Efficient Analysis and Detection of Intelligent Security Threats in Cloud Environment

*Ji Su Park**

Department of Computer Science and Engineering, Jeonju University, Korea
jisupark@jj.ac.kr

Abstract

Recently, as cloud environments have spread, security technologies have strengthened preemptive defense technologies that predict and prevent various attacks. For example, security intelligence solutions, which have experienced considerable challenges, do not have sufficient reference data to operate, but many companies are introducing reference operational solutions. Among them, the analysis of the correlation between the log and parsing log heterogeneity requires considerable time and manpower. In this study, we create a rule for the security scenario based on parsing techniques to extract and parse the log of the log using only meaningful data and propose a method for purifying that can detect an intelligent security threat.

Keywords: SIEM, Security intelligence, Security log Correlation analysis, Cloud

1 Introduction

With recent changes in the cloud environment, security attacks have been carried out continuously over a long period with a designated goal. As this is not a general attack method, it is difficult to respond to it using the pattern-matching method of existing signature-based detection methods [1-5]. The signature-based detection method registers a malicious code (Malicious Code: Malware) pattern, determines the malicious code and Uniform Resource Locator (URL) patterns, and blocks it depending on whether it matches the registered pattern [2-4]. Detection techniques based on signatures have the problem of not being able to properly respond to newly occurring malicious codes and intelligent variant malware [1-5]. To address these issues, further research is required to detect and defend against new threats.

To defend against attacks with increasing technological capabilities and ensure security visibility, companies operate by applying Security Information Event Management (SIEM), which collects and analyzes all internal event logs, and using Security Information Management (SIM) that integrates Security Event Management (SEM) [6-7]. SIEM supports the optimization of the IT operating environment through event collection and analysis, and integrates and analyzes the logs generated from all IT infrastructures. The main function of the SIEM is to collect logs generated from network flow data, security solutions, servers, and applications, and manage the life cycle of logs by standardizing them through internal log analysis. The collected and processed logs provide an intuitive security status through threat detection and statistical analysis through correlation analysis [8].

Although companies use SIEM to increase the level of security defense, there is no proper guide for the correlation and correlation analysis of information generated from heterogeneous logs, and event analysis is difficult because of interconnection problems with non-standard zed logs.

In this paper, we propose a method for detecting security threats based on scenarios through a correlation analysis of security logs generated from heterogeneous sources.

2 Problem Formulation

2.1 Analysis of Recent Security Threats

Recently, the demand for security intelligence has increased in companies, regardless of the size of the organization. Compliance and obligations are strengthening, and data threats and breaches continue to pose challenges for technical security [9-11]. With the development of social media, protection of personal information has become an issue. Security threats are becoming more serious owing to the increasing number of users and the large amount of data and events generated by IT infrastructure [12-13]. Recently, security incidents, such as the leakage of large amounts of personal information and large-scale system failures, have occurred frequently because of external attacks. Attacks against security threats are difficult to detect even with current security technologies, and because they occur over a long period of time, even if the security infrastructure is well designed, they are not safe from attack threats [14]. Figure 1 shows various types of attacks [12-13].

To preemptively respond to changes in the threat environment of advanced and diverse attacks, the importance of log analysis, which is not properly utilized even though it is the most basic data, is being recognized, as well as the efficient real-time collection of massive events occurring in security solutions, networks, servers, etc. We are at a time when the need for analysis is becoming a greater issue than ever before. Therefore, a technique is required to integrate and analyze various events and logs generated in IT infrastructures and effectively respond to security threats.

*Corresponding Author: Ji Su Park; E-mail: jisupark@jj.ac.kr
DOI: https://doi.org/10.70003/160792642024072504013

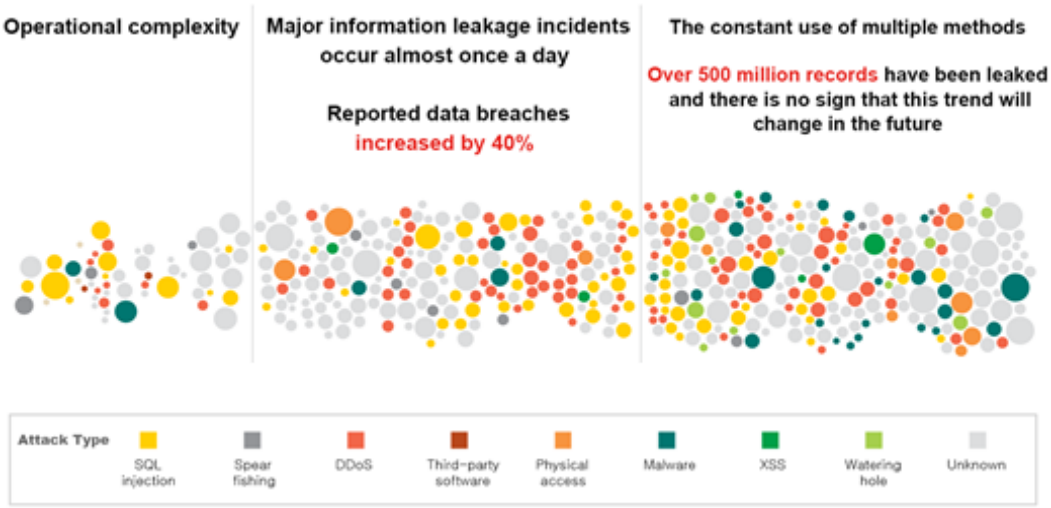

Figure 1. Continuous increase in types of attacks and security incidents

2.2 SIME

SIEM is a next-generation integrated log management solution that performs functions such as storing, analyzing, and deleting logs. It has the ability to detect attacks by profiling internal assets, integrating logs, and applying its own rules [6-7, 13, 15].

The SIEM architecture is shown in Figure 2. Heterogeneous logs are collected and processed in the event/flow processing system, including formalization and rule processing, and then transferred to the correlation analysis system, where the collected events are analyzed. Although each is an independent form, it is composed of a single architecture and the incoming logs are stored in the database of the log/flow collection and processing system [6-7, 13, 15].

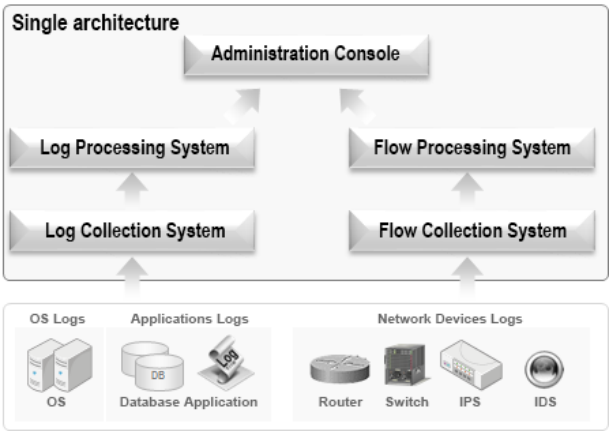

Figure 2. Security intelligence platform

3 Analysis and Detection Design Structure

The design structure for the security analysis and detection consisted of four steps. The first step is to collect the logs that are parsed and stored in the SIEM solution database. Step 2 analyzes the collected logs for correlation and defines the columns of meaningful data. In the third step, a security scenario was created based on the refined data. Step 4 involved creating information leakage detection scenarios and rules to detect each type. Figure 3 illustrates these four steps.

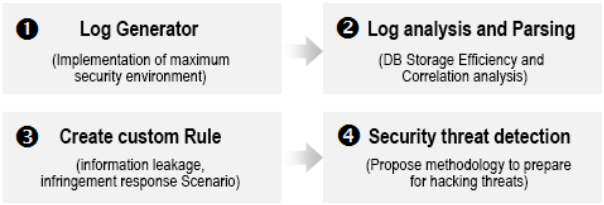

Figure 3. Flow chart of the experiment

In the process of collecting and storing raw logs, it was confirmed through the experiments in Section 4 that it is more efficient to extract and parse only meaningful parts, rather than standardize all log contents.

Through the parsing method shown in Figure 4, large amounts of log data flowing in real-time can be efficiently managed. This is the first step towards efficient storage and retrieval from a database for correlation analysis. Data must be segmented elaborately for correlation analysis and designed in a hybrid form that combines NoSQL databases, which form the basis of big data. It is a relational database that has advantages in terms of complexity, capacity limitations, and data correlation analysis, and is a core technology in data collection and analysis. In addition, in the case of domestic security solutions, there is no standardization of the log format when parsing logs; therefore, inefficient processing tasks increase from the log collection stage. To solve these problems, the logs must be standardized, core log definitions must be created from the logs of representative security solutions, and reference materials must be provided to parse meaningful logs. The log parsing method and rule creation data produced in this manner are used as reference materials for creating and applying security threat scenarios.

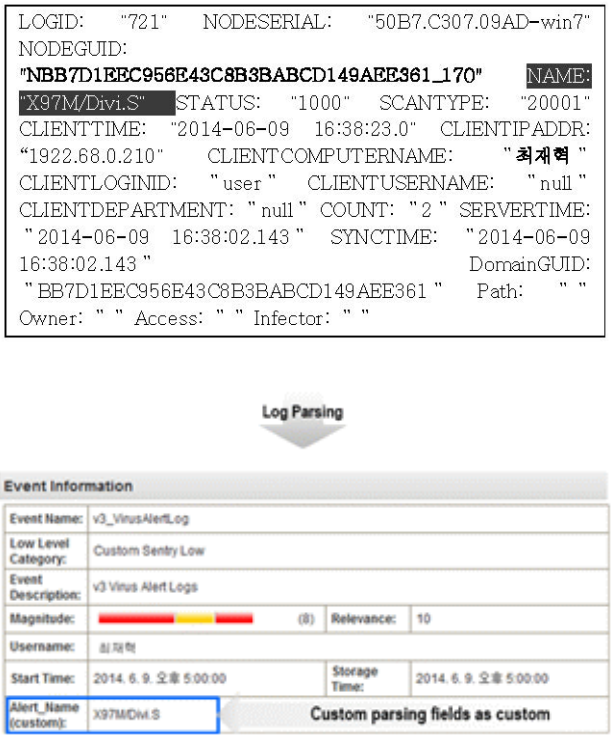

Figure 4. Example of raw log parsing

4 Experiment Result

4.1 Experimental Environment

The logs used in the experiment were linked to 12 logs of this type, including the network flow, as shown in Table 1. The experimental equipment was linked to Unified Threat Management (UTM): FortiGate, using the FG-200B model and a security printer (SINDOH D400) log. The hardware that generated the security log utilized a hypervisor (ESXi 5.0) for test efficiency. Multiple servers for one virtual machine (VM) were configured, and SIEM installation and the log generator were configured by building a Linux server and replacing it with equipment that generated logs through Tail2syslog and SFTP protocols.

Table 1. List of logs used in the experiment

Log source name	Description	Type
FortiGate	FortiNet FortiGate UTM	Syslog
Mail-i	SOMANSA Network DLP Mail-i	SFTP
MyGuard1	MyGuard FIle Use Monitering	SFTP
MyGuard2	MyGuard Processor Status Monitering	SFTP
Privacy-i	SOMANSA EndPoint Privacy-i	SFTP
SecurityPrinter	SINDOH Security Printer Security	JDBC
v3_NodeDetail	v3 Node Detail Clinet Last Update	SFTP
v3_VirusAlertLog	v3 Virus Alert Log	SFTP
Solaris10	Oracle SUN Solaris UNIX	Syslog
Windows2008	MicroSoft WINWOS2008	Syslog
Linux	RedHat Linux 5.8	Syslog
Network Flow	NetFlow Packet	Flow

To configure the basic system log by the type of operating system (OS), RedHat6.2, Windows2008, and Solaris10 x86 were installed, and the logs for each OS were linked using the Syslog protocol. The log configuration of the system is illustrated in Figure 5.

Name	Desc	Status	Protocol
FortiGate @ 192.168.0.1	FortiGate device	Success	Syslog
LinuxServer @ localhost	LinuxServer device	Success	Syslog
Maili @ 192.168.0.210	SOMANSA NDLP Mail-i Log	Success	Syslog
MyGuard1 @ 192.168.0.210	ITM MyGuard File Moniter	Success	LogFileProtocol
MyGuard2 @ 192.168.0.210	ITM MyGuard Process Kill Moniter	Success	LogFileProtocol
Privacy-i @ 192.168.0.210	SOMANSA EndoPoint DLP Privacy-i	Success	LogFileProtocol
SecurityPrinter @ 192.168.0.210	SecurityPrinter	Success	LogFileProtocol
Solaris10 @ 192.168.0.230	Oracle SUN solaris10 x86 OS Log	Success	Syslog
v3NodeDetail @ 192.168.0.210	Anlab V3 Node Detail	Success	LogFileProtocol
v3VirusAlertLog @ 192.168.0.210	Anlab V3 Virus Alert Log	Success	LogFileProtocol
Windows2008 @ 192.168.0.208	MicroSoft Windows AuthServer Device	Success	Syslog

Figure 5. Linked log sources screen

Network flow data were collected using a mirror method at the interface between the internal and external network sections and were linked for use when analyzing logs and correlations between different types of data.

The software used in this experiment was IBM QRadar 7.2.1 and ESXi5.0 evaluation versions. In addition to the actual equipment, the security solution log collects sample logs, retains the log structure of the original log, and contains random information that threatens security. The values were edited and linked in the form of text files using the SFTP and Tail2Syslog protocols. The experimental logical architecture is shown in Figure 6.

The hardware configuration consists of four VMs with an ESXi hypervisor in an IBM x3650 M3 box. It was created for each type of OS (Linux, Unix, and Windows) and linked using the Syslog protocol of the OS SYSTEM. Security sample logs collected on a Linux server were managed in one location and linked to an SIEM solution. The hardware configuration is shown in Figure 7.

The entire configuration, including the network, is configured to link the network flow data. At the top of the section entering the Internet, it is configured in the following order: “Router (local area network gateway) → Firewall (FortiGate UTM) → L2 switch → Server”. Network flow data is used to collect traffic data that flows both externally and internally. The port connected to the firewall was mirrored and interconnected at the L2 switch under the firewall, which is the point of contact with the Internet. The overall network configuration is shown in Figure 8.

4.2 Experimental Scenario

This security scenario is developed from the perspective of personal information leakage. This was created based on the scenario that detects information leaks through a correlation analysis of the logs related to personal information solutions. The experimental scenario used for personal-information leakage is shown in Table 2.

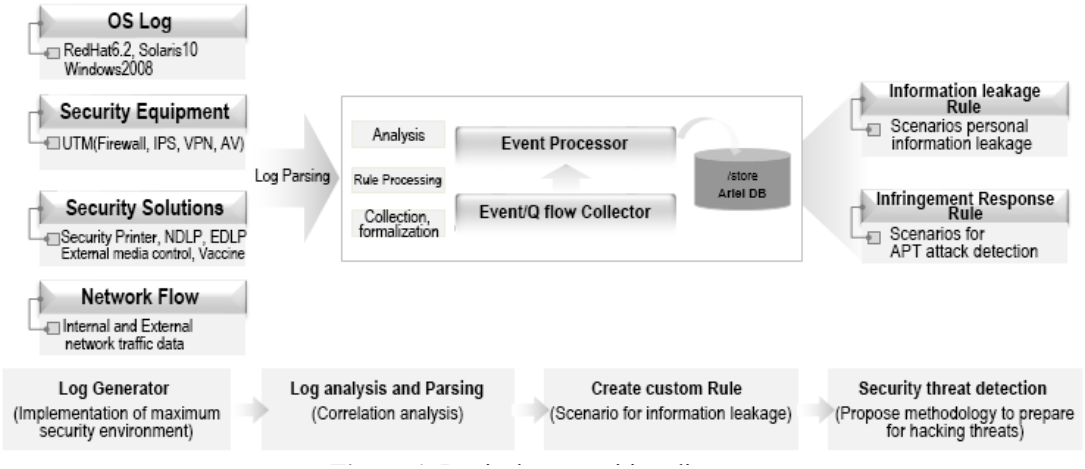

Figure 6. Logical composition diagram

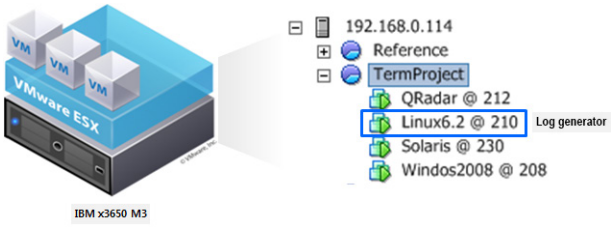

Figure 7. Hardware configuration

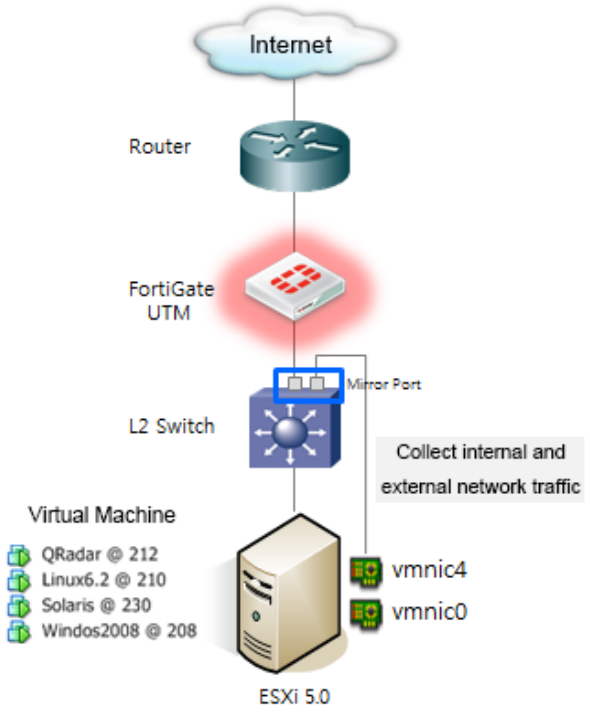

Figure 8. Physical configuration diagram

Table 2. Information leak detection scenario

Num	Description
1	When sensitive information is extracted from NDLP log
2	In EDLP, when more network traffic than usual occurs in a terminal without sensitive information
3	When a secure printer tries to write to external media in violation of its policy (scanning and printing privacy documents)

4.2.1 Extracting sensitive information from NDLP

In Network Data Loss Prevention (NDLP), risk levels can be assigned and managed for each item of personal information that may be sensitive, and a method for detecting this is proposed. The detection experiment is illustrated in Figure 9. In the case of sensitive information, the final risk (Ri or Value) is calculated by adding the results of multiplying the sensitive information by the weight for each company. For example, when "sensitive information" = "account number," it can be referenced or utilized in policies that can be managed by "risk," etc.

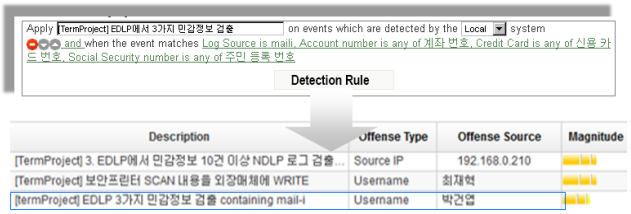

Figure 9. Detect leakage of sensitive personal information

4.2.2 Sensitive information Detection and Correlation Analysis in EDLP

When sensitive personal information is detected above a certain threshold in the Endpoint Data Loss Prevention (EDLP) log, a rule is created by linking the detection process and network traffic generation information through a correlation analysis of the network flow data. Content detected in this manner can be considered an act of personal information leakage. This algorithm combines rules by analyzing the correlations between log sources and network flow data. An explanation for this is shown in Figure 10.

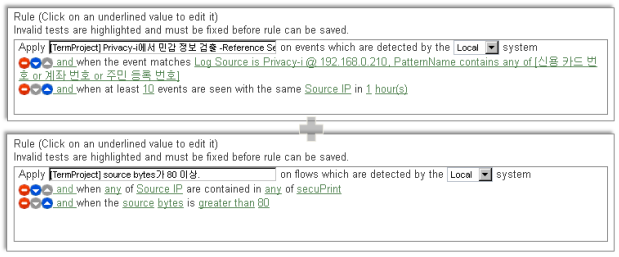

Figure 10. Example of creating log and flow data rules for correlation

4.2.3 Log Correlation for Secure Printers and Media Control

In this scenario, a user has violated the privacy policy of a secure printer by attempting to write to a USB. This policy scans documents related to personal information through security printer logs, and detects attempts to leak information through USB storage media.

As shown in Figure 11, a rule is first created using a security printer log. After filtering out users who violate the security printer policy, the log field value "Job" type detects the "SCAN" condition and registers it in the reference set. Additionally, the rule for using an external media control log is when a user who violates the security printer log based on filter conditions attempts to write to the USB. This is an example of creating a rule by analyzing the correlation between log sources.

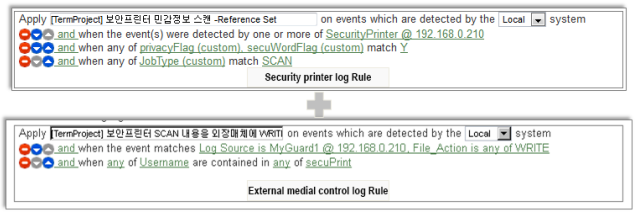

Figure 11. Example of correlation analysis using a combination of logs and logs and a combination of rules and rules

By combining these scenarios, correlation and correlation analyses between logs can be performed. In this scenario, if a correlation is made with the security printer log among the linked log sources, information can be leaked if the scanning or printing behavior of the printer or an attempt to write to an external medium is temporally analyzed from the source IP that violates the EDLP solution.

5 Conclusion

In this study, we analyze recent security threats and examine various security issues through related research. The types of information handled in security threats are diversifying, and the complexity of the security management environment is increasing owing to compliance requirements for large amounts of data. Extracting, purifying, and parsing meaningful parts from large amounts of data flowing in real-time using existing methods, as well as forming relationships and correlation analyses of specific field values of logs with other logs, require considerable manpower and time. However, as confirmed through experiments, if data on scenarios and rule creation from various perspectives can be accumulated and utilized, it can be used as an excellent reference material for security managers. In addition, when parsing logs, a significant amount of effort can be saved if sample logs and loss sources are defined and shared for each security solution.

We verified this through experiments, starting with a method of parsing raw logs and creating and detecting security scenarios. Security scenarios were created from the perspectives of information leakage and infringement responses. In detecting information leaks, it was confirmed that false positives can be reduced, and accurate detection is possible only when the correlation between security solution logs at the endpoint level is well combined. This is only possible through an understanding of internal policies and security solutions. Even if the forms and types of security solutions are different, the major categories and scenarios are not significantly different. In addition, because each company has different solutions for internal intrusion prevention systems and web firewall systems in terms of the types of infringements, the content of creating security scenario rules is an excellent reference material when establishing internal security policies, even if the log parsing method is different. The expectation is that this can be put to good use.

To operate security efficiently, an infrastructure must be established that can standardize the process of collecting and storing numerous heterogeneous logs and analyzing an integrated security system rather than analyzing individual security solutions. In addition, security knowledge that can publicly refer to cases and various references is required for a system that can predict and respond to advanced attacks through a correlation analysis between heterogeneous equipment and logs.

In this study, we focused on experiments from the perspective of information leakage; however, through future experiments and research on security attacks conducted over a long period of time, we plan to analyze attack forms for various types of infringements and use them as reference materials.

Acknowledgements

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support pro-gram (IITP-2023-2020-0-01789) and the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT. (RS-2023-00267476).

References

[1] P. Chen, L. Desmet, C. Huygens, A study on Advanced Persistent Threats, *15th IFIP TC 6/TC 11 International Conference*, *CMS 2014*, Aveiro, Portugal, September 25-26, 2014, p. 63-72. https://doi.org/10.1007/978-3-662-44885-4_5

[2] O. E. L. Castro, X. Deng, J. H. Park, Comprehensive Survey on AI-Based Technologies for Enhancing IoT Privacy and Security: Trends, Challenges, and Solutions, Article number, *Human-centric Computing and Information Sciences*, Vol. 13, Article No. 39, August, 2023. https://doi.org/10.22967/HCIS.2023.13.039

[3] K. S. Yu, S. H. Im, H. B. Kim, Technology Trends and Development Direction of the Integrated Log Management System, *The Korea Institute of Information Security and Cryptology*, Vol. 23, No. 6, pp. 90-99, December, 2013. https://www.dbpia.co.kr/pdf/pdfView.do?nodeId=NODE02334146&googleIPSandBox=false&mark=0&ipRange=false&b2cLoginYN=false&isPDFSizeAllowed=true&accessgl=Y&language=ko_KR&hasTopBanner=true

[4] S. W. Kim, J. Shin, H. B. Bang, Development of Behavior-based Malicious code etection·Blocking System, *Security Engineering Research Journal*, Vol. 9, No. 2, pp. 163-176, January, 2012. https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE10878580

[5] I. Ghafir, V. Prenosil, Advanced persistent threat attack detection: an overview, *International Journal of Advancements in Computer Networks and Its Security-IJCNS*, Vol. 4, No. 4, pp. 50-54, December, 2014.

[6] M. Sheeraz, M. A. Paracha, M. U. Haque, M. H. Durad, S. M. Mohsin, S. S. Band, A. Mosavi, Effective Security Monitoring Using Efficient SIEM Architecture, *Human-centric Computing and Information Sciences*, Vol. 13, Article No. 17, April, 2023. https://doi.org/10.22967/HCIS.2023.13.023

[7] I. Jeon, K. Han, D. Kim, J. Choi, Using the SIEM Software vulnerability detection model proposed, *Journal of The Korea Institute of Information Security and Cryptology*, Vol. 25, No. 4, pp. 961-974, August, 2015. https://doi.org/10.13089/JKIISC.2015.25.4.961

[8] A. Soroka, IBM X-Force 2013 Mid-Year Trend and Risk Report, *IBM*, October, 2013.

[9] W. S. Hwang, J. G. Shon, J. S. Park, Web Session Hijacking Defense Technique Using User Information, *Human-centric Computing and Information Sciences*, Vol. 12, Article No. 16, April, 2022. https://doi.org/10.22967/HCIS.2022.12.016

[10] H. H. Kim, J. Yoo, Analysis of Security Vulnerabilities for IoT Devices, *Journal of Information Processing Systems*, Vol. 18, No. 4, pp. 489-499, August, 2022. https://doi.org/10.3745/JIPS.03.0178

[11] S. Jeong, S. Kang, S. Kim, A Methodology for Integrating Security into the Automotive Development Process, *KIPS Transactions on Software and Data Engineering*, Vol. 9, No. 12, pp. 387-402, December, 2020. https://doi.org/10.3745/KTSDE.2020.9.12.387

[12] S. H. Jee, J. S. Park, J. G. Shon, Security in Network Virtualization: A Survey, *Journal of Information Processing Systems*, Vol. 17, No. 4, pp. 801-817, August, 2021. https://doi.org/10.3745/JIPS.04.0220

[13] A. Garofalo, C. Di Sarno, I. Matteucci, M. Vallini, V. Formicola, *Closing the loop of SIEM analysis to Secure Critical Infrastructures*, May, 2014. https://arxiv.org/abs/1405.2995

[14] A. R. Zope, A. Vidhate, N. Harale, Data Mining Approach in Security Information and Event Management, *International Journal of Future Computer and Communication*, Vol. 2, No. 2, pp. 80-84, April, 2013. https://doi.org/10.7763/IJFCC.2013.V2.126

[15] S.-W. Chen, C.-J. Tsai, C.-H. Liu, W. C.-C. Chu, C.-T. Tsai, Development of an Intelligent Defect Detection System for Gummy Candy under Edge Computing, *Journal of Internet Technology*, Vol. 23, No. 5, pp. 981-988, September, 2022. https://doi.org/10.53106/160792642022092305006

Biography

Ji Su Park received his B.S., M.S. degrees in Computer Science from Korea National Open University, Korea, in 2003, 2005, respectively and Ph.D. degrees in Computer Science Education from Korea University, 2013. He is currently a Professor in Dept. of Computer Science and Engineering from Jeonju University in Korea. His research interests are in Grid computing, Mobile cloud computing, Cloud computing, Distributed system, Computer education, and AIoT. He is employed as associate editor of Human-centric Computing and Information Sciences (HCIS) by Springer, The Journal of Information

Processing Systems (JIPS) by KIPS. He has also served as the chair, program committee chair or organizing committee chair at international conferences and workshops. He has received “best paper” awards from the CSA2018 conferences and “outstanding service” awards from CUTE2019 and BIC2020. Email: jisupark@jj.ac.kr