, r.r_time AS r_time WHERE 20160901 >r.time >= 20160101 RETURN kind, count(distinct rid), r_time;	1.83	0.05
Query#3	SELECT rid, type_id, type, value FROM tb_resource_id WHERE rid IN (SELECT DISTINCT a.rid FROM (SELECT r_from rid FROM tb_resource_relationship WHERE r_to = 68347 AND relationship = 'mapping' AND r_time >= 20160101 AND r_time <20160901 LIMIT 15 UNION SELECT r_to rid FROM tb_resource_relationship WHERE r_from = 68347 AND relationship = 'mapping' AND r_time >= 20160101 AND r_time <20160901 LIMIT 15) a);	MATCH (a:RESOURCE)-[r:HAS_REL]-(b:RESOURCE) WHERE a.rid = 68347 AND r.relationship = 'mapping' AND 20160901 >r.time >= 20160101 RETURN b.rid, b.type_id, b.type, b.value	1.95	0.06

Acknowledgments

This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT)(2017-0-00158, Development of Cyber Threat Intelligence(CTI) analysis and information sharing technology for national cyber incident response.)

References

[1] https://www.fireeye.kr/company/press-releases/2016/fireeye-announces-acquisition-of-isight-partners.html
[2] https://www.symantec.com/services/cyber-security-services/deepsight-intelligence
[3] https://www.checkpoint.com/threat-prevention-resources/
[4] https://www.facebook.com/threatexchange
[5] IBM Security, X-Force Threat Intelligence, https://www-03.ibm.com/security/xforce/
[6] https://www.dni.gov/index.php/about/organization/ctiic-who-we-are
[7] https://www.dhs.gov/national-cybersecurity-and-communications-integration-center
[8] http://www.yonhapnews.co.kr/bulletin/2016/06/22/-0200000000AKR20160622166700017.HTML
[9] https://www.oasis-open.org/standards#stix1.2.1

[10] https://www.splunk.com/ko_kr/resources/video.NnczN4MjE6qVYs873PJ4NL2w8Vp8DTj5.html
[11] An Introduction to threat intelligence, CERT UK, 2015.
[12] Cyber Threat Intelligence, Nettitude, 2016.6.16. http://www.thinklabsmedical.com/.
[13] White Paper, Intelligence-Led Security, IDC, Robert Ayoub, Christina Richmond, Michael Versace, March, 2016.
[14] Operational Threat Intelligence, Technical Operations & Program Integration, Mandiant, A FireEye Company, 2016.
[15] Evolution of Cyber Threat Intelligence, BLUE COAT, Bret Jordan, ENISA 2015.
[16] Automated Network Defense through Threat Intelligence and Knowledge Management, Christopher O'brien, 2015.
[17] Implementation Framework Cyber Threat Prioritization, Troy Townsend, Jay McAllister, 2013.
[18] Conceptual Framework for Cyber Defense Information Sharing within Trust Relationships, Diego Fernandez Vazquez, Oscar Pastor Acosta, 4th International Conference on Cyber Conflict, 2012.
[19] Cybersecurity Information Sharing: A framework for information security management in U.K. SME supply chains, Lewis, Riyana et al, Twenty Second European Conference on Information Systems, Tel Aviv 2014.
[20] Investigating the Dark Cyberspace: Profiling, Threat-Based Analysis and Correlation, Claude Fachkha, Elias Bou-Harb, Amine Boukhtouta, Son Dinh, Farkhund Iqbal, Mourad Debbabi, 7th International Conference on Risks and Security of Internet and Systems (CRiSIS), 2012.
[21] Cyber security information exchange to gain insight into the effects of cyber threats and incidents, F. Fransen, A. Smulders, R. Kerkdijk, Elektrotechnik & Informationstechnik DOI 10.1007/s00502- 015-0289-2, 2015.
[22] https://www.gartner.com/doc/2487216/definition-threat-intelligence
[23] http://www.cybersquared.com
[24] http://security-architect.com/is-threat-intelligence-a-misnomer/

Biographies

NAKHYUN KIM received the B.S. degree in Computer Science from the University of Seoil, Korea, in 2008, the M.S. degree in Network Security from the University of SoongSil, Korea, in 2011, respectively. Currently, He is a Deputy General Researcher of Security Technology R&D Team 1 at Korea Internet & Security Agency. His research areas include cyber threat analysis, cyber attack related data correlation, and Sensor Network Security. Nakhyun Kim may be reached at knh@kisa.or.kr

BYUNG-IK KIM received the B.S. degree in Computer Science from the University of Ajou, Korea, in 2010. Currently, He is a Deputy General Researcher of Security Technology R&D Team 1 at Korea Internet & Security Agency. His research areas include cyber threat analysis, cyber attack related data correlation, and Sensor. Byung-Ik Kim may be reached at kbi1983@kisa.or.kr

SEULGI LEE received the B.S. degree in Computer Science from the University of Chungnam, Korea, in 2013. Currently, He is a Researcher of Security Technology R&D Team 1 at Korea Internet & Security Agency. His research areas include cyber threat analysis, cyber attack related data correlation and machine learning algorithm. Seulgi Lee may be reached at sglee@kisa.or.kr

HYEISUN CHO received the B.S. degree in Computer Science from the University of Sejong, Korea, in 2013, the M.S. degree in Information Security from the University of SungKyunKwan, Korea, in 2017, respectively. Currently, She is a Researcher of Security Technology R&D Team 1 at Korea Internet & Security Agency. Her research areas include cyber threat analysis, cyber attack related data correlation and cyber threat reputation analysis. Hyeisun Cho may be reached at hscho@kisa.or.kr

JUN-HYUNG PARK received the B.S. degree in Computer Science from the University of Ajou, Korea, in 1999, the M.S. degree in Multimedia from the University of Chonnam, Korea, in 2002, the PhD degree in Information Security from University of Chonnam, Korea, in 2004, respectively. Currently, He is a Manager of Security Technology R&D Team 1 at Korea Internet & Security Agency. His research areas include cyber threat analysis, malware analysis and mobile billing fraud detection. Junhyung Park may be reached at junpark@kisa.or.kr