RESEARCH ARTICLE OPEN ACCESS

Database Forensic Analysis Using Log Files

Mr. Jitendra R Chavan*, Prof. Harmeet Kaur Khanuja**
*(Department of Computer Engineering, MMCOE, Pune University
Email: jitendra.rchavan@gmail.com)
** (Department of Computer Engineering, MMCOE, Pune University
Email: harmeetkaurkhanuja@mmcoe.edu.in)

ABSTRACT
Most organization and company maintain their business related data in database system. The misuse of data within database system may cause consequences for company. Intruders are both unauthorized user and insider privileged users. It is difficult to detect high privileged user's misuse of database within organization. Database contains the number of location where it maintain the evidence of fraudulent activity carried on database. Database forensic is the process of analyzing database and their metadata. Forensic analysis of database can help to determine intruder in system, the fraudulent activity that is carried on Database system. The goal of paper is to present survey on Database forensic and proposed framework for forensic analysis of database using various log files.

Keywords – Audit log, Database Forensic, Digital Forensic, Digital Notarization, MySql.

I. Introduction

Secure data storage is need of company or organization to do business. For many organizations, if data changed by an outsider or an inside intruder, it could cause consequences for the company. To achieve accountability and integrity organization conduct auditing through third party auditor. Due to Collusion between auditors and the companies they audit helped to introduce several federal laws e.g. Health Insurance Portability and Accountability Act (HIPAA), Sarbanes-Oxley Act to controls on electronic data. The Health Insurance Portability and Accountability Act is a federal law designed to improve the efficiency and effectiveness of the health care system [11] [12]. HIPPA act Impose restrictions on the use and disclosure of personal health information. Database audit is the process of monitoring access to and modification of selected database objects. Auditing should be ultimately enabling a separation of duty. Database systems make numerous redundant copies of sensitive data, so when data is deleted, it is not destroyed completely but it often present on disk. Forensic analysis does the process of extracting information and data from database internals like logs, data files, Meta data, view etc.

II. Digital Forensic

Digital forensic is multi-stage process which consist of following investigating process [1].

I. The first stage of the digital forensic process is the identification of relevant digital evidence. It identifies sources of digital storage component.

II. The second stage of the digital forensic is to acquire the identified evidences and to preserve it. Goal of this stage is to recover as much evidence without altering the crime scene. Standard hash functions are used to verify integrity of the evidences.

III. Goal of Evidence examination and analysis stage of digital forensic process is to extract data from the acquired evidence. Perform the analysis on copy of evidence not on original. Forensic analysis analyzed the evidence sources to determine the sequence of events that occur in the crime scene. The evidence can be examined using various forensic tools like FTK, log miner etc.

IV. Evidence documentation stage of digital forensic process makes documentation of each stage carried out during investigation. This documentation is useful in presenting the evidence in the court of law.

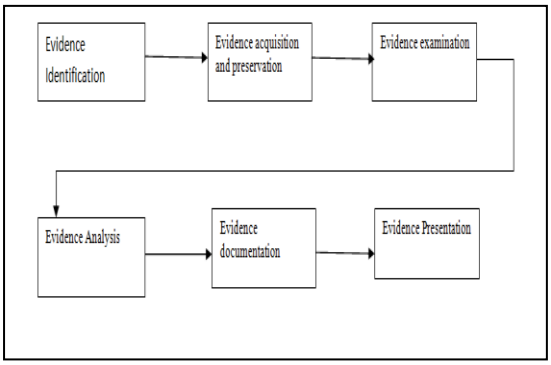

Fig. 1.Digital forensic multi-stage process

Database forensic is branch of Digital forensic. Due to complexity of database and different characteristic of database we cannot follow all Digital forensic procedure on database. Database forensic can apply some of the digital forensic technique as it is [2]. These techniques are mentioned as follows.

i. Acquisition and preservation of data in database forensics: There are three techniques to collect data from database after database has been compromised. Acquisition techniques are Live acquisition, Dead acquisition or Hybrid acquisition

ii. Collection and analysis of artifacts in database forensics: In database there are number of location where information about database is present. Identify and collect data from these locations. E.g. Plan cache, transaction log,

iii. Database forensics investigation process. Database forensic analysis normally focuses on schema, Metadata, data files in database. We can also perform database forensic using transaction log in database system or application.

III. Database Artifact for Forensic Analysis

In this paper we study MySql Database artifact for forensic analysis. MySql is open source relational database [19]. When intruder delete a table or data in database it is not completely remove from database. It is replicated in number of places in database. We can recover deleted data from those places. MySql contains number of log file from which we can identify activity carried out on database. After tamper detection we can collect evidence from various log files and data files.

3.1 Mysql log file

MySql maintain following log file:

i. Error log

ii. General query log

iii. Binary log and the binary log index file

iv. Update log

v. slow-query log

vi. The error log

This log contains a record of server startups and shutdowns, as well as messages about problems or exceptional conditions occur in database system.

i. The general query log

This log contains a record of client connections, statements received from clients. It is useful for monitoring server activity: who is connecting, from where, and what they are doing. Using general query log we can find out which statement use and on which object the operation perform while performing tampering.

Example of general query:

Mysql>SET GLOBAL general_log = 'ON';

Mysql> show tables;

Mysql> create table sample (col1 int);

Mysql> insert into sample values (3);

Mysql> commit;

Mysql> exit;

Mysql>select * from mysql.general_log;

Time Id Command Argument

130327 10:55:07 3 Connect root@localhost on

3 Query select @@version_comment limit 1

130327 10:55:26 3 Query show tables

130327 10:55:50 3 Query create table sample (col1 int)

130327 10:56:08 3 Query insert into sample values (3)

130327 10:56:13 3 Query commit

130327 13:33:41 3 Quit

ii. The binary log and the binary log index file

This log consists of one or more files that record statement that modifies the database. It contains a record of statements such as DELETE, INSERT, REPLACE, CREATE TABLE, DROP TABLE, GRANT, and REVOKE. Binary log contents are written as SQL statements encoded in binary format. The binary log files are accompanied by an index file that lists which binary log files exist on the server.

iii. The update log

The update log is similar to the binary log, but it is stored in text format and does not contain as much information.

iv. The slow-query log

This log's purpose is to help you identify statements that may be in need of being rewritten for better performance.The server maintains a long_query_time system variable that defines "slow" queries. If a query takes more than that many seconds of real time, it is considered slow and is recorded in the slow-query log. The slow-query log also can be used to log queries for which no indexes were used.

IV. Related Work

Audit log maintain the activity carried out on database. In this [3] paper they use cryptographically strong one way hash function to detect the tampering in audit log. They use transaction time table to maintain the audit log. Tamper detection in audit log is based on strong cryptographic hashing, partial result authentication codes, and off-site digital notarization service.

Forensic analysis algorithms are used to find out when and what data tampered in database. In this [4] [5] paper proposed Monochromatic, RGB, and Polychromatic forensic algorithm. They used cryptographic hash functions to detect database tampering. They calculate additional hash chains to improve forensic analysis. Once the corruption has been detected, a forensic analyzer determines the corruption region. Corruption region indicate "where" and "when" of the corruption. The Monochromatic Algorithm uses only the cumulative hash chain. It is the simplest algorithm in terms of implementation. Problem with this algorithm is it cannot differentiate postdating from backdating events separately or from a data-only corruption event. It can find only one corruption event. The RGB Algorithm extend Monochromatic algorithm by using additional three new types of chain to the original chain in the Monochromatic Algorithm. These hash chains are computed in parallel. All hash chain consists of linked sequences of hash values of individual transactions in commit order. Problem with this algorithm is it can find only two corruption event. Advantage of this algorithm is there is no false positive. Polychromatic algorithm extends red and blue with k partial red and blue chains. It also finds out only two corruption event. There is no false positive.

The tiled bitmap forensic algorithm is more efficient than previous forensic algorithm. It finds out multiple corruption events in the database. The algorithm compute candidate set over database. The candidate set contains all possible location of detected tampering. The disadvantage of algorithm is false positive [6].

Database contains number of location where information about data processing is stored. Forensic analysis can find out past activities and recover deleted data using this information [18].MySql store table information on disk in .frm format. The file has the same name as the table, with an .frm extension. Writing a program we can read and reconstruct the table structure from the .frm file [7].

In this [8] paper they have shown that Forensic analysis of InnoDB databases by using the redo logs. Privileged users cannot delete information present in redo logs. We can recover Insert, Delete and Update statements issued against a database using redo logs.

V. Framework for Forensic Analysis

Framework for forensic analysis is as shown in fig.2 is divided into two part .First part is user application and database. User perform transaction database using application.

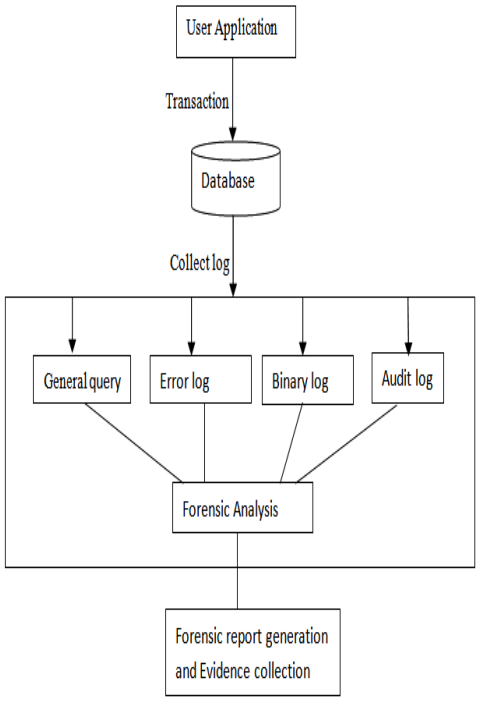

Fig. 2.Digital forensic multi-stage process

Second part is central server for forensic analysis of database. In proposed system we collect database log files in central server. Sever process each file separately and apply pattern matching technique to find out valid evidence. We can also collect the audit log to central server. Audit log maintain transaction performed by user using application. We can create transaction profile to find out fraudulent transaction. We can also create attack profile using error log to find out type of attack and unauthorized access of system.

VI. Conclusion

The Database forensic analysis is difficult due to complexity of database structure. Database contains audit logs and data file which maintain the information about activity carried out on database. There should be best practices to secure an organization's databases from internal as well as external threats. Forensic analysis performed when a crime has been detected in system. Forensic analysis finds out when, who, and from where the intruder performed illegal activity. In database there are many places where parts of the data are temporarily stored using this data we can reveal past activities, and recover deleted data. We will be to collect various log files after data tampering in database and use data mining techniques to perform the analysis.

REFERENCES

[1] Sriram Raghavan, "DIGITAL FORENSIC RESEARCH: CURRENT STATE OF THE ART" *Springer CSIT (March 2013) 1(1):91–114 DOI 10.1007/s40012-012-0008-7.*

[2] O.M. Fasan and M.S. Olivier, "ON DIMENSIONS OF RECONSTRUCTION IN DATABASE FORENSICS*" Seventh International workshop on Digital Forensics & Incident Analysis (WDFIA)2012.*

[3] R.T. Snodgrass, S.S. Yao, and C. Collberg, "TAMPER DETECTION IN AUDIT LOGS," *Proc. Int'l Conf. Very Large Databases, pp. 504-515, Sept. 2004.*

[4] K.E. Pavlou and R.T. Snodgrass, "FORENSIC ANALYSIS OF DATABASE TAMPERING," *Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 109-120, June 2006.*

[5] K.E. Pavlou and R.T. Snodgrass, "FORENSIC ANALYSIS OF DATABASE TAMPERING", *ACM Trans*

[6] Kyriacos E. Pavlou and Richard T. Snodgrass, "THE TILED BITMAP FORENSIC ANALYSIS ALGORITHM", *IEEE transaction on knowledge and data engineering, Vol. 22, pp no.590-601, April 2010.*

[7] Peter Frühwirt, Markus Huber, Martin Mulazzani, Edgar R. Weippl, "INNODB DATABASE FORENSICS" *2010 24th IEEE International Conference on Advanced Information Networking and Applications*

[8] Peter Fr¨uhwirt, Peter Kieseberg, Sebastian Schrittwieser, Markus Huber, and Edgar Weippl, "INNODB DATABASE FORENSICS:RECONSTRUCTING DATA MANIPULATION QUERIES FROM REDO LOGS" *2012 Seventh International Conference on Availability, Reliability and Security*

[9] Martin S. Olivier, "ON METADATA CONTEXT IN DATABASE FORENSICS" *Science Direct Digital investigation 5(2009) 115 – 123.*

[10] M. Malmgren, "AN INFRASTRUCTURE FOR DATABASE TAMPER DETECTION AND FORENSIC" *Univ. of Arizona, http://www.cs.arizona.edu/projects/tau/tbdb/MelindaMalmgrenThesis.pdf, 2009.*

[11] U.S. Dept. of Health & Human Services, The Health Insurance Portability and Accountability Act (HIPAA), http://www.cms.hhs.gov/HIPAAGenInfo/

[12] http://www.soxlaw.com/

[13] "MySQL 5.5 Reference Manual,"www.dev.mysql.com/doc/refman/5.5/"

[14] K.E. Pavlou and R.T. Snodgrass, "TEMPORAL IMPLICATION OF DATABASE INFORMATION ACCOUNTABILITY"*2012, IEEE*

[15] Mohammad wazid ,Avita katal ,"HACKTIVISIM TRENDS,DIGITAL FORENSIC TOOLS AND CHALLENGES: A SURVEY",*2013,IEEE.*

[16] Florian Buchholz, Eugene Spafford, "ON THE ROLE OF FILE SYSTEM METADATA IN DIGITAL FORENSICS" *Digital Investigation (2004) 1, 298e309 Elsevier.com*

[17] Kevvie Fowler, "FORENSIC ANALYSIS OF A SQL SERVER 2005 DATABASE SERVER" *April 1, 2007 GCFA Gold Certification.*

[18] Patrick Stahlberg , Gerome Miklau, and Brian Neil Levine, "THREATS TO PRIVACY IN THE FORENSIC ANALYSIS OF DATABASE SYSTEMS" *In proceedings of the 2007 ACM Sigmoid International Conference on Management of Data, pp.91–102.*