J Supercomput (2019) 75:4267–4294
https://doi.org/10.1007/s11227-018-2440-4

An effective security measures for nuclear power plant using big data analysis approach

Sangdo Lee[1,2] · Jun-Ho Huh[3]

Published online: 4 July 2018

Abstract Among the many hacking attempts carried out against information systems for the past few years, cyber-attacks that could lead to a national-level threat included attacks against nuclear facilities particularly nuclear power stations. Two of the typical examples are the Stuxnet attack against an Iranian nuclear facility and the cyber threat against Korea Hydro and Nuclear Power in December 2015. The former has proven that a direct cyber-attack can actually stop the nuclear power station, and the latter has shown that people can be terrorized with only a (cyber) threat. After these incidents, security measures for cyber-attacks against industrial control systems have been strengthened. The nuclear power stations also changed their passive concept of executing security measures by operating the plant with an isolated network to prepare for the cyber-attacks carried out by malicious codes. The difference between the two concepts is that the latter has been formulated based on the possibility that most of the control systems can be targets of cyber-attacks. Threats against control systems are gradually increasing nowadays, so the relevant industries are implementing some measures to identify/develop safe and reliable digital equipment and identify risks to establish effective cyber security plans. Thus, this paper proposes a security measure based on the classification of past attack incidents against control systems and the big data analysis technique that processes the data generated from individual security

✉ Jun-Ho Huh
72networks@pukyong.ac.kr; 72networks@cup.ac.kr

[1] Department of Computer Science and Engineering, Soongsil University, Seoul, Republic of Korea

[2] Security and ICT Department, Cyber Security Control Team, Korea Hydro and Nuclear Power (KHNP) Co. LTD, Gyeongju, Republic of Korea

[3] Department of Software, Catholic University of Pusan, Busan, Republic of Korea

equipment. The security of control systems is expected to be strengthened through such effective measure.

Keywords Nuclear power plant · Big data analysis · AI · Artificial intelligence · Security

1 Introduction

One of the recent forms of security attacks is advanced persistent threat (APT), which persistently targets a specific object over a long period of time. Since this kind of attack is not carried out based on the latest unknown vulnerability or with a routine attack method, it is not easy to respond to it using the existing pattern matching method based on the signatures. For this reason, the situation that the security controller at TOC should concentrate on and analyze is often an unfamiliar or unknown one [1–3].

The well-known attacks are usually blocked automatically by signature-based detection methods or vaccines. In signature-based detection methods, the unique patterns of malicious codes are registered in advance to find similar patterns in the network to block them or the URLs containing them. However, that such signature-based defense mechanism has a problem of not being able to respond to the newly generated malicious codes or intelligent mutant codes in real time. To improve such situation, the security centers are using a SIEM-based system that can analyze logs or introducing equipment that analyzes Big Data [4–8].

Nevertheless, the problem still lies in another area. The method of analysis can vary depending on the users of such system or equipment. The reduction of response time and improvement of analysis capability against new vulnerabilities have been the prominent questions for security control. The emerging method of addressing these problems is the utilization of AI and Big Data analysis [9–14].

AI allows intelligent judgment or autonomous action based on the analysis of available information. It has a self-learning function to make adequate judgment or adapt to the surrounding situation while developing into a technology that the people can use to undertake the tasks that they used to perform.

The area where AI is being actively used is the cancer diagnosis area. In this area, AI assists doctors in making an accurate diagnosis by studying the X-rays or cases involving similar symptoms. AI is used in the security field as well since both security and medical fields require human judgment as well as an information-learning process. Although a number of machine-learning mechanisms are being developed for this purpose, it is still at an early stage for the security area. Watson (IBM), which has already been commercialized for the analysis systems at the security control centers in the US, provides functions that analyze and diagnose security threats and possibility of malicious hacking attempts. Thus, in this study, Big Data analysis was performed for the capabilities of the current security AIs as well as their limitations and considerations.

This paper describes the SIEM analysis method performed for security control and presents the AI-based analysis method to be used as a major analysis tool in the future by recognizing its capabilities and limitations. This will change the paradigm of the current defense mechanism against cyber-attacks.

The method of responding to the hackings and malicious codes targeting nuclear power plants is also presented in this paper. By using the machine-learning mechanism, leaks of personal information and hackings through unsecure web services can be prevented. It is possible to respond to persistent and intelligent APT as well. Additionally, by using Big Data, the logs will be analyzed to determine the possibility of threats. All of these methods will further advance the existing security control measures and analysis methods.

2 Related works

2.1 Big data analysis method: security information and event management (SIEM)

The log content is a set of event records describing what had happened to the specific equipment or system. The common mechanism of recording the contents can be logically defined/separated as log transmission, logging grammar, and components of login setting.

In order to categorize security threats and develop the relevant scenarios, the analysis of raw logs is essential. It is necessary to categorize what each field means through the log definition of the relevant solution to understand the type of raw log. In addition, to create a security scenario, the characteristics and meaning of the log should be extracted including the correlation with other security solutions. As the data volume and variety increase, not only the number of simple access logs but also the network flow data volume and number of entire full-packet logs and atypical logs have increased as well. These are commonly referred to as Big Data, and some equipment and analysis methods have been developed; still, it does not mean that the number of security analysis methods has increased as well. Rather, it simply means that the volume of data that can be analyzed has increased and become segmentalized [15–22].

As a next-generation integrated log management solution developed from the current integrated log management system, SIEM (Security Information and Event Management) performs functions such as storage, analysis, and deletion of logs. SIEM also profiles internal properties and integrates logs to detect attacks autonomously by applying self-established rules. SIEM is adopted by many companies to analyze events and improve their security defense systems. Nevertheless, in addition to the problem of parsing of non-standardized logs, there are no guidelines for analyzing the correlations between the information generated from the heterogeneous logs; hence, the many problems in designing a detection policy for the security threats.

This study presented the guideline for the scenario wherein the necessary information is extracted from a huge database using SIEM and security threats are detected to respond intuitively to each threat.

2.2 Artificial intelligence

The core principle of Artificial Intelligence (AI) is a machine-learning process such as Alpha-Go or Deep Learning [15] wherein the machine itself collects and abstracts information for learning without any additional guidelines for judgment by humans.

AI is a technology that implements the intellectual capacity of a human such as thinking or learning through computers. Conceptually, it can be divided into Strong AI and Weak AI. The former refers to AI with an ego that allows liberal thinking like human beings and performs a variety of tasks, so it is also called Artificial General Intelligence (AGI). The latter refers to AI that does not have any ego, and it is mainly used to complement human limitations and enhance productivity after being developed into a form specializing in a certain area. One of the typical AIs of this category is Watson, which is often used in Go or quiz games as well as in medical fields. Most of the AIs developed so far are Weak AIs; AI with a strong ego has yet to come. Thus, AI is also a technology that implements learning, inference, perception, and language understanding abilities through computer programs [23–28].

2.3 Limitation of AI

Watson transmits five kinds of basic information to the Watson Cloud after encrypting them. Although the information necessary for more precise security data analysis includes port number, user agent, AV signature, E-mail address, and file name, only the minimum information will be transmitted to the cloud. This is to enable minimum information control and avoid possible personal information leak or other security problems [29–34].

As for the security problems, there are many cases of intrusions by external unauthorized hackers, but threats from insiders are increasing nowadays. Information concerning the logs generated from an inside user PC or an Agent/ERP server is often too small to use external clouds, and company executives or information analysts are not receptive to sending out an important log history to the outside systems as they are concerned about the possibility of information leak. Much time and cost are needed to set the security rules for SIEM accurately when attempting to analyze the insider user policy or detect unusual behaviors based on machine-learning; in addition, reducing the rate of misdetections through its learning process is expected to take much time.

3 Proposed method

This paper proposes a detailed analysis technique for network hackings and intrusions against control facilities.

3.1 Cyber security framework against cyber-attacks targeting control facilities

Among many hacking attempts against information systems, those that would have resulted in a national-level disaster were the cyber-attacks against nuclear facilities and power plants, such as the Stuxnet attack against an Iranian nuclear facility and the cyber-attack launched on Korea Hydro and Nuclear Power Co., Ltd. in the Republic of Korea (ROK). The former showed that a single direct cyber-attack alone could stop the nuclear power plant, whereas the latter demonstrated that the people can be terror-stricken with just a threat of cyber-hacking. After these incidents, the security of industrial control facilities had been tightened. For instance, the management of nuclear power plants changed their existing concept of operating/managing control facilities. They have realized that security cannot be guaranteed just by separating their operating network from external Internet networks, and it is possible to attack a control facility with malicious codes or cyber-attacks. In this regard, control systems in the US have already set up a strong security framework for individual control systems [25, 35–41].

In this paper, the instances of cyber-attacks against control systems are studied, followed by the analysis of a response plan for the individual attack scenarios to strengthen the security level when designing a domestic security framework by considering the major security elements involved.

There was a serious threat (hacking attempt) to Korean Hydro and Nuclear Power Co., Ltd. in December 2014. Korean people who remember the Fukushima nuclear accident were terror-stricken by this incident as the same could occur in their nuclear facilities. Such psychological issue was the result of careless security management (information leak) by one of the subcontractors.

The report from the Korean joint investigation task force revealed that it was an act committed by a certain hacker organization to cause social unrest by making people doubt the safety of their nuclear facilities. Although most of the information was confirmed to have been leaked through E-mails of the subcontracting company, the integrity of the security system of nuclear facilities in ROK had not been ensured from the early stages. What the nuclear power plant control system boasted of was that its control system had been completely separated from external Internet networks, thereby establishing a closed network. However, that such arrangement is no longer safe as shown in the case of the Stuxnet attack against the Iranian nuclear facility control system. Diagnosis for the individual traffic, utility, and nuclear control systems must be carried out urgently, and systematic security standards must be established along with reliable security frameworks. Nevertheless, it has been easy to deal with security vulnerabilities as to what kind of impact the security measures will have on the control system. In other words, the security system itself could cause safety accidents. Meanwhile, the US has enacted the Federal Information Security Management Act (FISMA) in 2002 for the protection of national service infrastructure (e.g., energy, traffic, power, utility, communication, nuclear, and other core systems) against terror attacks. Based on this, the Obama administration has established the Comprehensive National Cybersecurity Initiative (CNCI) and developed it as the core element of the US's cyber security strategy.

Table 1 Status of nuclear power stations in ROK

Power plant	Power station (14)	Remarks (28 units)
Gori power plant, ROK	Gori 1 and 2 Sin-Gori 1 and 2	4 Power stations (8 units)
Hanbit power plant, ROK	Hanbit 1, 2 and 3	3 Power stations (6 units)
Wolseong power plant, ROK	Wolseong 1, 2 and 3	3 Power stations (6 units)
Hanwul power plant, ROK	Hanwul 1, 2 and 3 Sin-Hanwul 1	4 Power stations (8 units)

3.2 Status of Korean nuclear power plants and their network in OSI layers

The individual operation network of ROK's nuclear power plants is commonly divided into three zones: Internet Network, Internal Network, and Control and Monitoring Networks. Although the first two networks have been physically separated, the data can be transmitted through the inter-network data transmission system. Both control and monitoring networks are separated from the Internet and internal networks. However, that they are hierarchically (OSI) interconnected to receive operational information from the equipment that monitors its control facility; in this case, the information flow is unilateral. In other words, only simple operational information and measurements are transmitted from the high-level security control network to the internal network and finally to the collection server. The control facility is divided into server side, RTU, and inter-communication network, and the TCP/IP protocol-based system in the OSI layer is used for the measuring equipment. For this reason, the existing cyber threats can still be carried out, so the security vulnerabilities should be analyzed to devise a stronger security measure. Table 1 presents the status of nuclear power stations in ROK. Figure 1 shows the OSI-hierarchical structure of the nuclear power plant network.

The operation facility consists of digital equipment such as computers, terminals, and data communication network, and the network has a closed form. Measuring facilities transmit data to the equipment connected to the operating network unilaterally, and only the control or measurement signals flow from the closed network to this operation network. Although these networks are separated from the external Internet, it is not totally impossible to carry out some sort of cyber-attacks against them, as observed in the Stuxnet attack against the Iranian nuclear facility or the malicious code-based attack against a Ukrainian power station.

The instrument and control system and the communication network, both of which were the initial nuclear power plant management system, did not include a comprehensive security enhancement measures for the possible cyber-attacks in its early designing phases and most of the system was analog.

Later, the system was converted into a digital system, and commercial OS started to be used. However, that such measure gave rise to vulnerabilities against cyber-attacks. Moreover, the Windows OS, which has higher risk of infection from a virus or a malicious code, required constant updates.

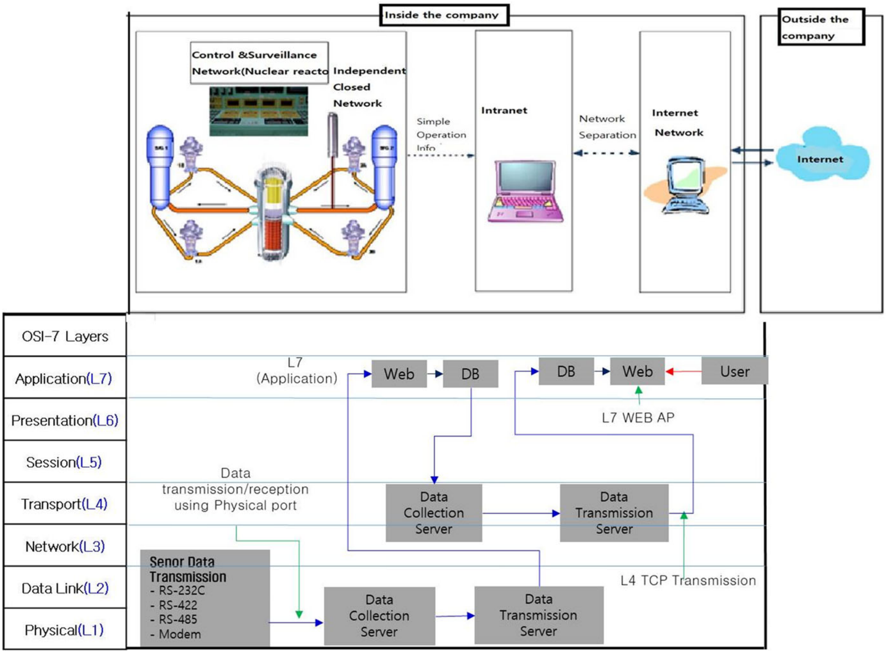


Fig. 1 OSI-hierarchical structure of the nuclear power plant network

Additionally, some measures against the virus infection of the control system software were demanded; because the system adopted a closed network, however, they had not been dealt with. Patching and maintenance are conducted only during the maintenance period, and only for the control system being operated currently. In other words, no maintenance works that would have even the smallest influence over the control system and other equipment works will be allowed except during the prescheduled/urgent maintenance period.

4 Big data analysis for security control: SIEM

4.1 Development phase of threat detection scenario

During the development phase of the detection scenario, the status of internal information leaks is analyzed followed by the determination of a specific feature in the hacking codes by performing static and dynamic analyses of security events to create a signature based on the Snort or Yara rule. The signature is applied to the security equipment to check a misdetection or a correct detection. If the created scenario is unable to make a correct detection, additional rule(s) will be added to repeat the detection process. The detection can be performed by establishing a basis for analyzing the logs in the information system first, and then applying the threat analysis rules in addition to additional rules generated through the verification scenario. When categorizing the events, items that caused misdetections should be excluded. A scenario

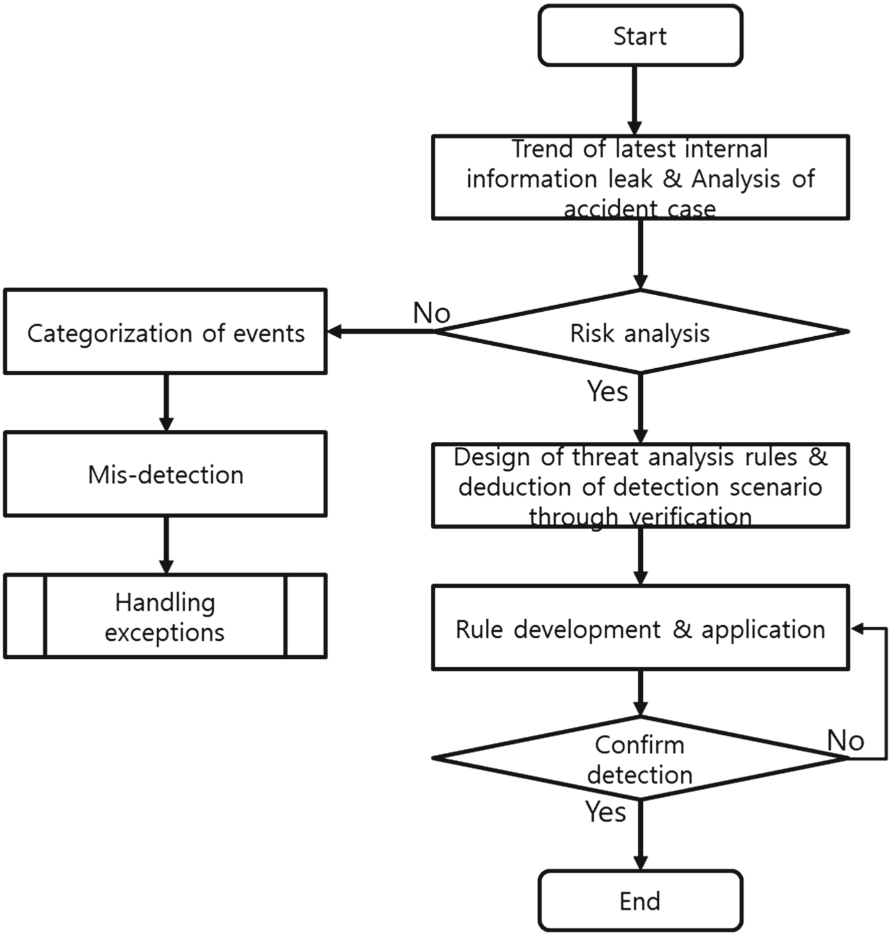

Start
Trend of latest internal information leak & Analysis of accident case
Risk analysis
No
Categorization of events
Mis-detection
Handling exceptions
Yes
Design of threat analysis rules & deduction of detection scenario through verification
Rule development & application
Confirm detection
No
Yes
End

Fig. 2 Flowchart of security scenario development

is then designed following the category defined in the threat scenario, log extracted based on the user definition, and specific keyword search. After completing the analyses of the status of information system assets as well as operation status, a number of hypotheses and theories are established to create a security scenario wherein security breach will be detected based on the rules initially given. Figure 2 shows the rule generation and detection flow in the security scenario.

4.2 Analysis process of SIEM logs

Approximately 400 million logs are commonly generated every 10,000 EPS (Events Per Second) daily at a ratio of 65% (business hours) to 35% (no-business hours). It will be difficult to create detection scenario and rules without analyzing the raw logs of the relevant log source and exceptions. Using examples, the method of extracting

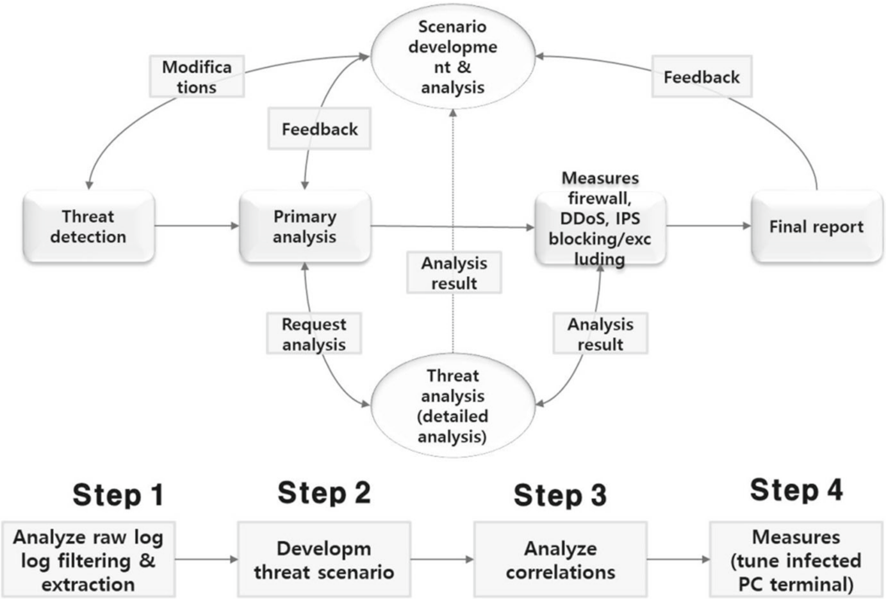

Scenario developme nt & analysis
Modifica tions
Feedback
Feedback
Threat detection
Primary analysis
Measures firewall, DDoS, IPS blocking/exc luding
Final report
Analysis result
Request analysis
Analysis result
Threat analysis (detailed analysis)
Step 1
Analyze raw log log filtering & extraction
Step 2
Developm threat scenario
Step 3
Analyze correlations
Step 4
Measures (tune infected PC terminal)

Fig. 3 Security threat detection process

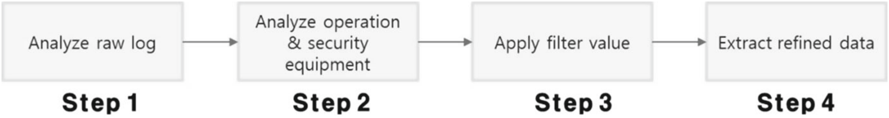
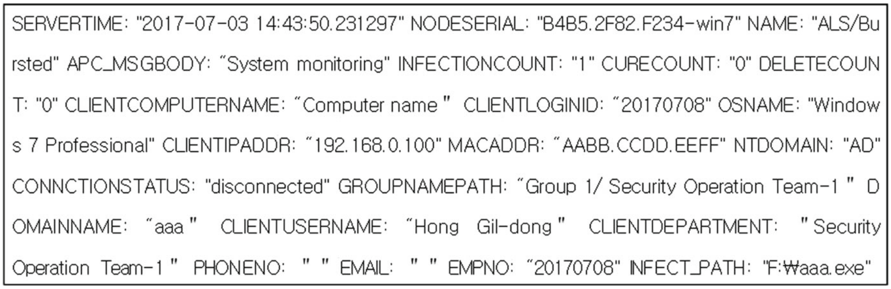


Fig. 4 Process of extracting meaningful data from the logs

meaningful data from the log source and the process of analyzing the SIEM logs will be described in this phase. The analysis of raw logs as a basic data is a primary phase of detecting threats. If important data cannot be filtered out in this phase, the entire process from scenario development to threat detection will be affected considerably. The threat detection process is defined below.

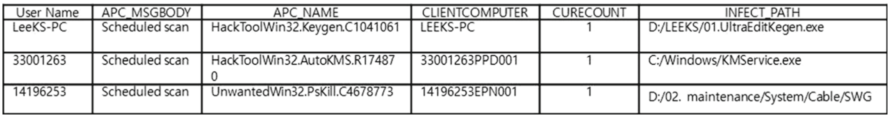
As described in Figs. 3 and 4, the first process of threat detection involves analyzing the raw logs generated from security equipment. This is also a process of analyzing the format form of a log before parsing it. This process seeks to confirm whether the data such as user, source IP, destination IP, log source, and time have been collected from the raw logs in real time or in the form of their deployment.

The second process involves the analysis of equipment tasks and categorization of each field and code value generated from the raw logs. Figure 5 is an example of the raw data generated by the endpoint vaccine. Each field is described as follows: SERVERTIME (time of detection by the vaccine server), NODESERIAL (registered client serial), NAME (designated name of the detected virus), INFECTIONCOUNT (number of files infected), CURECOUNT (number of files cured), DELETECOUNT (number of infected files deleted), CLIENTCOMPUTERNAME (name of the client computer), CLIENTLOGINID (login ID), OSNAME (name of



```
SERVERTIME: "2017-07-03 14:43:50.231297" NODESERIAL: "B4B5.2F82.F234-win7" NAME: "ALS/Bu
rsted" APC_MSGBODY: "System monitoring" INFECTIONCOUNT: "1" CURECOUNT: "0" DELETECOUN
T: "0" CLIENTCOMPUTERNAME: "Computer name " CLIENTLOGINID: "20170708" OSNAME: "Window
s 7 Professional" CLIENTIPADDR: "192.168.0.100" MACADDR: "AABB.CCDD.EEFF" NTDOMAIN: "AD"
CONNCTIONSTATUS: "disconnected" GROUPNAMEPATH: "Group 1/ Security Operation Team-1 " D
OMAINNAME: "aaa " CLIENTUSERNAME: "Hong Gil-dong " CLIENTDEPARTMENT: " Security
Operation Team-1 " PHONENO: " " EMAIL: " " EMPNO: "20170708" INFECT_PATH: "F:₩aaa.exe"
```

Fig. 5 Example of raw logs of a vaccine



User Name	APC_MSGBODY	APC_NAME	CLIENTCOMPUTER	CURECOUNT	INFECT_PATH
LeeKS-PC	Scheduled scan	HackToolWin32.Keygen.C1041061	LEEKS-PC	1	D:/LEEKS/01.UltraEditKegen.exe
33001263	Scheduled scan	HackToolWin32.AutoKMS.R174870	33001263PPD001	1	C:/Windows/KMService.exe
14196253	Scheduled scan	UnwantedWin32.PsKill.C4678773	14196253EPN001	1	D:/02. maintenance/System/Cable/SWG

Fig. 6 Extraction of refined data

the DS used), CLIENTIPADDR (IP address of the PC), MACADDR (MAC address of the PC), NTDOMAIN (active directory domain name), CONNECTIONSTATUS (connectivity), GROUPNAMEPATH (group name), DOMAINNAME (domain name), CLIENTUSERNAME (name of the user), CLIENTDEPARTMENT (name of the user group), PHONENO (telephone number of the user), EMAIL (user mail information), EMPNO (employment number of the user), RESTCOUNT(), and INFECT_PATH (means of infection). As such, this process categorizes the characteristics of the security equipment, field values, and code values.

A filter is applied to extract the necessary data in the third process. For statistical analysis, the analysis of the characteristics and task of the security equipment is required in the process of refining meaningful data. In Fig. 5, many data will be found when searching for the data prior to analyzing the meaning of each field value in the raw logs. The refined meaningful data can be extracted by categorizing tasks and field/code values and subsequently applying the filter value(s) depending on the threat scenario.

For example, when extracting some meaningful data according to the operating environment of a certain vaccine, filtering can be performed on the uncured PC (files infected but not deleted) or by excluding the viruses included in the TEMP directory. If any kinds of messengers are used within the company, filtering will be carried out without a messenger monitoring task. By applying these filtering processes, the refined data will finally be extracted.

Through the third process, the refined data can be extracted in the fourth process. Considering the first process (analysis of meanings of raw logs), only the meaningful data desired by the user will be extracted (Fig. 6), and these refined data can be used to analyze the threat scenarios and correlations that will ultimately allow the detection of possible final threats.

4.3 Analysis of control process in security control using AI

The major form of recent security attacks is the APT, which attacks the target persistently for a long period of time. Since this method does not employ regular attack patterns that are previously known or are based on the vulnerabilities, the existing signature-based pattern matching methods are not effective. For this reason, one of the main tasks of security controllers is to analyze the attack patterns that are new to them. The other known attacks can be blocked by using vaccines or other common measures. In the signature-based detection method, the pattern unique to each malicious code is registered in advance to block the URLs or malicious codes that follow those patterns in the network. Such technique cannot deal with new malignant codes or intelligent mutant codes in real time. To improve such situation, individual security control centers are adopting the SIEM system or introducing equipment that can analyze the Big Data log. Nevertheless, the effectiveness of these measures will vary depending on the users. Reducing the response time and enhancing the analysis level for the new vulnerabilities have been the issues in security control. Thus, adopting AI can be an effective new solution.

AI allows autonomous information analysis and intelligent judgment to make an adequate response or to adapt to the current situation, performing the tasks that people used to do.

The area where AIs are being used most actively is the medical diagnosis field. They are assisting doctors in their diagnosis by studying previous medical cases and relevant patient X-rays. AIs are also being used in the security area. These two areas have many similarities since both of them require human judgment and much learning. AIs perform them through the machine-learning function, which is developing rapidly. Some of the AI-based medical diagnosis systems have been commercialized already and are actually used in the field, whereas commercialization of similar products has just begun. The leading product for the security control analysis system is Watson, which analyzes security threats or determines the possibility of malicious attacks.

This study investigated the process of Big Data analysis for security control, with the process and results of the Big Data analysis in Chapter 3 compared with the analysis results obtained from an AI system to identify its performance level and limitations. With the comparison result, a technology that can change the paradigm of defense against new cyber-attacks will be proposed together with the future direction of AI use.

5 Leak detection via machine learning

Until now, most of the personal information leaks through the web servers have occurred by attacking their vulnerabilities such as SQL injection, for example. On the other hand, the number of leaks by internal or authorized users is increasing nowadays. Their method involves collecting the personal information gradually each day by approaching it in a normal way, so it was difficult to detect their excessive access to important information. The means for the security controller to access or monitor all the queries or authorized SQL queries were limited as well. Thus, research on the technique that can monitor them in real time and the method of blocking personal information leak through the White List-based approach policy was conducted in this study. Illegal access to the Internet web servers or inappropriate approach to the personal information database is expected to be monitored effectively and efficiently while securing the stability of the system.

Previously, the main goal of hacking attempts was to parade the individual hacker's own ability or to voice the political opinion of a certain group. However, that the latest cyber-attacks are being launched with monetary expectation. The ATP attacks clearly target the database of an individual or a firm for money or to trade personal/company information through threats. Ransomware is one of these types of attacks targeting the general population or companies. A few years back in ROK, the customer DB of a large company was leaked by an insider, leading to a lawsuit. What was unique about this incident was that access had been made through the escalation of the insider's privilege. Although he stole information continuously and gradually to trade them with the competing company, the company security team did not have any means of detecting the leaks as the traffic remained stable and the information had been collected over a few years through the privilege escalation of a normal user.

In this chapter, the access behaviors of all the users who had accessed the personal information processing system were analyzed by using a pattern classification technology (machine-learning) to deduce a number of normal patterns in order to detect abnormal accesses. If the current access behavior does not conform to the normal behavior of the user's profile, the system will identify it as a malignant user. For the implementation of this technique, the analysis method based on the threshold values was used in addition to the definitions of patterns established in advance.

5.1 Extraction of feature points and profile learning

The feature point extraction and profile formation process is composed of two steps. Specifically, a profile is created by using the information collected for a given period of time. If this profile is used for user authentication continuously, however, it will not be able to reflect the gradually changing behaviors after that period. Thus, to reflect the change, a fixed number will be used for (Eq. 1).

$$p_t = \sum_{i=i}^{T-1} w_{t-i} \cdot x_{t-i}, \quad \text{where} \begin{cases} \sum\limits_{i=1}^{T-1} w_{t-i} = 1 \\ T = 1 \sim N \end{cases} \tag{1}$$

- $t - i$: Entire period from today to ith.
- w_{t-i}: Graded weight for entire ith data. $w_{t-i} \leq w\_(t - i + 1)$, $\forall i$
- x_{t-i}: Profile of entire ith data
- T $(= N)$: Entire period where the graded weight applies
- p_t: Second-step profile which indicates the result obtained after the graded time weight has been applied for the developed product, $T{=}4$.

The profile is created in the first step. Equation 2 uses two methods: one that forms the user-behavior information by using the behavior pattern when attempting to establish access, and the other is to form an access pattern based on the behaviors demonstrated during a certain period of time after establishing access.

5.1.1 Classification of behavior patterns during access

The calculation was performed by using the Markov Chain, which models the shifting patterns of web pages in a time-stochastic manner.

Markov Chain Model (Probability Model): (2)

$$\begin{aligned} P &= \left(X_{\text{training}}\right) \\ &= P\left(x_1, x_2, \ldots, x_t\right) \\ &= P\left(x_1\right) P\left(x_2|x_1\right) P\left(x_3|x_2\right) \cdots P\left(x_t|x_{t-1}\right) \\ &= P\left(x_1\right) \prod_{i=1}^{T-1} P\left(x_{i+1}|x_i\right) \end{aligned} \tag{2}$$

- Time t is a discrete value: $0 \leq T \leq t$
- State Space X : $X = (1, 2, \ldots.., N)$
- When the state of Time t is x_t, the state transition vector is $(x_0, \ldots, x_t)$
- Probability that the state will become x_{t+1} at t + 1: $P(x_{t+1}|\, x\_0,\ x\_1, \ldots, x\_t) = P\,(x_{t+1}|\, x_t)$
- State transition probability: $P(x_{(t+1)}|x_t),\ \sum_{t=0}^{N} P(x_{t+1}|x_t) = 1$
- Initial state probability: $\pi_u = P(x_0),\ \sum_{u=0}^{N} \pi_u = 1$

5.1.2 Decision surface

The decision boundary value (threshold) of the pattern used when establishing access should be equivalent to or larger than the abnormal user's maximum aggregated value and larger or smaller than the normal user's minimum aggregated value. The aggregated value of membership analysis can be expressed as Eqs. (3) and (4).

$$\prod_{i=1,j=1}^{I\text{max}} m_{ij} \tag{3}$$

-

$$\alpha^i_{1^{st}} \prod_{i=1}^{I_{\text{max}-1}} \left(\beta^i_{\text{max}}\right) \leq \theta_{1^{\text{st}}} \text{ (threshold)} < \prod_{i=1}^{I\text{max}} \left(\beta^i_{\text{min}}\right) \tag{4}$$

Through the analysis of SQL between Java application and jdbc, only the White SQL is authorized, and others are blocked (Fig. 7).

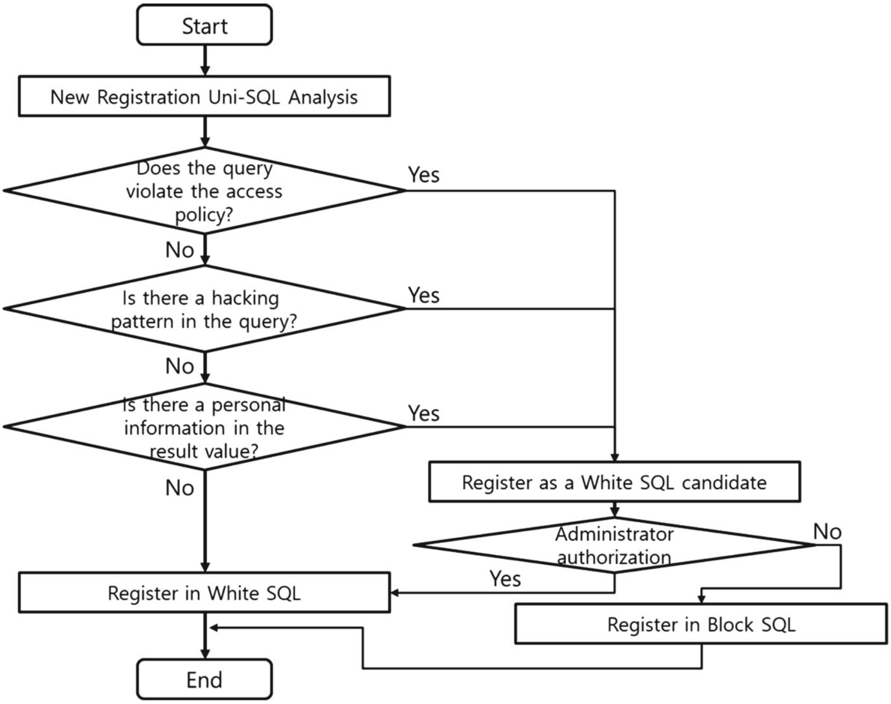

Start
New Registration Uni-SQL Analysis
Does the query violate the access policy?
Yes
No
Is there a hacking pattern in the query?
Yes
No
Is there a personal information in the result value?
Yes
No
Register as a White SQL candidate
Administrator authorization
No
Yes
Register in White SQL
Register in Block SQL
End

Fig. 7 Flowchart of personal information search

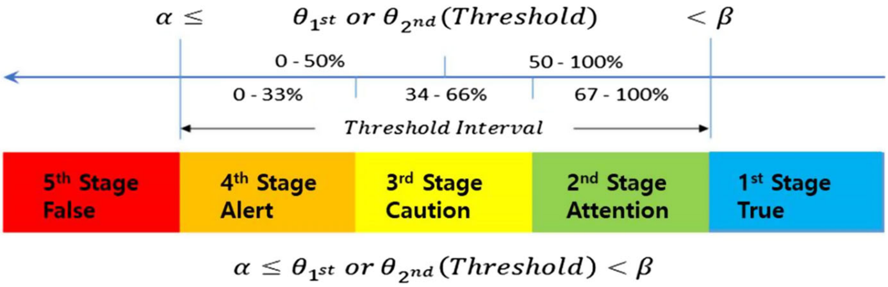


Fig. 8 Warning levels of individual thresholds

5.2 Performance evaluation

As one of the criteria for detecting normal users, a threshold value exists in each model; the warning level depending on each threshold is determined as shown in Fig. 8 (5 levels).

Based on the user profiles, a score is assigned for each level after analyzing the user in real time. Table 2 shows the standards of comprehensive distinction.

The user-behavior pattern can be identified even with only small data via parsing and filtering of log data. The analysis showed higher results (Table 3, Fig. 9).

Table 2 Standards of comprehensive distinction

Level	Score	Level	Criteria
1	Normal (true)	1	When larger than the least value
2	Attention	2	Staying within the range of 67–100%
3	Caution	3	Staying within the range of 34–66%
4	Alert	4	Staying within the range of 0–33%
5	Serious (false)	5	Staying under the threshold range

Table 3 Measured value of accuracy/Reduction rate

	Target value (%)	Measured value (%)
Accuracy	95	98.97
Reduction rate	66.60	77.57

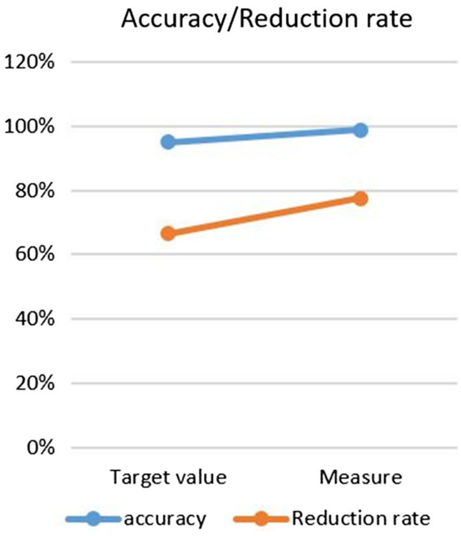


Fig. 9 Accuracy/reduction rate

In the past, most of the malignant behaviors were detected by using the signature or rule-based detection method, but it was not possible to detect the abnormal behaviors of a normal user as these methods could be used. Since it is necessary to extract the behavior-based approaches that monitor personal/internal information or accesses by the abnormal users, misdetection must be reduced. In addition, those SQL queries that are not authorized as White SQL need to be blocked; other authorized queries should also be monitored for their specifics by applying a proper threshold.

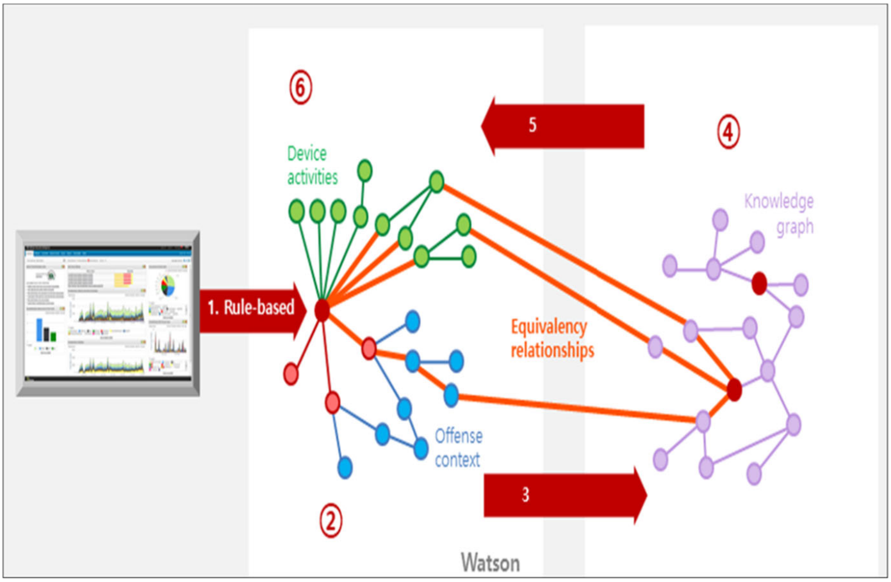


Fig. 10 Operating order of IBM's Watson

5.3 Security analysis

5.3.1 AI analysis method and flow

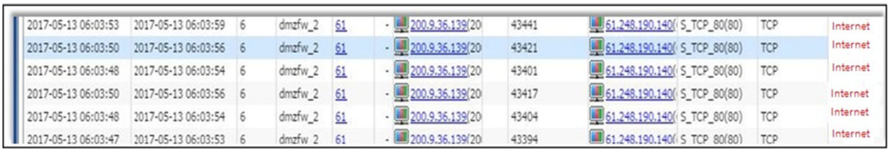
In this chapter, the security data of a nuclear power plant in ROK was analyzed (big data analysis) through experiments by using IBM's Watson. As for the order of operation, among the information detected in SIEM using the rule-based method, only the basic information such as S-IP, D-IP, and hashes are sent to Watson first. Second, the local context is acquired, and a strategy is formed. Third, a relationship is formed between the content analyzed through the reputation lookup and the knowledge DB. Fourth, the threat is studied to develop specialized knowledge. Fifth, all infection routes and warning signs are indicated with a map. Lastly, the intelligence collected through investigation is applied, and eligibility is given to the incident. Figure 10 shows the operating order of IBM's Watson.

Three advantages were expected from introducing Watson into the Korean nuclear power plant. First is speed. Matters investigated through reputation lookup (X-Force, VirusTotal, etc.), security article, and threat feed can be searched faster than the security analysts.

Second is consistency. While information analysts use different information sources depending on their preference, Watson offers consistent information among many sources. Last is insight. Watson can show correlations in a large volume of data in visual form speedily.

Table 4 Network transmission values for Watson to perform analysis with the cloud

Subject	Content	Transmission
Source IP	External source IP related to the offense following the criteria defined by the network layer structure of SIEM	Yes
Destination IP	External destination IP related to the offense following the criteria defined by the network layer structure of SIEM	Yes
File hash	Hash value of a suspicious file	Yes
URL	URL of malicious code or infection route	Yes
Domain	Area or C&C server infected with malicious code	Yes



```
2017-05-13 06:03:53  2017-05-13 06:03:59  6  dmzfw_2  61  - 200.9.36.139(20  43441  61.248.190.140( S_TCP_80(80)  TCP  Internet
2017-05-13 06:03:50  2017-05-13 06:03:56  6  dmzfw_2  61  - 200.9.36.139(20  43421  61.248.190.140( S_TCP_80(80)  TCP  Internet
2017-05-13 06:03:48  2017-05-13 06:03:54  6  dmzfw_2  61  - 200.9.36.139(20  43401  61.248.190.140( S_TCP_80(80)  TCP  Internet
2017-05-13 06:03:50  2017-05-13 06:03:56  6  dmzfw_2  61  - 200.9.36.139(20  43417  61.248.190.140( S_TCP_80(80)  TCP  Internet
2017-05-13 06:03:48  2017-05-13 06:03:54  6  dmzfw_2  61  - 200.9.36.139(20  43404  61.248.190.140( S_TCP_80(80)  TCP  Internet
2017-05-13 06:03:47  2017-05-13 06:03:53  6  dmzfw_2  61  - 200.9.36.139(20  43394  61.248.190.140( S_TCP_80(80)  TCP  Internet
```

Fig. 11 Raw log of firewall equipment

5.3.2 Data transmission by Watson AI

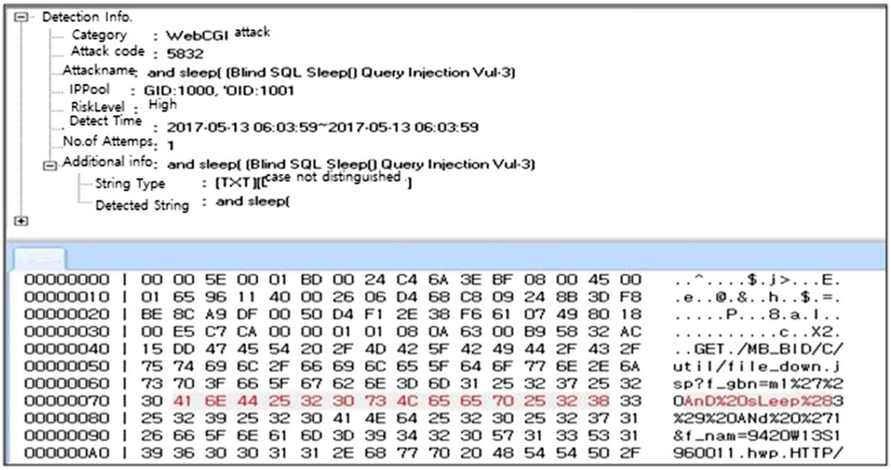
Watson is installed in applet form in the SIEM (QRadar) of IBM. According to its Offense Rule, the risk level is indicated with its scoring rule, and then the risk itself is analyzed by linking with the external Watson Cloud. For the analysis, the basic information in Table 3 is encrypted and sent to the cloud. Watson Cloud accumulates the threat information acquired from the reputation lookups for the malicious DBs and represents each site's newly requested basic information with nodes to show their correlations (i.e., correlations of as many as 50+ units).

Table 4 shows the network transmission values for Watson to perform analysis with the cloud.

In this system, a cognitive study is performed for the underlying causes and additional elements in addition to the speedy processing of matters that are easily missed by humans.

5.3.3 Comparison of threat analysis results between SIEM and AI

This chapter evaluates performance through an experiment conducted to compare the threat analysis results obtained between the SIEM system and AI system. The SIEM system employs the rule-based analysis method, and the same event was used for both systems. The final experiment results of both analyses were compared in terms of speed, consistency, and insightfulness. Finally, the last picture shows the parsed log generated from the threat. The equipment where the relevant log was generated indicates that the event was generated in IPS and firewall (Fig. 11).



```
⊟ Detection Info.
     Category      : WebCGI attack
     Attack code   : 5832
    Attackname : and sleep( (Blind SQL Sleep() Query Injection Vul-3)
     IPPool     : GID:1000, 'OID:1001
     RiskLevel  : High
     Detect Time   : 2017-05-13 06:03:59~2017-05-13 06:03:59
    No.of Attemps : 1
  ⊟ Additional info : and sleep( (Blind SQL Sleep() Query Injection Vul-3)
        String Type      : [TXT][[case not distinguished ]
        Detected String  : and sleep(
⊞

00000000 | 00 00 5E 00 01 BD 00 24 C4 6A 3E BF 08 00 45 00   ..^....$.j>...E.
00000010 | 01 65 96 11 40 00 26 06 D4 68 C8 09 24 8B 3D F8   .e..@.&..h..$.=.
00000020 | BE 8C A9 DF 00 50 D4 F1 2E 38 F6 61 07 49 80 18   .....P...8.a.I..
00000030 | 00 E5 C7 CA 00 00 01 01 08 0A 63 00 B9 58 32 AC   ..........c..X2.
00000040 | 15 DD 47 45 54 20 2F 4D 42 5F 42 49 44 2F 43 2F   ..GET./MB_BID/C/
00000050 | 75 74 69 6C 2F 66 69 6C 65 5F 64 6F 77 6E 2E 6A   util/file_down.j
00000060 | 73 70 3F 66 5F 67 62 6E 3D 6D 31 25 32 37 25 32   sp?f_gbn=m1%27%2
00000070 | 30 41 6E 44 25 32 30 73 4C 65 65 70 25 32 38 33   0AnD%20sLeep%283
00000080 | 25 32 39 25 32 30 41 4E 64 25 32 30 25 32 37 31   %29%20ANd%20%271
00000090 | 26 66 5F 6E 61 6D 3D 39 34 32 30 57 31 33 53 31   &f_nam=9420W13S1
000000A0 | 39 36 30 30 31 31 2E 68 77 70 20 48 54 54 50 2F   960011.hwp.HTTP/
```

Fig. 12 Raw logs of IPS equipment

In the foregoing, the screen shows the detected event in X.X.X. 140 server and Internet IPS. After checking the threat, it can be seen that visibility was secured up to the Destination IP of 200.9.36.139 (Costa Rica). When analyzing with the SIEM system, a person has to perform directly the task of analyzing both reputation lookup and the threat blocked in the IPS. In order to analyze the correlation of the threat, a correlation analysis should be performed for the IP included in the threat, firewall log, and IPS log as shown in Figs. 11, 12. Such task will take about 50 min to 2 h if a person is to perform it. Moreover, for securing visibility to carry out a preemptive response, the human analyst will be able to check the threat occurring in Cost Rica only.

What should be anticipated is that the time required to analyze the IP information generated from the threat and associated log as well as its reputation lookup will increase depending on the situations. Elementary analysis such as reputation lookup, reading of related articles, and investigation of the hits recorded in the threat feed would take at least more than one hour even if SIEM has been used. In terms of consistency, there is a possibility that the analysis result will be varied due to the subjectivity of the analyst, who may use different types of information. Moreover, it will be difficult for the analyst to understand the correlations in a large volume of data as various logs need to be analyzed. Much human resources and time will be required to perform such task.

The following is a description of the comparative analysis carried out for the same event. Figures 13, 14 show that the event has been detected in the IPS and firewall, and the analysis by AI describes that better visibility has been secured since it shows that the Destination IP 200.9.36.139 (Costa Rica) is connected to the domain hotel-southbeachjaco.com. In addition, AI confirms that the Destination IP is connected to the malicious attack IP 50.62.160.99 to execute a spam activity. Because of such extended visibility, the analyst will be able to prepare a preemptive response system for the malicious IP.

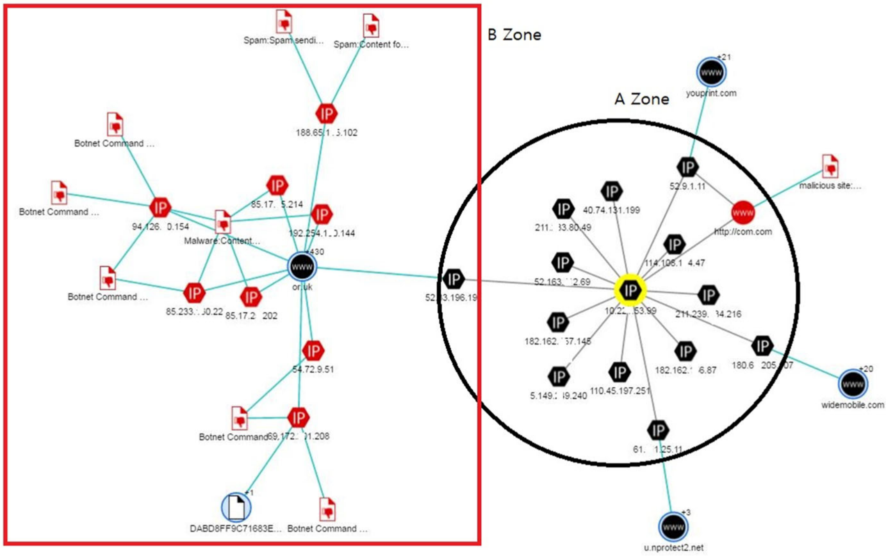
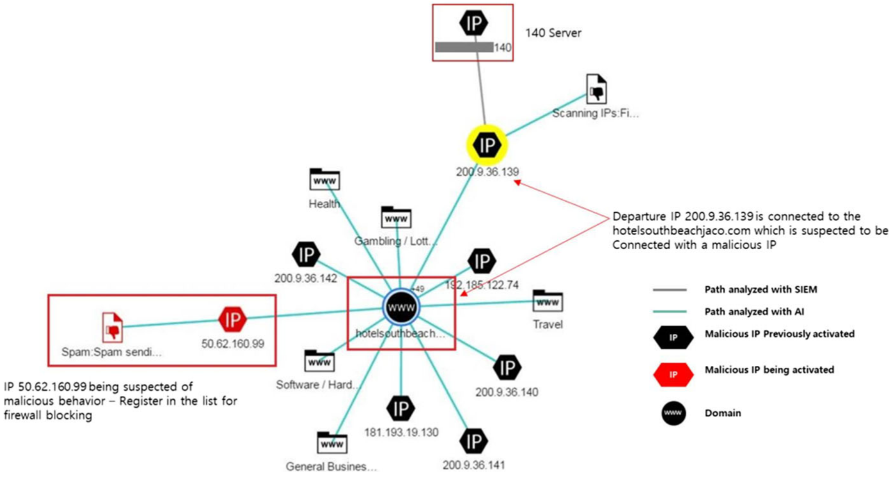


Fig. 13 Connecting flow of a malicious code analyzed by AI (1)



Departure IP 200.9.36.139 is connected to the hotelsouthbeachjaco.com which is suspected to be Connected with a malicious IP

IP 50.62.160.99 being suspected of malicious behavior – Register in the list for firewall blocking

Path analyzed with SIEM
Path analyzed with AI
Malicious IP Previously activated
Malicious IP being activated
Domain

Fig. 14 Connecting flow of a malicious code analyzed by AI (2)

The black circle in Fig. 13 shows the in-company logs and their links to the client log. In other words, the connection between a certain client and the infected log(s) and their communication routes are shown on the screen. The red box shows the connections to the malicious code originating from outside of the company through the Internet. To perform such task, the data associated with the logs that are being linked to the malicious code should be collected in real time and sent to the global cloud where the connections will be described with a map. Ultimately, the IP located at the intersection of red box and black circle will be the passageway of infection in

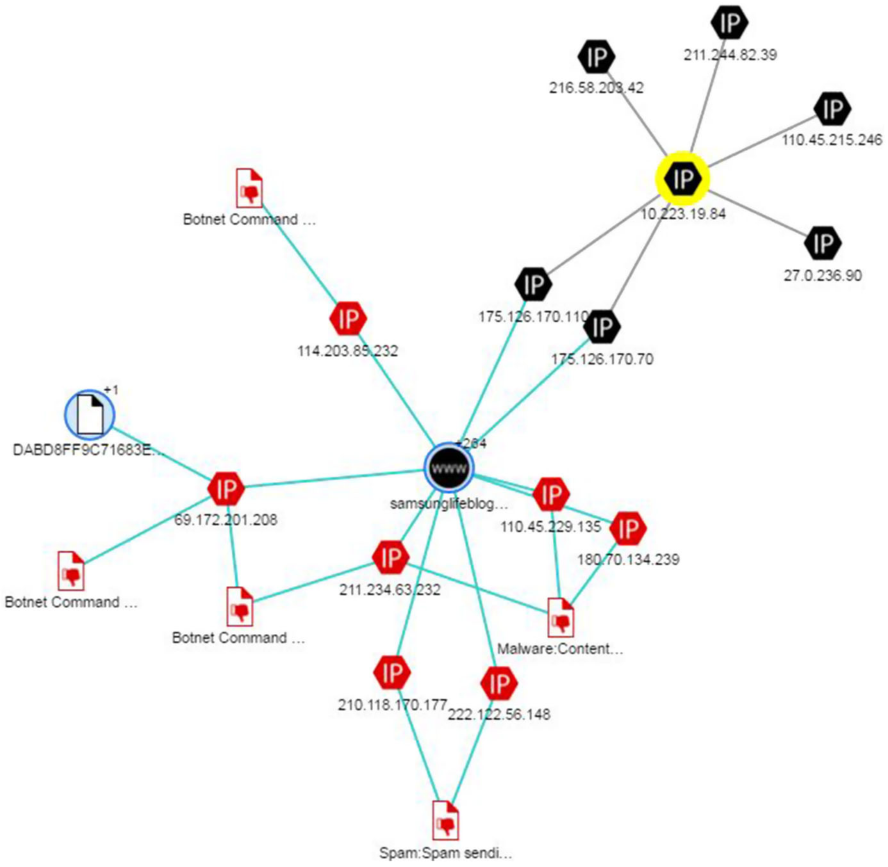


Fig. 15 Connecting flow of a malicious code analyzed by AI (3)

the company. It would take much time to find the trace of connection between the external malicious code and internal logs since it is not easy for the current analysis systems to find the connecting logs in real time and trace the route. On the other hand, AI can be much more effective since it refers to the past event history by connecting with the DB containing existing malicious codes and delivers the result to the security analyst intuitively on a real time basis.

In this picture, the internal IP 200.9.36.139 has been first infected via web surfing by connecting to the malignant IP 50.62.*.99 subsequently spreading to the others. The security controller can block the internal IP first, and then the malignant host IP initially causing the infection to clear the fundamental problem (Fig. 14).

This picture shows that the additional risk remains if only the internal IP used as an infection route is blocked. The original infection code can continuously infiltrate through other IPs, so it is imperative that the infector or the disseminator be completely blocked as well (Fig. 15). Thus, this time, the effect of SIEM and AI analysis for the same event was analyzed by adding the cost element to the three criteria presented in Table 5.

Table 5 Comparison of the effects between SIEM and AI

Term	SIEM	AI
Speed	Early response time—50 min	5 min
Consistency	Different information sources are used depending on the analyst, affecting the results	Consistent information can be provided
Insight	Too much data for the analyst to analyze correlations	The connections can be rapidly represented visually
Cost	*2 analysts High skill level: 5,973,240 won (M/M) Intermediate skill level: 4,757,277 won (M/M) (2016 SW skill wage statistics)	No other analyst is required

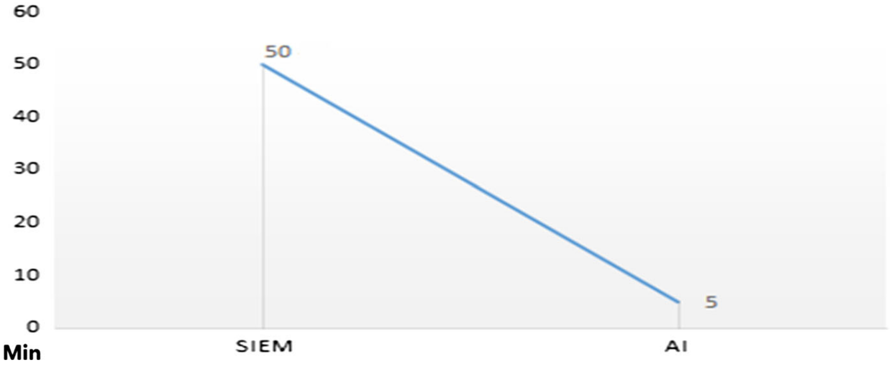


Fig. 16 Graph comparing the analysis speeds of SIEM and AI

For the cost calculation, there are two security analysts, and their skills are at the level where they can carry out a primary response to the threats by performing SIEM or Big Data analysis.

Figure 16 is a graph comparing the analysis speeds of SIEM and AI. Considering the development stage of security control, the first phase involved the evaluation of the security event occurring in the security equipment to assess the risk. The method of detecting the log is based on the signatures (e.g., Snort rule, PCRE-rule, etc.). In the second phase, the SIEM or the Big Data system comprehensively analyzes the event detected primarily, or the network flow, in the equipment. The third phase involves AI, which not only analyzes a single relation but also shows the original source together with the possibility of additional occurrences. AIs can distinguish unknown malware, find other potential infecting hosts, or determine whether the event has been falsely detected or not (Fig. 17).

Lastly, the fourth phase is the future direction of AI-based analysis wherein the risks are analyzed through the machine-learning process and the next steps will be suggested. For the analysis of cyber threats, current AI systems can analyze and describe correlations but the function of suggesting the adequate next step to be taken, like Alpha-Go

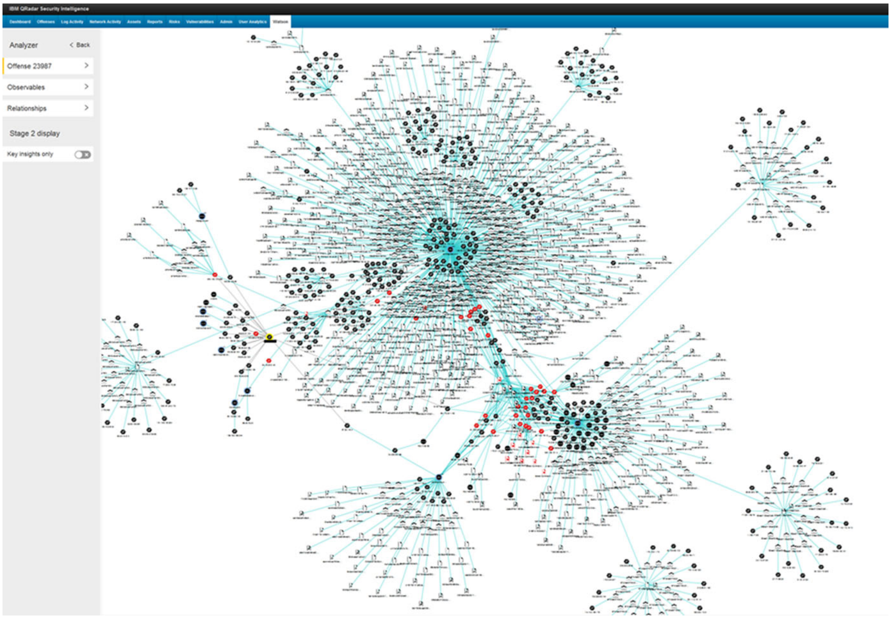


Fig. 17 Topology of all infection routes by malicious codes (1)

[15], for example, is still in the early stages. When this phase is reached, AIs will be able to detect threats and risks autonomously, perform analysis for countermeasure, and brief the security controllers or analysts on the results so that they will be able to outpace the hackers (Fig. 18).

In addition, it will be possible to reduce the security costs considerably by performing preventive inspection and predicting accidents through Zero-Day analysis or inspection of one's own information system. The costs due to hackings can be huge due to lawsuits, and the company could lose customers' trust.

Detection accuracy Basically, SIEM collects event log information of various devices at Smart Connector, and processes parsing in a regular pattern. It is followed by the analysis of correlation in the analysis system to bring about Alert. In the past, ESM (Enterprise Security Management) is an integrated management of log of various devices. SIEM collects data from various security devices and a big data concept is added I which correlation analysis is carried out to past data as well as present data.

Most vendors of security devices distribute signature each month. If there is a signature that needs to be distributed quickly such as Zero-day weakness, it is distributed from time to time. If the relevant signatures are set as 'Enable', an excessive event is detected that may cause trouble by the device load, and that is not efficient in terms of monitoring. Accordingly, signatures distributed by vendors are set as 'Enable/Disable' according to the service environments to detect relevant events.

It is set according to the asset type, OS and software which are found in the weakness diagnosis, and each signature provides affected software and system information together. The events detected by signature set in the device is transmitted to SIEM

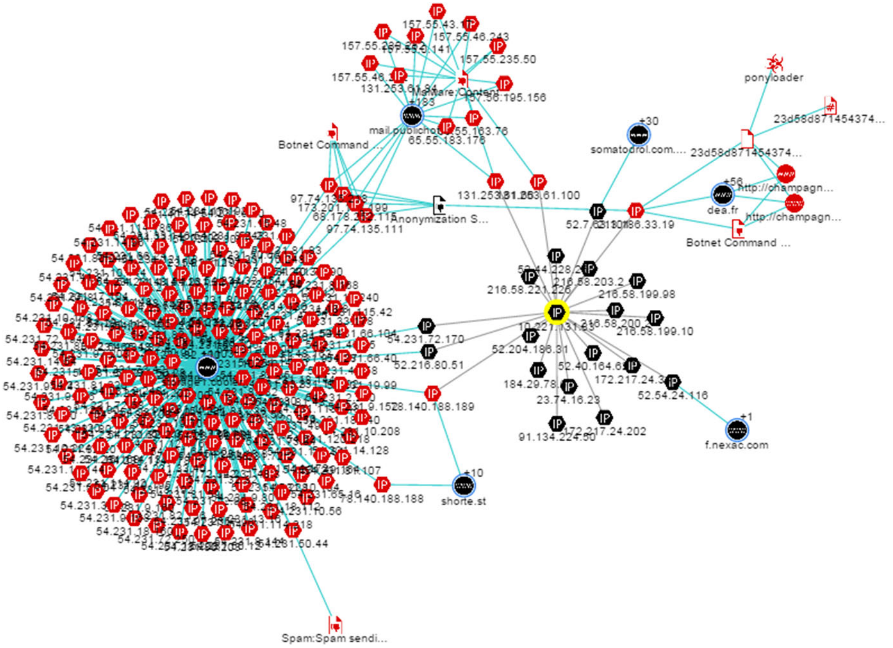


Fig. 18 Topology of all infection routes by the malicious codes (2)

and finally Alert is given to matters that the monitoring staff must check according to the event processing algorithm. If SIEM is perceived from the perspective of Rules, detected events pass through the filter of SIEM. Risk scoring method is also applied to Alert method like the rules connected to filter.

Figure 19 shows the display of insider information accessing botnet IP. Like the emergency detection events such as Zero-day, there is a rapid increasing attack by utilizing specific weaknesses globally due to the appearance of new Worms. If there is software weakness on the relevant assets and if Alert is required to all attacks utilizing relevant weaknesses, the detection can be made by connecting signature of security device and CVE information. Meanwhile, AI can figure out attacks that are less known than SIEM. AI (Watson) of IBM does not recognize the attempt of access using botnet IP as False-Positive, but analyzes and notifies accurately by comparing with various information.

In the event of detection, AI (Watson) conducts reputation search at external virus information sites such as X-force, VirusTotal etc. and checks if whether there is malicious code or not. And, it determines information from security related articles or from weakness related sites and enhances accuracy on the malicious code by surveying hit rates on harmful IP or has information registered in the threat survey sites. Therefore, it has far higher accuracy than SIEM. And, it is fare effective as the analysis speed and real-time reflection are quick.

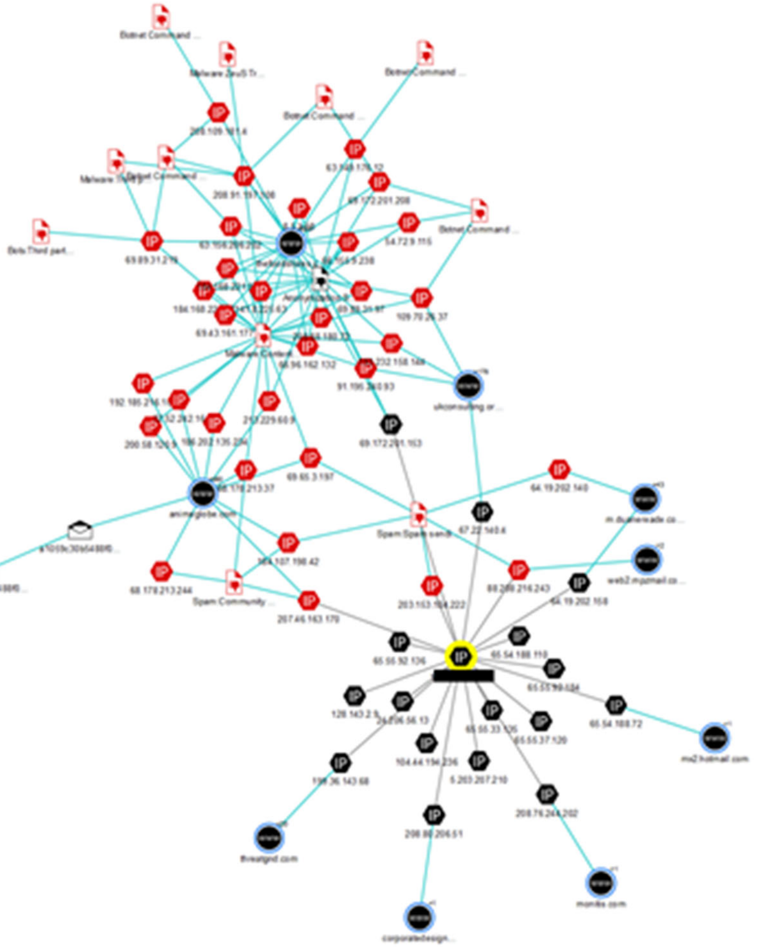


Fig. 19 Display of insider information accessing botnet IP

The type of error for each method While SIEM detects more threats by correlation analysis of various events, it cannot respond to software weaknesses that each asset contains.

Multiple number of security devices respond to security threats through SIEM. However, like weakness analysis tool, the standard on the signature weakness to detect is different to each other. Accordingly, it cannot recognize threats on software weaknesses that are in a specific asset and cannot exclude the detection on weakness removed assets. A monitoring security analyzer must understand service environment and registers it manually to SIEM with the weakness diagnosis results. It means weakness diagnosis result cannot be applied to SIEM flexibly.

In the monitoring system, the issue of False-Positive and False-Negative is very sensitive. When False-Positive is reduced, False-Negative increases. Therefore, an appropriate balance is required between False-Positive and False-Negative.

Table 6 Detection accuracy and type of error of SIEM and AI

	SIEM	AI
Detection accuracy	Alert incurred from security device is analyzed according to an algorithm. As it is poor in the analysis of significance, it is limited to see if the False-Positive takes place accurately by the information analyzer	It shows why detected information appears and how it is related so that it can reduce the possibility of False-Positive
Type of error	It has False-Positive and False-Negative and it needs to control it properly	It is more accurate than SIEM. It may have False-Positive and False-Negative to some degree of cognitive learning, but it can be reduced gradually

There might be False-Positive in the events detected in the security system. Thus, even if software weakness is known through the weakness diagnosis, and there is event to detect weakness of the relevant software, meaningless Alert rings if the event itself is False-Positive. The issue of False-Positive and False-Negative is a basic issue of security systems. The existing issue of False-Positive and False-Negative is not improved.

When SIEM is perceived from the perspective of Rules, detected events from the device passes through the filter of SIEM. The filter classifies the attack type and selects threat events between various events with conditions such as AND, OR etc. In the filter, filtering is made based on numerous information supplied from the device such as existence or nonexistence of a specific string and event count. And, correlation analysis on events of different equipment can be carried out.

Filtered events give Alert to the monitoring staff finally through Rules. Rules are part to give Alert to the monitoring staff by connecting filtered events and analyzing them. Finally, it sets information shown to workers. Table 6 shows detection accuracy and type of error of SIEM and AI.

In the series process, the vendor enables some distributed signatures for the service environment to reduce its number and classifies threats through filtering at the traffic where many events take place. Finally, Alert is given to the monitoring staff through Rules. Yet, there are two problems here.

Firstly, while more threats are detected than past through correlation analysis of various events in SIEM, it cannot respond software weakness held by assets. As filter is set according to the overall service which enables to cope with overall threats, it cannot respond to some assets and software.

Secondly, while many security devices respond to security threats through SIEM, there are different standards on the signature weakness to detect like weakness analysis tools. Accordingly, it cannot recognize threats on software weaknesses that are in a specific asset and cannot exclude the detection on weakness removed assets. A monitoring security analyzer must understand service environment and registers it manually to SIEM with the weakness diagnosis results. It means weakness diagnosis result cannot be applied to SIEM flexibly.

6 Conclusion

The existing behavior analyses have used a passive method that depends on the past data, but this study proposed an alternative method that can quickly detect threats based on the analysis of their patterns to prevent the leak of personal information. This way, it will be possible to extract behaviors that attempt to leak critical information by monitoring access by the web users and to trace back the hacker. This method also complies with the obligation of personal information access record management stipulated in the Personal Information Protection Act. By modularizing the system, the operator will be able to detect information leak due to an external hacker based on Big Data analysis and also detect an inadequate behavior of the internal user. The future study involves research on the automated data collection system using AI for user-behavior analysis.

AI is used in many fields ranging from voice recognition and pathfinding to medical diagnosis, thereby rapidly replacing human functions. Security analysis is one of the jobs requiring expert knowledge. While the importance of the security control center is recognized as a front-line key facility in protecting critical information, the technical skills of the controllers vary considerably and affect their response capabilities. Since AI can realize the reduction of such variation and conduct immediate analysis/response, it best fits with the security control tasks. The AI Watson is currently commercialized for cancer diagnosis, and it is achieving good results. AI is also being used for the security control area, and the commercialized version Watson for Security has already been introduced on the market. Watson analyzes the security events by interlinking the existing SIEM-based analysis data with the security cloud. Although it is playing only a secondary role by assisting the security analysts, it will be able to minimize hacking damages caused by global cyber-attacks using Ransomware such as WannaCry. Moreover, if it can perform adequate prediction diagnosis, and it is equipped with a self-learning capability like Alpha-Go, a proactive approach can be taken against cyber threats.

Acknowledgements This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korea government (MSIT) (No. 2017R1C1B5077157).

References

1. Lee S, Huh JH (2017) An efficient nuclear power plant security measure using big data analysis approach. In: The 2017 international conference on future information technology, applications and services, IFIT 2017, p 1
2. Huh JH, Kim TJ (2018) A location-based mobile health care facility search system for senior citizens. J Supercomput. https://doi.org/10.1007/s11227-018-2342-5
3. Tankard C (2012) Big data security. Netw Secur 7:5–8
4. Kim GH, Trimi S, Chung JH (2014) Big-data applications in the government sector. Commun ACM 57(3):78–85
5. Eom S, Huh JH (2018) Group signature with restrictive linkability: minimizing privacy exposure in ubiquitous environment. J Ambient Intell Human Comput. https://doi.org/10.1007/s12652-018-0698-2
6. Hirose K (2012) 2011 Fukushima Dai-ichi nuclear power plant accident: summary of regional radioactive deposition monitoring results. J Environ Radioactivity 111:13–17

7. Jakóbik A (2016) Big data security. In: Pop F, Kołodziej J, Di Martino B (eds) Resource management for big data platforms. Springer, Cham, pp 241–261
8. NIST Special Publication 1500-4 NIST big data interoperability. In: Security and privacy. NIST Big Data Public Working Group
9. He D, Jiajun B, Chan S (2013) Handauth: efficient handover authentication with conditional privacy for wireless networks. IEEE Trans Comput IEEE 62(3):616–622
10. Schneier B (2007) Applied cryptography protocols, algorithms, and source code in C. Wiley, New York, pp 1–34
11. Stallings W (2003) Cryptography and network security: principles and practice. Pearson, London, pp 1–16
12. Kitchin R (2014) The real-time city? Big data and smart urbanism. GeoJournal 79(1):1–14
13. Sola J, Sevilla J (1997) Importance of input data normalization for the application of neural networks to complex industrial problems. IEEE Trans Nuclear Sci 44(3):1464–1468
14. Pazhanirajaa N, Victer Paula P, Saleem Bashab MS, Dhavachelvanc P (2015) Big data and hadoop—a study in security perspective. In: Procedia computer science, 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15), vol 50. Elsevier, pp 596–601
15. Birkenmeier GF, Park JK, Rizvi ST (2008) Ring hulls of semi prime homomorphic images. In: Brzeziński T, Gómez Pardo JL, Shestakov I, Smith PF (eds) Modules and comodules. Springer, New York, pp 101–111
16. Birkenmeier GF, Park JK, Rizvi ST (2010) Principally quasi-Baer ring hulls. In: Van Huynh D, López-Permouth SR (eds) Advances in ring theory. Springer, Verlag Basel/Switzerland, pp 47–61
17. Lantz B (2013) Machine learning with R. Packt Publishing Ltd., Birmingham
18. Hao F, Park DS, Woo SY, Min SD, Park S (2016) Treatment planning in smart medical: a sustainable strategy. J Inf Process Syst 12(4):711–723
19. Silver David et al (2016) Mastering the game of Go with deep neural networks and tree search. Nature 529:484–489
20. Joo JW, Lee JK, Park JH (2015) Security considerations for a connected car. J Converg 6:1–9
21. Sharma PK, Moon SY, Park JH (2017) Block-VN: a distributed blockchain based vehicular network architecture in smart city. J Inf Process Syst 13:184–195
22. González-Aparicio MT, Ogunyadeka A, Younas M, Tuya J, Casado R (2017) Transaction processing in consistency-aware user's applications deployed on NoSQL databases. Hum Centric Comput Inf Sci 7(7):1–12
23. Ngu HCV, Huh J-H (2016) B + -tree construction on massive data with Hadoop. In: Hariri S (ed) Cluster computing. Springer, New York, pp 1–11
24. Lu T, Guo X, Xu B, Zhao L, Peng Y, Yang H (2013) Next big thing in big data: the security of the ICT supply chain. In: International Conference on Social Computing (SocialCom). IEEE, pp 1066–1073
25. Cardenas AA, Manadhata PK, Rajan SP (2013) Big data analytics for security. IEEE Secur Priv 11(6):74–76
26. Huh Jun-Ho, Seo Kyungryong (2016) Design and test bed experiments of server operation system using virtualization technology. Hum Centric Comput Inf Sci 6(1):1–21
27. Zhao G, Rong Ch, Gilje Jaatun M, Sandnes FE (2012) Reference deployment models for eliminating user concerns on cloud security. J Supercomput 61(2):337–352
28. Huh J-H (2017) PLC-based design of monitoring system for ICT-integrated vertical fish farm. Hum Centric Comput Inf Sci 7(20):1–19
29. Patil HK, Seshadri R (2014) Big data security and privacy issues in healthcare. In: 2014 IEEE International Congress on Big Data (BigData Congress). IEEE, pp 762–765
30. Huh JH (2018) Implementation of lightweight intrusion detection model for security of smart green house and vertical farm. Int J Distrib Sens Netw 14(4):1–11
31. Vosoughi S, Roy D, Aral S (2018) The spread of true and false news online. Science 359:1146–1151
32. Jordan MI, Mitchell TM (2015) Machine learning: trends, perspectives, and prospects. Science 349:255–260
33. Huh Jun-Ho (2018) Big data analysis for personalized health activities: machine learning processing for automatic keyword extraction approach. Symmetry 10(4):1–30
34. Moon Seo Yeon, Park Jong Hyuk (2016) Efficient hardware-based code convertor of a quantum computer. J Converg 7:1–9
35. Liu H, Gegov A, Cocea A (2016) Rule based systems for big data a machine learning approach. Springer, New York, pp 1–43

36. Medeiros J, Schirru R (2008) Identification of nuclear power plant transients using the Particle Swarm Optimization algorithm. Ann Nucl Energy 35(4):576–582
37. Huh J-H, Otgonchimeg S, Seo K (2016) Advanced metering infrastructure design and test bed experiment using intelligent agents: focusing on the PLC network base technology for Smart Grid system. J Supercomput 72(5):1862–1877
38. Feng D, Zhang M, Li H (2014) Big data security and privacy protection. Chin J Comput 37(1):246–258
39. Huh J-H (2017) Smart grid test bed using OPNET and power line communication. In: Naumann F, Shasha D, Vossen G (eds) Advances in computer and electrical engineering. IGI Global, Pennsylvania, pp 1–425
40. Hashem IAT, Yaqoob I, Anuar NB, Mokhtar S, Gani A, Khan SU (2015) The rise of "big data" on cloud computing: review and open research issues. Inf Syst 47:98–115
41. Hewitt C (1991) Open information systems semantics for distributed artificial intelligence. Artif Intell 47(1–3):79–106