Security Framework using Hbase and Log Management Technology

Nan ju Kim
University of Hannam
Department of Computer Engineering
Republic of Korea
91knj@naver.com

Joon Woo Park
University of Hannam
Department of Computer Engineering
Republic of Korea
zickblue@naver.com

Yu Jin Kim
University of Hannam
Department of Computer Engineering
Republic of Korea
rladb4343@naver.com

Eui In Choi
University of Hannam
Department of Computer Engineering
Republic of Korea
eichoi@hnu.kr

Seul Gi Lim
University of Hannam
Department of Computer Engineering
Republic of Korea
seres92@naver.com

Abstract—The explosive development of the IT field, the number of structured and unstructured data is increasing in geometrical. According to this tendency, the importance of Big Data and analysis techniques has received attention. Also, using these data a number of cyber-crime has occurred. So, the security threat in many areas is a serious problem. But, Efforts about security are showing a low level relatively. Recently, Security intelligence that defined by Gartner Group has attracted attention as a new concept for the cyber threats. That is based on technology utilizing a Big Data processing, analysis technology for a variety of large data. In this paper, we will learn about SIEM as security system management solution and Big Data processing, analysis technology that received attention as security response based technology. We proposed a new security technique that complements traditional security techniques.

Keywords- Big Data Security; Hbase; SEIM; Big Data technology

I. INTRODUCTION

As Mobile and social media, smart devices, Cloud, Internet of Things (IoT) will evolve unstructured data such as photos, images, videos and structured data increasing rapidly. However, valuable data among numerous data is handful. The need of technology for finding the data appeared. So, interest for Big Data is increasing.

Security threats are increasing by the open network use and the explosive growth in data. Contain complex security issues of to be solved in the field of multiple such as data protection, resource management, ensuring the availability, Privacy. So, Big Data and information security cannot think of situation they were separated.

Recently, over the years, domestic and foreign have been found in many cases. In July 2010, In case of Stuxnet hacked Iran nuclear power facility, 20% of the centrifuge facility was shut down. In April 2011 Korea Nonghyup computer network Resources damaged and in the same year in July, attacks such as leakage of personal information of 3500 ten thousand people in Nate caused extensive damage of large-scale. [1] Thus, Security threat began known attack such as Worm and viruses, and intelligent attack changed. So, it is threatened monetary damages as well as the social chaos, national security.

In addition, recently, Security intelligence that defined by Gartner Group has attracted attention as a new concept for the cyber threats. That is based on technology utilizing a Big data processing, analysis technology for a variety of large data. That is focused on predicting in advance and defending against unknown security threats. Therefore, it is estimated to be positioned as the core concept of future cyber defense technology. [2]

In this paper, we list about definition and technique of Big Data in IT field buzzword. Also, we will learn about SIEM as security system management solution that can respond to security threats by combining event and log data. Such security threats, security measures and proposing security mechanism are described. And, we proposed a new security technique that complements traditional security techniques.

II. Big data definition and technology

A. Big Data Definition

Big Data (Wikipedia) means a large number of structured and unstructured dataset beyond capability of the data collection, management, and analysis of traditional database managing tool and extract value from these data and technology to analyze the results. [3] In addition, according to the report of Mackinsey, Big Data means data of scale exceeds range of data that can be collected, stored, managed, analysis of traditional database managing tool. [4]

B. Big Data Technology

The three elements of Big Data 3V has been described Section 2.1. The various implementation techniques such as systems, software and storage devices are needed to provide 3V. These various techniques can be classified into following five configurations step significantly. [5, 6]

1) Data Collection/Integration Step

Technologies of data collection and integration step include new data generating, disparate data consolidation of Internal and external, external data collection scattered in network and the like. It means technology to secure the data that irrespective of the type of data and the material. Technique such as collection robot, data virtualization and logging station is used in the data collection and integration step.

2) Data Preprocessing Step

Data preprocessing step includes technology that performs the following function. Continuously generated unstructured stream data such as SNS and sensing information in collected information from first step is purified. Thereafter, that data transforms structured data that possible form of analysis. Then, raise the accuracy of analysis and makes possible to depth analysis.

3) Data Storage/Management Step

Data storage and management technology means the distributed computing technology that various types of data such as web data, social media, business data and sensing information can be stored and managed in real-time. It means core technology of Big Data platform. NoSQL (Not Only SQL) of big data storage technology beyond the RDB (Relational Data Base) is new database concept for Big Data storage. It provides interface of various types. And, it is non-standard state. Types are MongoDB, Cassandra, and Hbase.

4) Data Analysis Step

Data analysis technique is a variety depth analysis techniques for extracting value contained in the Big Data. It is utilized technology such as large-scale statistical, data mining, graph mining, machine learning, artificial intelligence and the like.

5) Data Visualizing Step

This is technology showing visually the analyzed data. Also, it means technology expressing easily to know feature or meaning of the analysis data.

III. Security in big data environment

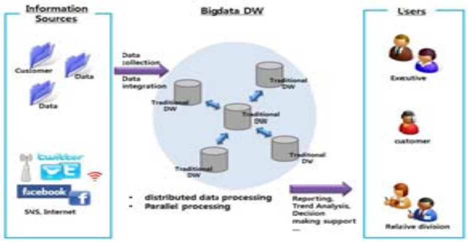

Figure 1. Big Data processing

As shown in Figure 1, the first step is process that collecting producing data through a variety of sources. The second process is operation and distributed storage of data for a distributed processing and parallel processing. Final process is showing reuse as service through data analysis and secondary data generation. [5] Section 3, According to these, Security threats and security technology will be described. And propose new security technology.

A. Big Data collection section

Generated and collected large amounts data through a variety of paths are accompanied by security threats of large number. Recently, deepening cyber targeted attack threats generate social and national risk. ATP (Advanced Persistent Threats) is being utilizing attacks method such as cyber terror, Cyber Warfare and hacktivism. It is clear systematic attacks targeted. It is targeting government and corporate mainly the for the purpose of information deodorization such as military secrets, industry secrets. [1] Accordingly, increases concerns about the authenticity and integrity of data. Also, Company utilize smart phone and tablet PC as means of adjuvant business. Such companies are on the rise. So, the number of collected data through individual IT terminal is increasing. Therefore, BYOD (Bring Your Own Device) threats are on the rise. BYOD means the phenomenon of to utilize privately owned IT devices as business. Consequently, Major security threats has personal information leakage, account leakage by the staff low security awareness, data loss caused by lost or stolen devices, corporate information leakage caused by vulnerability of the terminal with malicious code and loss of IT control of corporate. [7] May be cause personal information problems that collected and used personal data indiscriminately due to leakage of personal information. To solve these two problems have been conducted various studies. Electronic signatures, various filtering techniques, anti-spam, anti-phishing, and at least to personal information being collection such techniques are applying.

B. Big Data analysis section

This section is process, store, operation and analysis to the data collected through various paths. So, this part may be exposed to attack threat from the inside as well as outside. Security measures method through Big Data analysis technology is explain.

As follows:

6) Intelligent security management system

Intelligent security management means technology that corresponds to the unknown deadly attack. It analyzes the relevance of the data and security event arising from network, system and application service of major IT-based Facilities. That is, it means next-generation security information analysis technology to improve the security management intelligence. [1]

7) Unit Security System

Typical unit security systems are intrusion prevention system and intrusion detection system. Next-generation intrusion prevention system overcomes the limitations of existing security solutions. User-oriented Detection analysis and control functions in application layer can be performed [7].

8) Integrated Security 2.0

Typically, the integrated security means system that event and log of security solution as intrusion prevention system, intrusion detection system and virtual private networks integrated. In this paper, the integration security of big data former era was defined as integrated security 1.0. Integrated security 1.0 is difficult that real-time analysis of the data. But, Integrated security 2.0 is possible that real-time analysis of the data by using long-term data through to the distributed parallel processing. Also, Integrated security 2.0 is possible a parallel processing by considering collection performance and analysis performance.

9) Hadoop Security

There are two known security vulnerabilities in Hadoop. One is issue using single symmetric key cipher key. The other objects issue is that do not support Hadoop security in Hadoop ecosystem system. Recently Hadoop 1.0 was released. So, it provides Kerberos authentication, RPC digest method, SASL (Simple Authentication and Security Layer) through GSSAPI (Generic Services Application Program Interface). [8]

10) STAP (Specialized Threat Analysis)

STAP technique contains signature-based anti-malware code, Intrusion prevention and Protection system. STAP products are operating in network layer or endpoints or both. It scans and detects the abnormality activity for incoming and outgoing traffic containing C&C traffic and Botnet to inform typical infiltration state. [9]

11) SIEM (Security Information and Event management)

SIEM get log information and security event information from security solution, server and network equipment of firewall, IDS/IPS and anti-virus. And, it is correlation analysis between this information. So, it provides the ability that to perform security situational awareness, rapid incident response and log management [10]. SIEM will be in important role as effective integrated log management, risk detection, incident response, forensic and security compliance. But, SIEM operates in generally the rule-based. Therefore, you should not rely on one technique called SIEM. It needs to security technology combination that detection possible about unknown attack.

C. Big data secondary data creation and use section

Creation and use process of secondary data is needed in order to extract the user desired data. When creating the secondary data, invasion of privacy and confidentiality of the data is the risk of exposure. Anonymization and encryption techniques should be introduced surely to protect your privacy. Technology such as keyword-based search technology and PPDM (Privacy Preserving Data Mining) has been studied. The method of PPDM is converted to protect your personal information, or is used method that it can protect. Then, performing data mining and that result obtained. [8]

IV. PROPOSED SECURITY SYSTEM

Intelligent security threats such as target clear APT attacks are increasing. So, to defend using traditional security technology is difficult. As mentioned earlier, research and developed for security response technology is underway. Security response technology using Big Data system, storage and analysis techniques exist. Accordingly we propose a new security technique that complements traditional security techniques.

The objective of the Intelligent Security Technology is an integrated security management technology that consist the security infrastructure. Security infrastructure is interlocking network and system event using variety and numerous data.

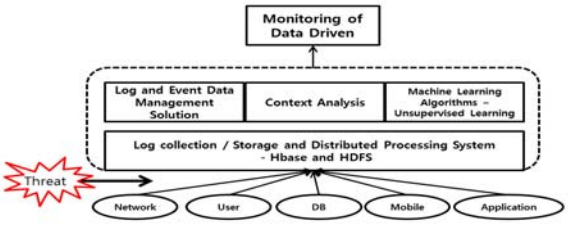
Monitoring of Data Driven
Log and Event Data Management Solution
Context Analysis
Machine Learning Algorithms – Unsupervised Learning
Log collection / Storage and Distributed Processing System - Hbase and HDFS
Threat
Network
User
DB
Mobile
Application

Figure 2. Proposed security architecture

Looking at Figure 2, when the threat invade, log and event data management is using SIEM technology. Also, to use a machine learning algorithm can determine whether or not normal and abnormal with respect events occurring in real-time. Unsupervised learning is effective in a new type of attack detected. In addition, part about log collection, storage and distributed processing is managed Big Data platform such as Hbase and HDFS. The HDFS is used for data storage and replication. The Hbase is charge of data distribution and monitoring. The maximum advantage of Hbase is short downtime even if a failure occurs. Also, it is ensured the data integrity about data recovery.

V. CONCLUSION

Security threats are increasing due to the development of Internet, IT technology and smart terminals. Due to these environments, the trend of security threat was changing. The initial attack has short-term and the purpose of joke. But security threat was changed as long-term, intelligent, the purpose of social confuse. Detection to the current security threat is difficult as traditional security technology. Interest in data security with the activation of Big Data advantage is growing. Security information and event management solutions are emerging as an essential component in security infrastructure in your organization. SIEM will be in important role as effective integrated log management, risk detection, incident response, forensic and security compliance. Also, Shall examine the traffic and log data in order to an unknown attack detection. So, Should take advantage of the Big Data platform because it is difficult to analyze the log and traffic in the tens of thousands.

The problem of traditional security technologies and currently being developed security technologies should analyze the pros and cons. And it is necessary to develop security technology using it. Also, the current security threat is occurring a problem in the company as well as the national level. So, platform development need for data protection from within government agencies and enterprise.

ACKNOWLEDGMENTS

This work was supported by the Security Engineering Research Center granted by the Ministry of Trade, Industry and Energy.

REFERENCES

[1] J.H. Kim, S.H. Lim, I.K. Kim, H.S. Cho, B.K. No, 2013, Technical Trends of Cyber Security with Big Data, *2013 Electronics and Telecommunications Trends*, ETRI, pp.19-29.

[2] I.K Kim, 2014, Big Data analysis technology and cyber security, *ETRI Special Report*, pp.16-22.

[3] Wikipedia, Big Data.

[4] James Manyika & Michael Chui, 2011, Big data: The next frontier for innovation, competition, and productivity, *McKinsey Global Instituted.*

[5] K.I. Jeong, H.N. Park, B.G. Jung, J.S. Jang, M.A. Jung, Big Data and Information Security, *Journal of Korea Institute of Information Technology*, 10(3), pp. 17-22.

[6] Y.Y. Joe, 2013, Understanding of Big Data and Major Issues, *The Korean Association for Regional Information Society*, 16(3).

[7] D.S. Choi, Y.M. Kim, 2012, Big Data and Integrated Security 2.0, *Journal of Information Science*, pp. 65-72.

[8] D.W. Lee, 2012, Security Issues of Big Data Period.

[9] IDG IT WORLD, Company's Security Strategy Corresponding to APT Threats, *IDG Tech Focus*.

[10] Byung-chul Kim, 2013, Big Data Security Technology and Response Study, *The Journal of Digital Policy & Management*, 11(10), pp.445-451.