International Journal of Multimedia and Ubiquitous Engineering
Vol.10, No.11 (2015), pp.105-114
http://dx.doi.org/10.14257/ijmue.2015.10.11.10

The Detection Model of Malignant Query and Personal Information Leakage based on Log Analysis

Gei-Young Kim[1], Kyung-Jin Jung[2], Yongtae Shin[3],
Sangphil Kim[4] and Jong-Bae Kim[5*]

[1,2,3]Dept. of IT Policy and Mgmt., Graduate School of Soongsil Univ., Seoul 156-743, Korea
[4, 5]Graduate School of Software, Soongsil University, Sangdo-dong, Dongjak-gu, Seoul, Korea*
[1]geiyoungkim@ligcorp.com, [2]kjjung@wise.co.kr, [3]shin@ssu.ac.kr,
[4]sangphilkim@ssu.ac.kr, [5]kjb123@ssu.ac.kr*

Abstract

Many behaviors happen in information protection control, threatening from unauthorized change, destruction, and exposure to integrity, confidentiality, and availability of database, which is the final and core object of control. Like this it approaches database through numerous paths like many applications and home pages and execute query which search, modify, and delete the data. Some of it executes normal queries, but sometimes it maliciously executes the queries for leakage of information, and gives load to database server by executing the query which uses large amount of hardware resources. Traditionally it has limits, using only to find the reason for the problems, such as malignant queries, by collecting security log. Analyzing malignant queries and personal information leakage in diversified views through multidimensional analysis of data is necessary in order to use security log in more various ways. Therefore, this treatise is going to design multidimensional analysis modeling and suggest the technology to analyze in diversified views as an application plan of existing security log so that we can detect malignant queries and personal information leakage through security log analysis. We established the standard of analysis as follows for various analyses. First, we made linkage analysis available, which we cannot know with only simple history search, through analysis of database examination history. Second, we analyze if it repeatedly approached important table for a long time through detection of abnormal pattern or long term leakage via database abnormal access analysis. Third, we understood the flow of elements and data which weigh impact on specific database assets through database impact analysis and made analysis of database assets correlation and data flow analysis available. For analysis this treatise analyzed the log collected by using OLAP tools and used experiment data and operation data in order to verify the efficiency of database security log analysis technology suggested. Also we showed that the analysis method suggested by this treatise is excellent in availability and credibility in detection of malignant queries and personal information leakage, by comparing traditional data analysis method and the analysis method suggested by this treatise.

Keywords: *Database Security, Log Analysis, Malignant Query, Personal Information, Detection*

1. Introduction

[*]Corresponding author. Tel. : +82-10-9027-3148.
Email address: kjb123@ssu.ac.kr(Jong-Bae Kim).

ISSN: 1975-0080 IJMUE

Database security is meant to protect external persons or insiders from leaking the important information assets of an individual or an organization. The treats to data-base security occur by user's mistake, misuse, and insider's abuse of his/her authority and/or attack to the known weakness of database. More and more threats occur to information assets saved and managed in database.

Since the existing database weakness analysis is initiated after accidents such personal information leakage by malicious query and system down by service overload, it is late and thus database can be exposed to an attack.

Therefore, the present study enabled to register database attack queries in Meta format and detect abnormal symptoms through multi-dimensional analysis on data-base audit history and abnormal access history in collected log files, which makes it possible to cope with potential attack to database in pre-emptive way.

2. Related Works

There are related works on this study not only studies of security log [1-2] and [3], but studies of protect personal information [4-5].

3. Detection of Malignant Query and Personal Information Leakage through Database Security Log Analysis

3.1. Types of SQL Injection Attack Queries

An intruder can use attack query to steal account information and password or create new account or password for the purpose of stealing the important assets in data-base by falsifying query internally in an abnormal way. The model proposed in the present study registers such attack queries by type and manages them in Meta format. Therefore, abnormal query can be instantly detected for judgment when security log analysis is conducted.

Table 1. Type of SQL Injection Attack Queries

Attack Type	Attack Query
Access to Table Name	Having 1=1
Access to Field Name	Group By
Access to Field Type	Union
Account Creation	Insert
Stealing Version and Configuration Information	@@Version
Account Extraction	Type Convert Error
Stealing Account Password	Union
DB Server Instance Down	Shutdown

3.2. Types of Personal Information Leaking Queries

Normally, personal information leaks out by insider's malicious intent or accidental mistake to leak database outside or by an external intruder's implantation of attack queries maliciously intended to leak out personal information. Many damages can bring out by personal information leakage: illegal use of other´s name, account stealing, voice phishing, SPAM mail, privacy risk. To prevent and minimize such damages involving with personal information leakage, the present study enables the proposed protection model to manage objects related to personal information, which is the starting point of personal information leakage, and analyze it by object.

3.3. Designing and Composing Database Security Log Analysis

3.3.1. Multi-Dimensional Model of Security Log Analysis

This study composed a multi-dimensional model that can analyze the audit history of database security log data and abnormal access to them at multidimensional angles Figure 1.

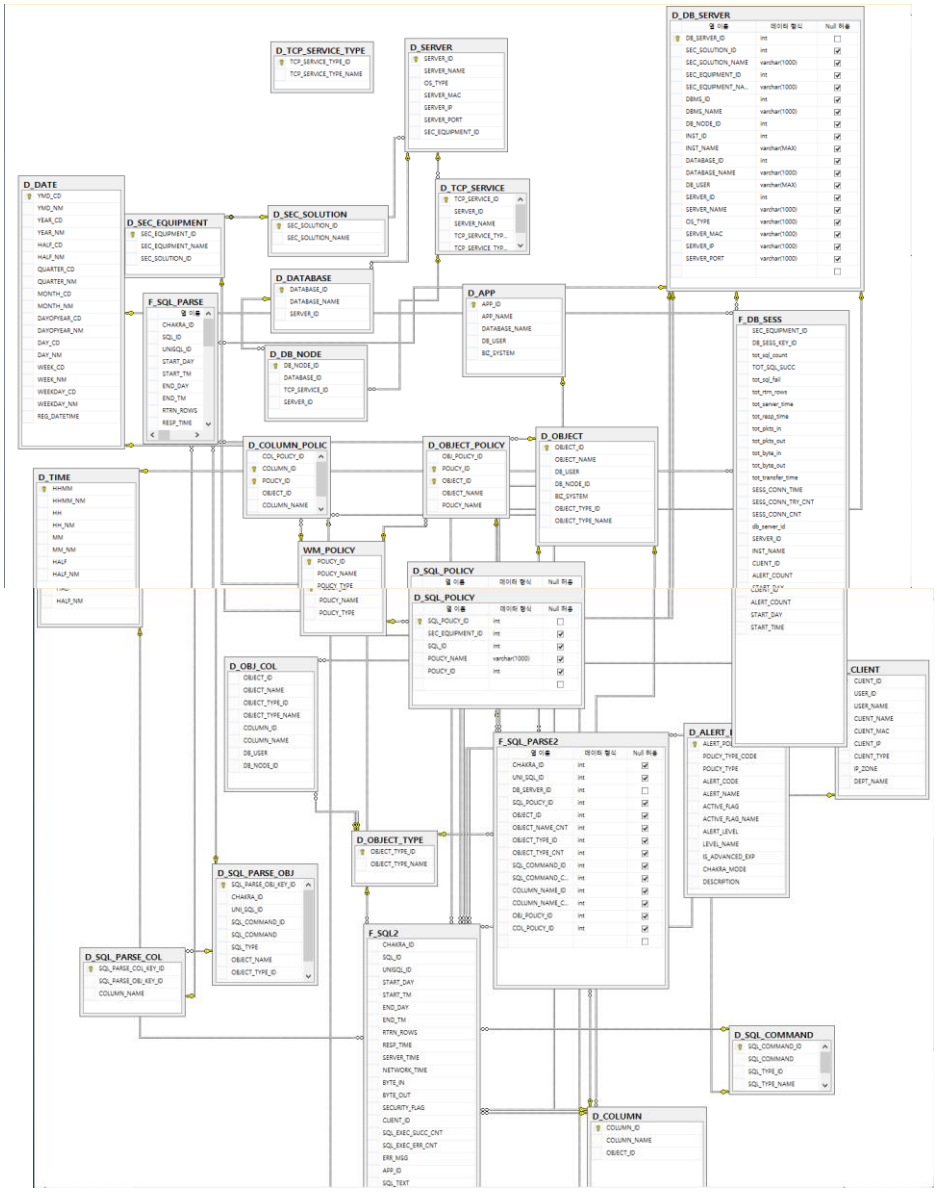

Figure 1. Multi-Dimensional Model of Security Log Analysis

This model is composed in a way to facilitate analyzing security log files by domain, such as DB server access analysis, SQL implementation analysis, SQL-based object analysis, SQL traffic analysis, and ALERT analysis. In addition, it can directly alter and create the elements necessary for analyses that can't be done in log files to improve the quality of analysis. DB server access analysis checks the current status of session and server access; identifies consistent and repeated access to long-lasting access records to a server as the prime objects to control; reversely analyzes them to know what queries are performed have been implemented to the concerned IP and user base; and checks if

malicious queries have been implemented during the uncontrolled time on DB server. SQL implementation and SQL-based object analysis can focus on important personal and security log files by object. Therefore, it can detect abnormal activities in network traffic as well as the results from query implementation Figure 2.

Subject Area	Measure	Remarks
DB Server Access Analysis	Number of times attempting access to session	Self alternation and creation
	Number of times of access to session	Self alternation and creation
	Time of access to session	Self alternation and creation
	Number of times attempting access to server	Self alternation and creation
	Number of times of access to server	Self alternation and creation
	Time of access to server	Self alternation and creation
	Uncontrolled time	Self alternation and creation
	Uncontrolled	Self alternation and creation
SQL Implementation Analysis	Number of times implementing SQL	Collecting security logs
	Number of times implementing normal SQL	Collecting security logs
	Number of times implementing error SQL	Collecting security logs
	Number of times of SQL implementation result	Collecting security logs
	Time implementing SQL (server + network)	Collecting security logs
	Time implementing SQL for server	Collecting security logs
	Time implementing SQL for network	Collecting security logs
	Time of implementing SQL for command	Self alternation and creation
SQL-Based Object Analysis	Number of times attempting use of column	Self alternation and creation
	Number of times attempting use of object	Self alternation and creation
	Number of times attempting use of instance	Self alternation and creation
	Number of times using column	Self alternation and creation
	Number of times using object	Self alternation and creation
	Number of times using instance	Self alternation and creation
	Time using column	Self alternation and creation

	Time using object	Self alternation and creation
	Time using instance	Self alternation and creation
SQL Traffic Analysis	Byte-in	Collecting security logs
	Byte-out	Collecting security logs
	Number of times of packet-in	Collecting security logs
	Number of times of packet-out	Collecting security logs
Alert Analysis	Number of times of alert	Self alternation and creation

Figure 2. Subject List of Security Log Analysis

The model in this study was designed to analyze security log files diversely by subject area. The model analyzed security log files in 7 dimensions: target to control, object, user, period, time, SQL and alert Figure 3.

Dimension	Characteristic	Reference (Class)
Target = Object to Control	Class for control	System > server > DBMS > Node > Instance > DB > Object > Column
Object Dimension	Class for control by object	Object > column
	Type of object	Table / View / SP, synonym, sequence (column dimension should not be used for the number of times using object) – meta register through transfer
	Table related to customer information	Object and column are meta-registered and automatically dimensioned.
	Column related to customer information	
	Table for security target	
	Column for security target	
User Dimension = User	User class by system type	System type > system > application > IP (Max Address)
	User class by department	Department > user (person in charge) > IP (Max Address)
	User class by work group	Work group > user > IP (Max Address)
	System type	Server, client, DB security management
	System name	
	Application	Client: it is like Todd and Orange, recognized by the name of execution file (.EXE) // Server: ERP, OLAP can't be recognized, but only access types like ETL and WAS are recognized

	IP	IP&IP band and system name, user, department mapping management
	Department	
	User name	
	Work group	
Period Dimension = Period		Yearly > half-yearly > quarterly > monthly > day > hour > minute
		Yearly > Ordinal Number of Week
		Day
Time Dimension = Time		Hour > minute
SQL Dimension =SQL	Command Class by SQL type	SQL type (DDL, PL...) > SQL command – the dimension applicable only for the measured number of times implementing SQL command
	SQL class of target to control	SQL target to control > SQL designated for control (designated control is pre-set in Shakra. If it is a target to control though not designated as target, it is done by us) – the number of times implementing SQL
	Injection SQL	It applies to the number of times implementing SQL (Inject type queries are registered and parsed and matched with SQL)
	SQL affecting performance	
Alert Dimension	Alert class	Alert type > alert name
	Alert type	
	Alert name	
	Alert status	
	Alert grade	
	Alert occurrence (hold)	

Figure 3. Dimensional Composition of Security Log Analysis

3.3.2. Meta-Management of Security Target

Target meta-management of a security target Figure 4 enables analysis by security target object Figure 5. Therefore, it can allow manage more focused and faster analysis of objects related to personal information and important information asset.

	OBJ_POLICY_ID	POLICY_ID	OBJECT_ID	OBJECT_NAME	POLICY_NAME
1	1	1	26	EMP	보안대상
2	2	1	27	AHN	보안대상

Figure 4. Metadata Registration of Security Target

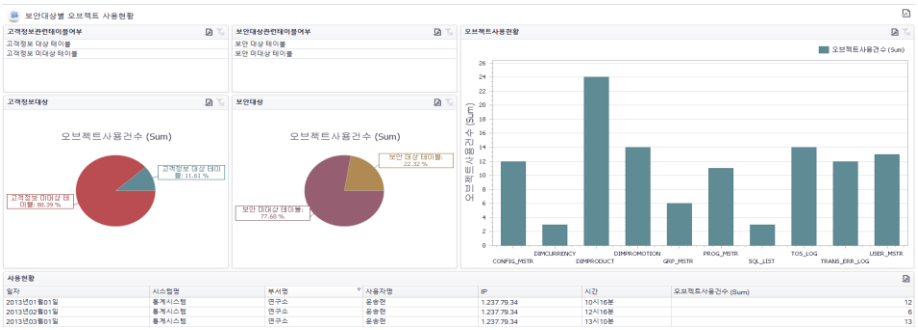

Figure 5. Current Status of Using Object by Security Object

3.4. Atypical Audit History Analysis Method

Since query text is atypical itself, it should be parsed so as to standardize it Figure 6. As a result, SQL became easy to analyze by type Figure 7.

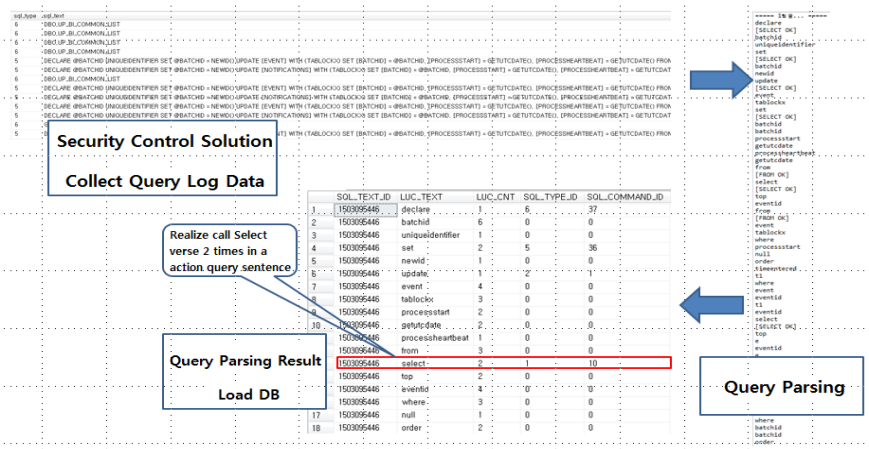

Figure 6. Query Parsing Flow

Figure 7. SQL Parsing Analysis

Because meta-data are attached, it became possible to analyze them by system name and user name, which can be intuitively recognizable Figure 8). Therefore, it can see at a glance what user performs what query in what system.

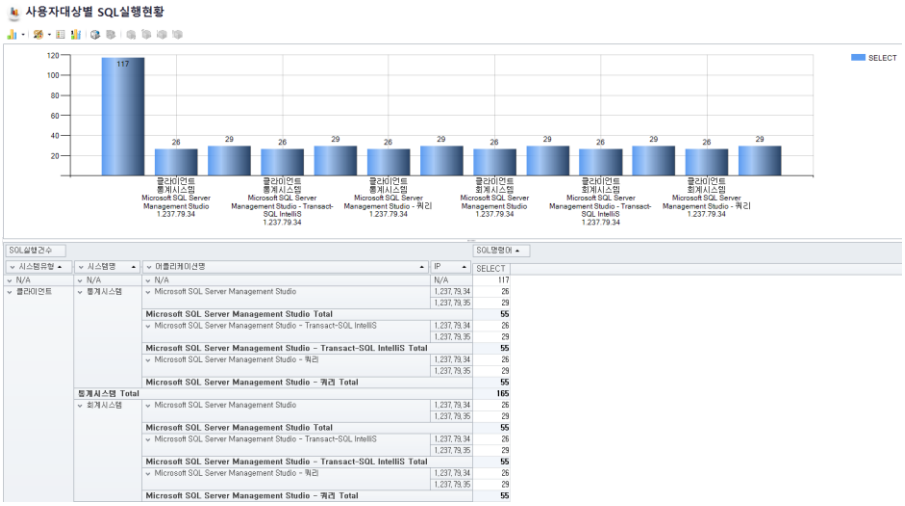

Figure 8. Current Status of SQL Implementation by User Target

3.5. Abnormal Pattern Detection Method

With patterns analyzed, it enables to detect unauthorized and abnormal access from a different band. It can identify information mapped differently by the flow of control target class to judge ‘normal’ or ‘abnormal’ access.

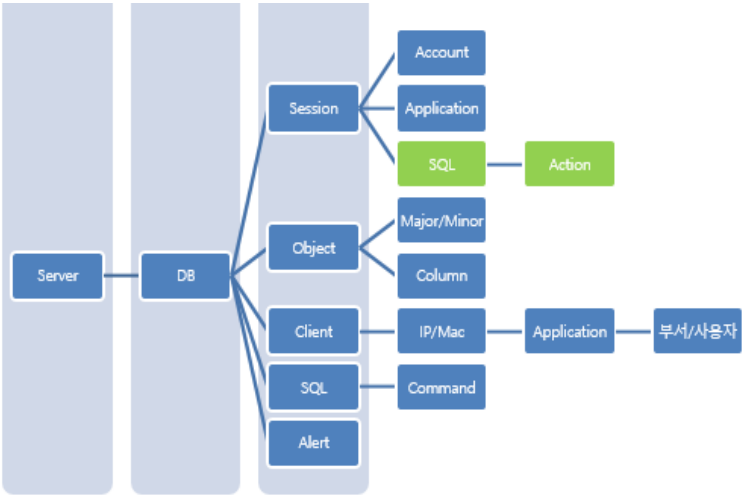

Figure 9. Control Target Hierarchy

3.6. Detection Method for Long-Term Information Leakage

Server access analysis can determine the presence of normal access (query) or long-term and repetitive access to important table.

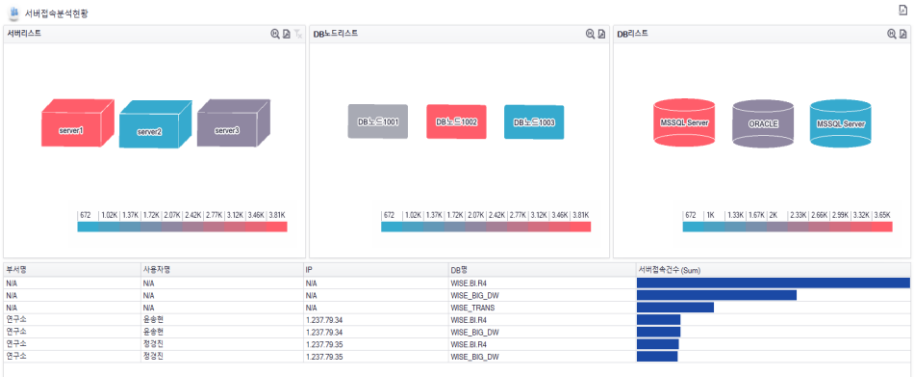

Figure 10. Current Status Analysis of Server Access

4. Conclusions

The proposed model demonstrated that it could collect database access control log data; analyze them; and detect abnormal patterns through DB audit history analysis and DB abnormal access analysis in a preemptive manner. In addition, the model was designed to handle and analyze bulky log data. Last, leakage analysis was possible: the connected analysis with other data than DB access control log data enabled to identify the factors that have an impact of the detection of malicious query and personal information leakage and this to reinforce security and manage DB assets.

References

[1] D. S. Choi, G. J. Mun, Y. M. Kim and B. N. Noh, "An Analysis of Large-Scale Security Log using MapReduce", JKIIT, vol. 9, no. 8, **(2011)**, pp.125-131.
[2] S. K. Ryeo, M. N. Shim and S. J. Lee, "The Threat Analysis and Security Guide for Private Information in Web Log", KIISC, vol. 19, no. 6, **(2009)**, pp.135-144.
[3] H. T. Chae and S. J. Lee, "Security Policy Proposals through PC Security Solution Log Analysis - Prevention Leakage of Personal Information", KIISC, vol. 24, no. 5, **(2014)**, pp.961-968.
[4] W. K. Park, "Solutions to Problems regarding Transfer of Korean Personal Information to the U.S. in the Cloud Computing Environment– With Analysis of The USA Patriot Act", Kyungpook Natl. Univ. Law Journal, vol. 38, **(2012)**, pp.455-478.
[5] S. K. Cho and M. S. Jun, "Privacy Leakage Monitoring System Design for Privacy Protection", KIISC, vol. 22, no. 1, **(2012)**, pp. 99-106.
[6] J. I. Baek and D. W. Park, "A Study on DB Security Problem Improvement of DB Masking by Security Grade", Journal of the Korea Society Computer and Information, vol. 14, no. 4, **(2009)**, pp.101-109.
[7] B. M. Lee and D. W. Park, "A Study on Intelligent Vulnerability DB Security System apply to Smart Grid", Journal of the Korea Society Computer and Information, no. 44, **(2011)**, pp.203-206.
[8] S. Choi, C. Kang and J. Cho, "Behavior analysis of entrance applicants using web log data", Journal of Korean Data & Information Science Society, vol. 20, no. 3, **(2009)**, pp.493-504.
[9] S. J. Jeong, "Legal Protection of Databases in Korea", Journal of Seoul National University Law Laboratory, vol. 37, no. 3-4, **(1996)**.
[10] T. H. Park, "Databases Encryption Policy", Journal of The Korea Society of Computer and Information, vol. 16, no. 1, **(2008)**, pp.61-72.

Authors

Gei-Young Kim, he obtained LG Masters of Business Administration (2005). He received his master's degree in health management information system from Yonsei University (2011). He is a PhD student studying Information Technology Policy & Management at the Soongsil University, while working as the head of public IT service department in LIG system. His current research focuses on topics such as MIS, Information Security, and mobile internet, and Public Information Policy.

Kyung-Jin Jung, he received his master's degree of Information Security Department from Soongsil University (2014). He is currently working as a data management and analysis company specializing in WISEiTech Research Institute senior researcher. His research interests focus on Information Security, Big Data analysis, Real-time analysis, Machine Learning.

Yongtae Shin, he is a Ph.D., professor in the School of Computer Science and Engineering, Soongsil University, Seoul, Korea. His research interests focus on Multicast, IoT, Information Security, Content Security, Mobile Internet, Next Generation Internet.

Sangphil Kim received his bachelor's degree of business administration in Yonsei University (2011). And he is studying his master's degree of software engineering in the Graduate School of Software, Soongsil University, Seoul. His current research interests include Open Source Software and Management of Technology.

Jong-Bae Kim, he received his bachelor's degree of Business Administration in University of Seoul,-Seoul(1995) and master's degree(2002), doctor's degree of Computer Science in Soongsil University, Seoul(2006). Now he is a professor in the Graduate School of Software, Soongsil University, Seoul, Korea. His research interests focus on Software Engineering, and Open Source Software.

