ISSN 2409-2665
Journal of Logistics, Informatics and Service Science
Vol. 9 (2022) No.1, pp.26-35
DOI:10.33168/LISS.2022.0103

An Empirical Study of Intelligent Security Analysis Methods Utilizing Big Data

Yang-ha Chun [1] and Moon-ki Cho [2]

[1] Yongin University, Cheoin-gu, Yongin-si, Gyeonggi-do, Korea.
[2+] Hanyang University, Sangnok-gu, Ansan-si, Gyeonggi-do, Korea.

poletopole@hanyang.ac.kr

Abstract. The frequency of cyber-terrorist attacks targeting broadcasting companies or key financial institutions, in which PCs affected by malicious code obstruct normal operations, has been increasing. With the widespread availability of internet technology, the frequency of intelligent cyber-attacks in virtual cyber environments, such as APT attacks, continues to rise. Through an intelligent analysis of cyber threats which are still unknown, it is possible to prevent security accidents before they occur through the use of intelligent security analysis methods based on big data. The researcher conducted this empirical study concerning intelligent security analysis methods utilizing big data which are capable of preventing cyber threats, while protecting information assets from risks through evaluation and prediction.

Keywords: Behavior-Based, Signature, APT, Big Data Analytics, ISO27001.

1. Introduction

Recently, broadcasting companies and key financial institutions have been targets of occasional cyber terrorist attacks, which render them incapable of normal operations due to the infection of their PCs with malicious codes (IDC, 2019). Cyber-attacks of these types are used to steal the information of an organization over a prolonged period. As such, tremendous damages are expected when an organization is attacked in this way (J. Feiman,2012, J. Manyika et al,2011).

Furthermore, it is impossible to detect these attacks before they occur because they use social engineering hacking methods, such as malicious codes. To respond to such threats, it is now necessary to employ intelligent security analysis, and is critical to study big data-based analysis technologies(N. MacDonald,2017). With the advance of new IT technologies capable of the real-time processing of large amounts of data, such as big data, it is now intended to develop a system of analysis, in which the unusual traffic packets and attack log events are gathered, so that the interrelations between them are analyzed(S. Curry et al, 2013).

This will make it possible to detect the intent of an attack and the possible targets, defending the system from the eventual attack. It is also intended here to analyze not only the unusual traffic and security system logs but also the composition information of the systems, networks, and applications and their status, as well as other multi-source data so that their interrelations can be analyzed to help to prevent such attacks. The analysis of unknown, recent attacks may also help such efforts(M. Nicolett and K.M. Kavanagh,2021). In particular, there is a need for methods to respond to and defend against the next generation of cyber-attacks, which disguise themselves as normal activities, as is the case with zero-day vulnerabilities and social-engineering methods. These attacks exploit the weakness of signature database-based detection technology, which thus should be complemented by new technology(M. Nicolett and J. Feiman, 2019).

In addition, there is a need for a specialized analysis method for the intelligent security analysis utilizing big data, as well as technologies for storing and high-speed processing of distributed searching to support it. In this study, the author performed an empirical study on an intelligent security analysis method utilizing big data, which has been gaining attention recently, while examining the changes in the technologies used in cyber-attacks, which are continuously growing smarter, and the technologies to protect against them(ANSI/ISA-TR99.00.01,2020, ANSI/ISA-TR99.00.02,2014).

In Chapter 2 of this article, the author conducted a literature review that describes the evolution of cyber-attack technologies, which is followed by Chapter 3 in which the intelligent security analysis method utilizing big data is presented. In Chapter 4, the author explains the intelligent behavioral analysis and verification methods. Finally, Chapter 5 provides the conclusion and suggestions for future research.

2. Literature review

2.1. The Evolution of Cyber-Attacks

With the evolution of the technologies used in cyber-attacks, they are now considered one of the major threats against the country and society. One of the most commonly used types of cyber-attack is the APT (Advanced Persistent Threat) attack. An APT attack continuously attacks its target with the goal of stealing industrial secrets, classified military intel, and customer information, etc. In addition, as this attack is used for major information leakage, control system attacks, and cyber terrorism, the threat of cyber-targeted attacks targeting major information and communication infrastructure is increasing.

Some of the recent examples of such attacks on overseas businesses include the attack on the five major US energy companies (Shell, Exxon Mobil, Marathon Oil, Conoco Phillips, and Baker Hughes), in which information on gas and oil production systems, financial documents related to oil exploration, and industrial control systems were stolen. One of the leading security companies, EMC RSA, also fell victim to a social engineering attack, and confidential information on its OTP (one-time-password) product, Secure ID was breached. Also, in July of the same year, the personal information of 35 million people stored in Nate's database was leaked. During the analysis of the attack on Lockheed Martin, a U.S. defense company, it was found that a total of 760 world-famous companies, including domestic telecommunication companies, were attacked. These attacks were all persistent attacks extended over time upon the PCs of the developers of the company.

2.2. The Current Status of Cyber Security in South Korea

Globally and particularly in the United States, the importance of cyber security is continuously increasing. The United States has become the de facto leader in global cyber security as a result of its diplomatic conflicts with the Arab countries in the Middle East. For this reason, the United States Department of Defense has focused on developing cyber weapons to prevent and prepare against cyber-terrorist attacks. China, North Korea, and South Korea are also developing policies emphasizing cyber security technologies to serve their interests (Defense Dept, 2015).

In South Korea, efforts are being made to create a safe environment by improving the laws, institutions, budgets, and organizations for information security in preparation for future cyber threats. The government has established a cyber crisis management system plan and is preparing for the future war over information security. For this purpose, the South Korean Military established the Cyber Command HQ under the Ministry of Defense, as the dedicated organization for information security and has been implementing information protection policies (Jason Stamp, 2005). However, these efforts are failing to keep pace with the highly advanced and intelligent cyber-attacks. The intent has been to focus on establishing

the procedures and systems that will protect the information system and respond to cyber threats (P.A.S. Ralston, J. H. Graham, J. L. Hieb, 2017).

Mckinsey defined big data as a group of data that goes beyond the capability of typical database management tools to gather, store, manage and analyze. Gartner also defined it as having the properties of high variety, volume, velocity, and complexity. Big data analyses can be divided into big data processing methods and big data analysis methods. Big data processing methods are the kind of technologies that are capable of unformatted, large volume mass data processing using distributed computing technologies based on Hadoop and MapReduce technologies (Nir Kshetri,2019). Big data analysis methods are technologies that support the analysis of large volumes of data using various advanced analysis techniques that apply open-source software for data analysis represented by R. Big data analysis technologies include, in particular, advanced data mining, text mining, opinion mining, and social mining.

3. Intelligent Security Analysis Using Big Data

Intelligent security can be defined as a next generation security information analysis technology that analyses the correlations between the data and security events happening in key IT infra network systems and application services, in order to respond to unknown critical attacks such as an APT attack. With the IT environment changing at a dizzying pace, cyber-attacks are becoming harder to detect, and are meticulously organized. As a result, the existing security approaches are revealing their limitations. For this reason, it is important to detect the correlations between different threat elements, rather than trying to block them individually, to detect a stealth attack underway. This involves the analysis of correlations to detect new forms of attacks that have not been previously observed, which is the weakness of the pattern-based attack control methods (Nir Kshetri,2019).

3.1. The Key Elements of Big Data Analysis

This is the architecture that supports the high-speed gathering, processing, and analysis of formatted or non-formatted data that are extracted from security events or abnormal network traffic data. The key elements of the studies on intelligent security analysis technologies utilizing big data processing technologies are shown in Table 1 below.

Table 1. Big Data Analysis Core Elements

Core Elements	Detailed Description
Association rule Learning	It is a technique for finding association rules of interesting topics from various variables in a large database, consisting of a set of algorithms that

	generate and test potential rules.
Classification	One of the methods of data mining that can identify the category to which new data belongs based on a training data set that includes already identified data
Cluster Analysis	A statistical method for dividing into small groups of similar entities in a state where the characteristics of similarity, etc. are not known in advance. For example, it is used to group customers into groups with self-similarity for target marketing
Fusion & Integration	A technology that integrates and analyzes data from multiple sources to obtain more accurate and efficient insights than the results of analysis from a single source
Data Mining	A technique for extracting specific patterns from large data sets by combining database management with statistical and machine learning methods
Ensemble Learning	A method of learning a new hypothesis by creating multiple classifiers and combining their predictions through the classification method of machine learning
Genetic Algorithm	A method of solving optimization problems as a computational model based on the evolutionary process of the natural world
Visualization	Technology used in image diagrams, animations, etc. to express the results of data analysis and improve the level of understanding

3.2. Intelligent Security Analysis Methods Using Big Data

The existing technologies against cyber-attacks utilize network-based security devices such as an IDS (intrusion detection system) or IPS (intrusion prevention system), as well as the technologies that are designed to prevent information leakage from the inside, such as anti-virus, database encryption, or other technologies to prevent information theft. In addition, such conventional technologies use integrated security management methods that manage the logs of security products on different models. In particular, firewalls, intrusion detection/prevention systems, and anti-virus technologies are focused on the development of the platform technology needed to apply signature or blacklist-based detection methods to 10Gbp-level high-performance networks.

Also, the changes in the key technologies to counter cyber-attacks were based

on integrated security management technologies, and it is expected that an intelligent security analysis model using big data will be the prominent topic of the future. By seamlessly integrating security equipment that has previously been operated separately and managing them as a whole, it will be possible to conduct an intelligent analysis of security threats, which will allow us to predict future problems in parallel with behavior-based analysis, to prevent and remove security risks.

Big data security analysis methods will allow the threats of attacks that have not been resolved to be addressed, utilizing the security system that will ensure seamless responses to the latest cyber-attacks. The correlation analysis of big data through these security systems will reveal the unknown patterns of the attacks and support decision-making. By implementing a high-performance pattern matching algorithm through the behavior-based signature database, it would be possible to make comparisons at a higher speed. Using composite data processing technologies, it would be possible to understand the characteristics of a large amount of accumulated data from multiple media. The security analysis methods that use big data are shown in Table 2 below.

Table 2. Security Analysis Method using Big Data

Division	Detailed Description
Realtime monitoring	Technology that collects and manages data from various sources to track and analyze the attack situation on system components or to monitor user activity in application programs
Threat intelligence	It is an up-to-date information system on various threats and attack patterns that enables more accurate recognition of abnormal activities. Technology that recognizes a small amount of outbound traffic to an external IP as threat information related to attack control when it is disguised as normal traffic and can be overlooked
Behavior profiling	When the conditions for anomalies are well defined, it is possible to define an association rule to find a set of specific conditions.) detection analysis is the main technology
Data & User Monitoring	Monitoring data/user activity, including user and data context, is an essential skill for intrusion detection and misuse detection, and is fundamentally required for monitoring privileged users and sensitive data access
Application monitoring	Since application vulnerabilities are the main target of targeted attacks, the activity

	monitoring technology of abnormal applications
Analytics	A core technology for big data security analysis that analyzes the characteristics of various source information and determines whether there is an abnormality using machine learning, data mining, and network mining techniques, which are traditional methods of big data analysis

4. Behavior-Based Intelligent Security Analysis Methods and Verification Results

In this study, security threats are classified by reconstructing the records of persistent attacks recorded in security solutions, servers, applications, and other events to enable a response to cyber-attacks in real time. That is, the researcher intends to study a behavior-based intelligent security analysis method in which detection rules can be used to understand the relationships of the events before and after an attack. This behavior-based technology will allow attacks to be predicted before they happen, making it possible to make an immediate decision.

4.1. Behavior-based Intelligent Security Analysis

An attacker uses various cyber-attack tools to get the information he or she wants. Some well-known examples of these include malicious codes, phishing, spam, DDOS, and viruses. However, we are always responding to these threats reactively. Now, it is time for us to find a new way to analyze, determine, and respond to these security risks differently.

One such approach can be a behavior-based, correlated analysis method for security threats which the author intends to study herein. The security threats of cyber-terrorist attacks are similar to those of the real world. Once an act is committed, it leaves a log entry or a packet. Analysts use these as evidence to track down the source. Furthermore, identical actions can happen in different places. Behavior-based correlation analysis allows analysts to look for events with similar patterns, in terms of locations, time, evidence, and behaviors. In this way, it would be possible to link two or more events, identified as interlinked. As such, cyber security attackers also leave traces that remain after their attacks. The resources, environments, purposes, stages, and results of their attacks are analyzed to identify correlations, revealing an identical attacker or frequently used resources, to prevent further damages. Correlation analysis methods for security threats can be classified into three types.

First, the information on malicious code. One of the most common tools of attackers is malicious code. By analyzing malicious code, relationships with similar malicious code can be found. Second, the information on the use of the network.

The network information that has been used is analyzed to find links with the network information used in other security threat elements. Third, the analysis of the trends of key attackers and their attack resources. Entities which are suspected to be attackers are monitored. Based on how they secure the resources for attacking and their activation, threats of targeted attacks can be predicted and reported.

4.2. Behavior-based Signatures

Behavior-based signatures are defined using the characteristics of the traffic (i.e., the entity) and the traffic analysis units (i.e., the flow) as their attributes. The destination IP, destination port, transmission layer protocol, and the payload of the first N bites used in behavior-based signatures are the four traffic characteristics. By using only the first N bytes, the use of using the payload of a private information protection and a fixed position (offset, length) among the payload information containing the key to identify the application solves the computational complexity problem thru using more than 10 bytes for the value of N to prevent extraction of method keywords only in the case of HTTP traffic, and thru setting more than 2 bytes in case of the non-Http traffic, respectively. With this approach, it is possible to solve the complexity of the calculation issue. The first packet of various flows is inspected and the signature is applied to the inquiring packet so that real-time control and accurate traffic analysis can be enabled. The behavior-based signature (BS) is composed of a combination of the entries and other attributes. Multiple entries have the characteristics of traffic, and the formula below represents a behavior-based signature and an entry, respectively.

$$\begin{aligned} \mathrm{BS} &= \{\mathrm{A,T,I\ ,E1,E2,...E}n \mid n \geq 2,\ \mathrm{Src(E1\ )} \\ &\quad = \mathrm{Src(E2\ )} = \ldots = \mathrm{Src(E}n\ \mathrm{)}\} \\ \mathrm{E} &= \{\mathrm{X} \mid \mathrm{X} \subset \{ip,\ port,\ prot,\ payload\},\ \mathrm{X} \neq \varnothing\} \end{aligned}$$

Since a specific host is used as the reference, the starting IP of all entries must be the same.

4.3. Extracting Behavior-Based Signatures

The traffic is converted from packages to five tuples of the flow (starting IP, starting port, destination IP, destination port, and transmission layer protocol) so that the relationships between flows and the statistical information such as the size and the duration of the flow are used to analyze the correlation in traffic.

The entries defined in the behavior-based signature are extracted from the initial inquiry packages of the input packages, which are sorted by their flows. Then, the entries extracted in this manner are listed up and sorted in chronological order. This is followed by the input of the output entry list, to identify the candidate signature. The candidate signatures are identified within the threshold values which allow the maximum time interval and a maximum number of entries. From the

candidate signature extracted in this manner, only those that exceed the minimum number of hosts are finally selected as the behavior-based signatures Fig. 1.

5. Conclusion

In this study, the author conducted an empirical study on intelligent security analysis methods that use big data to protect against ever-changing, agile, and intelligent cyber-attack methods.

Recent cyber-terrorist attacks go beyond a handful of hackers showing off their skills for the fun of it, and now pursue a range of ends including threatening national security, causing social unrest, and inflicting financial damages. For this reason, it is necessary to increase our awareness of security issues and how we respond to security breaches. In the future, various cyber-terror attack analyses and studies will be conducted to mitigate damages or prevent them from happening through intelligent security technologies.

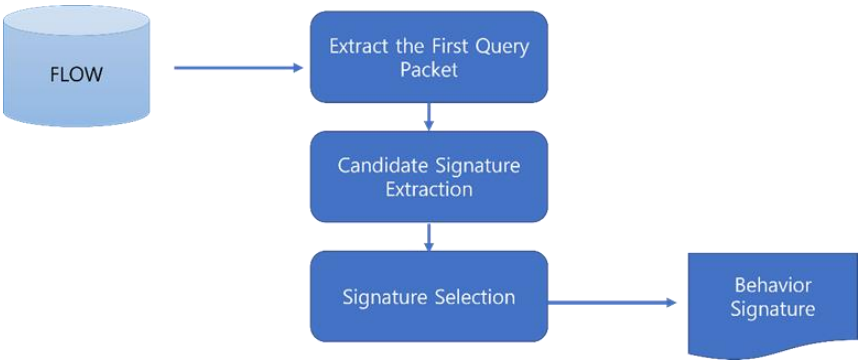

Fig. 1: Behavior-based Signature Extraction

Embracing big data analysis technologies is one approach to realizing this goal. The information gathered from logs of attack events, which can then be analyzed to find correlations, can aid in detecting and preventing attacks before they occur. Also, to respond to an APT attack, the members of the organization must launch a systematic response. To accomplish this, rather than simply combining individual security solutions with different functions, an organization should implement an advanced, intelligent security solution which can be used to prevent and respond to an APT attack. As South Korean technologies lag behind the intelligent security analysis technologies of key overseas products, it is believed that research and development in intelligent security analysis technologies are urgently needed.

References

IDC (2019). Korea Security Software 2012-2016 Forecast Update 2011 Review.

J. Feiman (2012). Hype Cycle for Application Security, Gartner Group, July 2012.

James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers (2011). Big Data: The Next Frontier for Innovation, Competition, and Productivity, Mckinsey Global Institute.

N. MacDonald (2012), Information Security Is Becoming a Big Data Analytics Problem, Gartner Research.

Sam Curry, Engin Kirda, SEddie Schwartz, William H. Stewart, Amit Yoran (2013). Big Data Fuels Intelligence-driven Security, RSA Security Brief.

M. Nicolett and K.M. Kavanagh (2021). Critical Capabilities for Security Information and Event Management, Gartner Group, May 2021.

M. Nicolett and J. Feiman (2019), SIEM Enables Enterprise Security Intelligence, Gartner Group, Jan. 2019.

General Dynamics (2010). R&D Support of DARPA Cyber Genome Program, http://publicintelligence.net/hbgary-general-dynamics-darpa-cyber-genome-program-proposal/

ANSI/ISA-TR99.00.01 (2020), Security Technologies for Manufacturing and control Systems, Instrument Society of America, 21-76.

ANSI/ISA-TR99.00.02 (2014), Integrating Electronic Security into the Manufacturing and Control System Environment, 60-69.

Defense Dept., Technical Support Working Group (TSWG) (2015). Securing Your SCADA and Industrial Control Systems, ISBN 0-16-075115-2, 12-36.

DOD (2005). Technical architecture framework information management Volume 6, Department of Defense Goal Security Architecture, Version 3.0, 3.1 - 3.3. Dominique Kilman, Jason Stamp, 2005

P.A.S. Ralston, J.H.Graham, J.L.Hieb (2017). Cyber security risk assessment for SCADA and DCS networks, *ISA Transactions* 46(4), 583-594.

Nir Kshetri (2019). Pattern of Grobal Cyber war and crime: A conceptual framework, *Journal of International Management* 11. 541-562.

Nir Kshetri (2020). Information and communications technologies, strategic asymmetry and national security, *Journal of International Management*, 11. 563-580.