RESEARCH ARTICLE OPEN ACCESS

On The Role of Log Based Metadata in Forensic Analysis of Database Attacks

Shraddha Suratkar *, Harmeet Khanuja**
**(Depart ment of Co mputer Engineering, PUNE Un iversity, India Email:*
*** (Depart ment of Co mputer Engineering, PUNE Un iversity, India Email: harmeet*

ABSTRACT-
Fro m the past few years the use of databases has increased exponentially. Almost all o f applications in world's largest organizations use database to manage their data. But with the increasing use of database, security and privacy issues are at the peak with respect to their significance. Large nu mbers of database security breaches are occurring at a very high rate on daily basis. One can never know when the confidentiality of user is being compro mised. Because of these serious issues, it is becoming extremely important for database investigators not only to confirm the occurrence of unauthorized database access but also to generate strong evidence against criminals for presenting it in the court of law in the form of who, when, why, what, how and where did the fraudulent transaction occur So, there is an imperative need in the field of database forensics to make several redundant copies of sensitive data found in database server artifacts, audit logs, cache, table storage etc. for analysis purposes. Large volume of metadata is available in database infrastructure for investigation purposes but most of the effort lies in the retrieval and analysis of that information fro m computing syst ems. Thus, in this paper primarily relevance of metadata in design of a generalized database forensics tool independent of DBMS used is focused. The various tools of database forensics along with the challenges faced are also discussed.

Keywords - Anti-forensics attacks, Database forensics, Digital notarization, linked hash technique, Metadata, Reconnaissance attack, SQL in jection, Trail obfuscation

I. INTRODUCTION

Digital forensics [1] is a branch of science which pertains to acquire, examine, analy ze, and possibly document and present the said artifacts and the reconstructed sequence of events as evidence in front of judiciary. Until recently, traditional dig ital investigations often excluded databases even though evidence can usually be found in them. Although the field is still in its early years, it is quickly becoming an important part of many investigations due to the increased volume of information that may be helpful in solving different crimes and the large nu mber of risks associated with the informat ion stored on many databases. Of major impo rtance in database forensics which is one of the branches of digital forensics has the ability to retrace the operations performed on a database and reconstruct deleted or compro mised informat ion on the database. This requirement affects how data is collected and analyzed during the forensics analysis of a database. Although new advance database forensics tools like Idea, Arbutus etc. are co ming into existence but the hackers are also becoming equally equipped with anti-fo rensics tools to erase those digital evidences or to produce delay in the digital evidence generation process. As a result the culprit is not getting punishment within the defined interval of time. Hence there is an imperat ive need of an open source database forensics tool which will forensically analy ze the database artifacts for generating evidence against major database attacks as well as anti-forensics attacks. Also, the existing database forensics tools depend on the specific DBMS system used. Hence design of forensics tool independent of specific DBMS system can be considered as the major challenge in this field.

II. COMPARATIVE STUDY OF DATABASE FORENSICS PROCESSES

Research in dig ital forensics has lead to the development of various techniques and process models. However, many of these techniques are not completely transferable to database forensics due to certain characteristics of databases which require them to be adapted for handling database forensics.

But still different techniques were proposed by researches in this field. Wong and Edwards (2005) presented a patent method for the forensics analysis of an Oracle database. The method consists of generic steps that a forensic investigator may try to follow to discover more information about an operation that was performed on a database (Olivier, 2009). This method segments a DBMS into four abstract layers that separates various levels of DBMS metadata and data. Another methodology focused on a damaged SQL server database was presented by

Fowler (2008). This method analyzes the system's volatile and non-volatile artifacts from the database. The methodology consists of four major steps: investigation preparedness, incident verification, artifact collection and artifact analysis. Thus, the database forensics can be viewed as a multi-staged process illustrated in paper [2] "Fig.1" consisting of following steps:

i. Acquisition and preservation of data in database forensics.
ii. Collection and analysis of artifacts in database forensics and
iii. Database forensics investigation process.

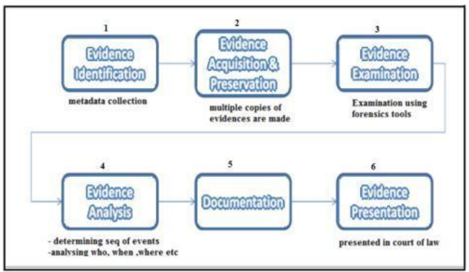

1
Evidence Identification
metadata collection
2
Evidence Acquisition & Preservation
multiple copies of evidences are made
3
Evidence Examination
Examination using forensics tools
4
Evidence Analysis
- determining seq of events
-analysing who, when ,where etc
5
Documentation
6
Evidence Presentation
presented in court of law

Figure 1. a formal process of database forensics.

Hence, the comparative table (TABLE 1) for the different methods available for carrying out the steps of database forensics can be shown as follows:

Table 1. Methods available in Database Forensics.

STEP NO.	STEPS IN DATABASE FORENSICS .	METHODS AVAILABLE
1)	Data Acquisition and Preservation	a) Live Data Acquisition. b) Dead Data Acquisition. c) Hybrid Data Acquisition.
2)	Collection and Analysis of Artifacts	a) From Transaction logs. b) Execution Plan Cache. c) Database log Files and Data Files. d) Web server logs. e) System Event logs of OS. f) Trace Files.
3)	Database Forensics Investigation Process	a) Oracle Log Miner in Oracle. b) SQL Trace in SQL. c) Olivier's (2009) method of segmenting a DBMS into four abstract layers that separates various levels of DBMS metadata and data. d) Fowler's (2008) methodology based on analyzing the system's volatile and non-volatile artifacts from the database.

III. CHALLENGES IN DATABASE FORENSICS

Database forensics is a field with multiple challenges due to lack of research on many aspects as revealed by the literature survey, which can be listed follows:

- For creating the metadata, there are different sources for collecting data from independent databases. First, is how to determine whether a database has been compromised, damaged, modified. Due to the multi-dimensional nature of database it is very difficult to predict the starting point of investigation. There is no heuristic on where to start an investigation in this case [4] .

 Another challenge often encountered in database forensics is in the large volu me of data that can be collected from a database. An investigator must determine wh ich data are pertinent to an investigation in order to reduce the size of metadata. Also, due to the number of different file formats in various databases, the process of elimination may prove a challenge as it may result in loss of some valuable data [2].

 In case of deleted data, the data recovery may prove cu mbersome due to the complexity of the database [5].

 Due to the technological advancements, the increasing volume of the digital evidences on daily basis can be the greatest challenge. It is very difficult to deal with terabytes of digital evidence [4].

 There are mult iple sources of digital evidences. One of the ways to manage the volume challenge would be to recognize this correlation across multip le digital evidence sources and automatically associate such evidence items which could lead to a reduction in the number of items for individual analysis [2].

 The literature recognizes the need for a comprehensive analysis framework which can adopt and support interpretation of a variety of dig ital evidence sources [1].

 Also, the integrity maintenance of generated metadata and digital evidence is a challenging task [12].

There is an abundance of metadata in today's digital systems and literature recognizes its value to digital forensics, particularly with regard to event reconstruction. Besides, metadata transcends data and formats and hence can bridge the diversity challenge between different DBMS's naturally. Hence sequencing these events across multip le diverse sources can be very valuable during an investigation.. Hence, we should design a system such that it overcomes all the specified drawbacks.

IV. PROPOSED SYSTEM ARCHITECTURE

4.1 PROPOSED ALGORITHM

The database forensics investigation process should in general include the following steps "Fig.2"

 Successful user authentication.

 User's query is then passed through the metadata file server.

 Monitor the transactions for any suspicious activity using pattern matching by inference rules.

 If the inco ming transaction matches the pattern i.e. any suspicious activity detected then collect volatile and non-volatile artifacts fro m the database for activity reconstruction and analysis else stop.

 Generate the evidence in form of who, when, why, where, how and what was done to database.

 Preserving and analyzing the generated evidence.

 Recover the log information in database using metadata file.

 Stop

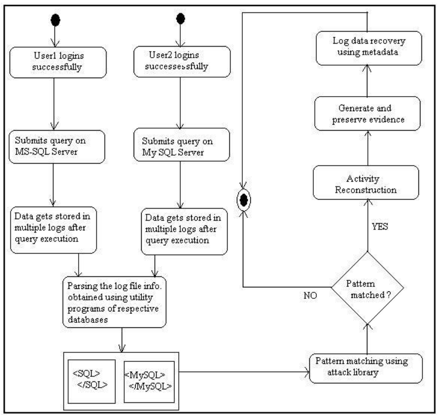

User1 logins successfully
User2 logins successessfully
Submits query on MS-SQL Server
Submits query on My SQL Server
Data gets stored in multiple logs after query execution
Data gets stored in multiple logs after query execution
Parsing the log file info. obtained using utility programs of respective databases
<SQL> </SQL>
<MySQL> </MySQL>
Pattern matching using attack library
Pattern matched ?
NO
YES
Activity Reconstruction
Generate and preserve evidence
Log data recovery using metadata

Figure 2. proposed algorithm for the database forensics process.

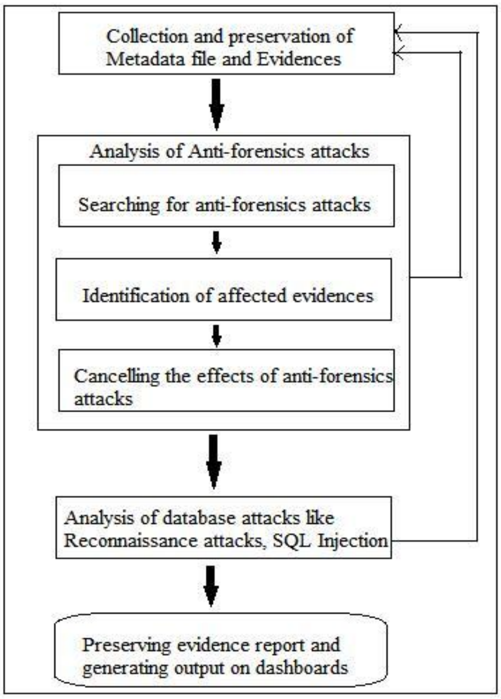

Collection and preservation of Metadata file and Evidences
Analysis of Anti-forensics attacks
Searching for anti-forensics attacks
Identification of affected evidences
Cancelling the effects of anti-forensics attacks
Analysis of database attacks like Reconnaissance attacks, SQL Injection
Preserving evidence report and generating output on dashboards

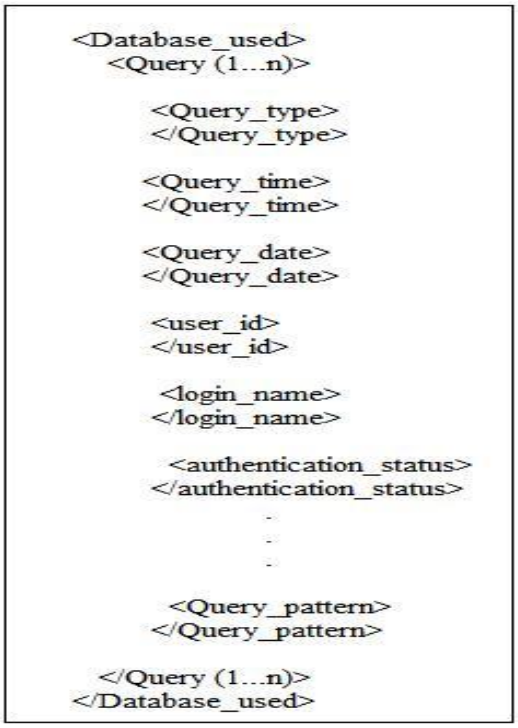
Figure 3. overall process of database forensics investigation.

4.2 A Fram Ework For Proposed System

A framework for the proposed algorithm as shown in “Fig. 2” and “Fig. 3” is carried out in two stages as indicated in “Fig. 5” and “Fig. 6” .

Stage 1(fig. 5):

In the first stage detailed mu ltip le logs of SQL, MySQL and operating system are made availab le at verification place which are used for attribution. Log files like database files, transaction logs, cache files fro m SQL, log files, text files, binary logs from MySQL and SQL server error logs “Fig.7”, memory dumps, trace files “Fig.8” and system event logs “Fig.9” fro m operating system are to be used. The logs contain useful informat ion it also contains routine operational data which may not be required at the time of analysis . Ext racting the useful informat ion from logs which are needed for analysis is a challenging task. These files of database are then parsed to give the relevant information as per the condition laid by the investigator. The expected outcome of the script is the metadata of logs “Fig. 4” having traces of actions from the mult iple files which helps to predict the identification and activities of unauthorized events carried out.



```
<Database_used>
   <Query (1...n)>

      <Query_type>
      </Query_type>

      <Query_time>
      </Query_time>

      <Query_date>
      </Query_date>

      <user_id>
      </user_id>

       <login_name>
      </login_name>

        <authentication_status>
      </authentication_status>
                .
                .
                .

        <Query_pattern>
      </Query_pattern>

   </Query (1...n)>
</Database_used>
```

Figure 4. XML schema of metadata file.

Then based on the Inference rules laid down using expert knowledge the decisions are taken fro m the stored informat ion in the derived metadata to get the relevant and filtered information for analysis. This is achieved by matching similar patterns and behavior of the system concluded fro m metadata. A pre-final log analysis report is generated at this stage.

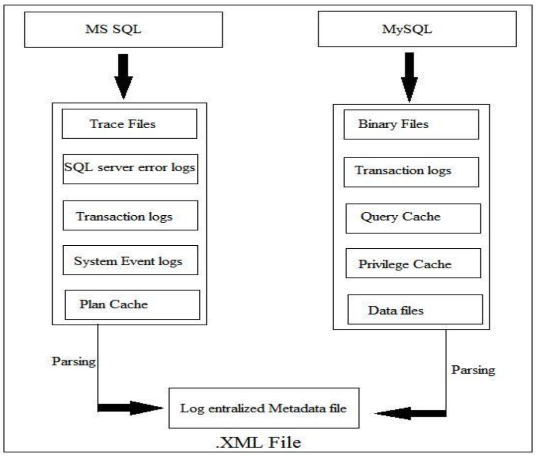


Figure 5. system architecture1 for forensics investigation.

Stage 2(fig. 6):

In the second stage, the reconstruction of events and activities fro m the collected volatile and nonvolatile artifacts is done for validation against the previously generated log analysis report to give a final forensic report. At this stage different database attacks like Sql in jection, brute force, buffer overflow, reconnaissance attacks will be detected on one side along with majo r anti-fo rensics attacks like Trail obfuscation, artifact wiping on the other. Database log informat ion will be recovered in case it is modified by any of the attacks. Then the evidence generation and preservation using linked hashed and digital notarization service will be carried out. The evidence will consist of who, why, when, what, how and where the malicious transaction were carried out.

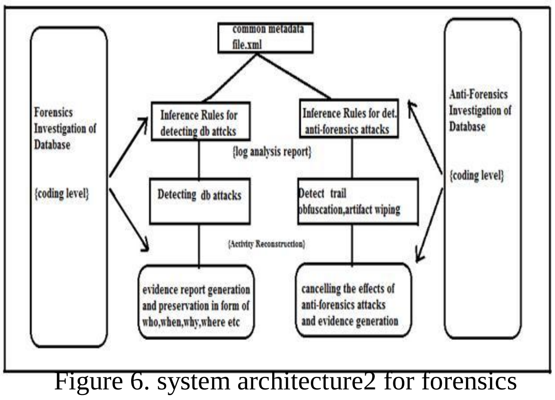

common metadata file.xml
Forensics Investigation of Database
{coding level}
Inference Rules for detecting db attcks
Inference Rules for det. anti-forensics attacks
{log analysis report}
Anti-Forensics Investigation of Database
{coding level}
Detecting db attacks
Detect trail obfuscation,artifact wiping
{Activity Reconstruction}
evidence report generation and preservation in form of who,when,why,where etc
cancelling the effects of anti-forensics attacks and evidence generation

Figure 6. system architecture2 for forensics investigation.

4.3 STATISTICS OF THE PROJECTED SYSTEM

System S can be defined as:

S = {Input, Output, intermed iate states, Inference rules}

- For genuine transaction as input, suppose output = 1, then

$$\boxed{O_i = \overset{n}{\underset{i=0}{\forall}} Q_i = 1}$$

- For non-genuine transaction as input, suppose output = 0, then

$$\boxed{O_i = \overset{n}{\underset{i=0}{\forall}} Q_i = 1}$$

- Where O: output of the system.

Q: set of queries given by user.

V. ADVANTAGES

The logs play a very important role in the field of database forensics. Whenever certain updates are done to database, they are stored in stable storage in a log file . So me of the major advantages of database forensics in the digital world are listed below:

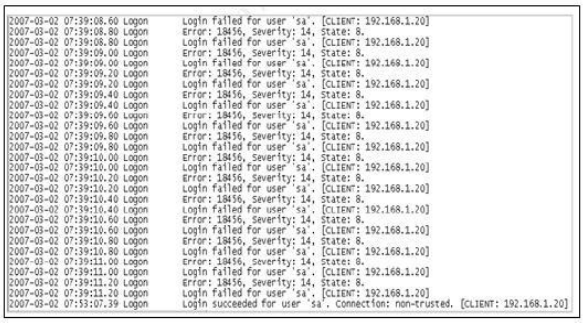
- Auditing is the ability to see what actions happened on the system and who performed those actions. Thus, auditing is the most important feature of database forensics which allo ws us to trace who, what, where, why, how and when did the malicious transaction take place.
- Also, database forensics provides an investigator with large amount of metadata fro m mu ltiple digital sources for analysis.
- With the help of logs an investigator can reconstruct the sequence of events and activities the user or system has carried out for validation against the previously generated log analysis report to give a final forensic report or ev idence file.
- Along with detecting major database and anti-forensics attacks, we can also generate an evidence to be presented in the court of law against the criminal for seeking him liable punishment for h is offence.
- A Database admin istrator (DBA) is the one who monitors and maintains database in an organized way. But database forensics allo ws us monitoring the database along with tracing the actions of a DBA if he misuses his access privileges.
- Thus, with the help of database forensics a large nu mber of co mputer security breaches can be identified and thus be prevented. Finally, there is no way for an attacker to escape without being punished.

VI. APPLICATIONS

- Database forensics helps in identifying and preventing cybercrimes on internet.
- In case of live analysis, fraud monetary transactions can be detected and blocked.
- In case, the database data is deleted by the criminal, it can be recovered with the help of metadata.

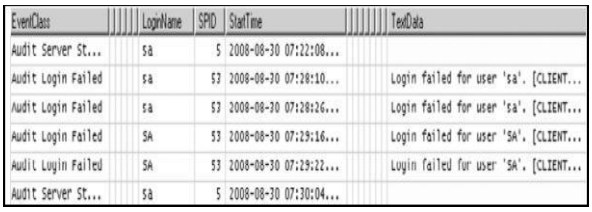
VII. RELEVANCE OF METADATA FOR FORENSICS

The advantages of the database stem fro m the fact that they contain metadata about the data they store within a source of digital evidence. Metadata can be defined at many levels but a particular type of metadata provides certain context informat ion that enables easy handling and management of the data contained and is hence very informative. Such metadata often describes the data so that programs that use the database are independent of the physical representation of the data. Clearly databases often contain informat ion that may be useful during many forensic investigations. This will be the case where a victim’s database may contain informat ion useful to the investigation, but also where a suspect has used a database to facilitate his or her suspicious acts. For forensics investigation, the main investigations a forensic investigator looks for are: who, what, when, how, where and why did the corruption events occur in the system. Most of the current work explains how to recover data fro m a system in one form or the other. If it were possible to record a system’s state register values, memo ry, t imers, and network events, interrupt informat ion, etc. -- for every single clock step, one could use that information to deterministically replay all events that took place on the system. One can also examine the forensics hard disk images, memory du mps, network captures and logs for creating metadata file for analysis. Thus, metadata helps in auditing the system and tracing actions to specific user “Fig.7” and “Fig.8” .



```
2007-03-02 07:39:08.60 Logon        Login failed for user 'sa'. [CLIENT: 192.168.1.20]
2007-03-02 07:39:08.80 Logon        Error: 18456, Severity: 14, State: 8.
2007-03-02 07:39:08.80 Logon        Login failed for user 'sa'. [CLIENT: 192.168.1.20]
2007-03-02 07:39:09.00 Logon        Error: 18456, Severity: 14, State: 8.
2007-03-02 07:39:09.00 Logon        Login failed for user 'sa'. [CLIENT: 192.168.1.20]
2007-03-02 07:39:09.20 Logon        Error: 18456, Severity: 14, State: 8.
2007-03-02 07:39:09.20 Logon        Login failed for user 'sa'. [CLIENT: 192.168.1.20]
2007-03-02 07:39:09.40 Logon        Error: 18456, Severity: 14, State: 8.
2007-03-02 07:39:09.40 Logon        Login failed for user 'sa'. [CLIENT: 192.168.1.20]
2007-03-02 07:39:09.60 Logon        Error: 18456, Severity: 14, State: 8.
2007-03-02 07:39:09.60 Logon        Login failed for user 'sa'. [CLIENT: 192.168.1.20]
2007-03-02 07:39:09.80 Logon        Error: 18456, Severity: 14, State: 8.
2007-03-02 07:39:09.80 Logon        Login failed for user 'sa'. [CLIENT: 192.168.1.20]
2007-03-02 07:39:10.00 Logon        Error: 18456, Severity: 14, State: 8.
2007-03-02 07:39:10.00 Logon        Login failed for user 'sa'. [CLIENT: 192.168.1.20]
2007-03-02 07:39:10.20 Logon        Error: 18456, Severity: 14, State: 8.
2007-03-02 07:39:10.20 Logon        Login failed for user 'sa'. [CLIENT: 192.168.1.20]
2007-03-02 07:39:10.40 Logon        Error: 18456, Severity: 14, State: 8.
2007-03-02 07:39:10.40 Logon        Login failed for user 'sa'. [CLIENT: 192.168.1.20]
2007-03-02 07:39:10.60 Logon        Error: 18456, Severity: 14, State: 8.
2007-03-02 07:39:10.60 Logon        Login failed for user 'sa'. [CLIENT: 192.168.1.20]
2007-03-02 07:39:10.80 Logon        Error: 18456, Severity: 14, State: 8.
2007-03-02 07:39:10.80 Logon        Login failed for user 'sa'. [CLIENT: 192.168.1.20]
2007-03-02 07:39:11.00 Logon        Error: 18456, Severity: 14, State: 8.
2007-03-02 07:39:11.00 Logon        Login failed for user 'sa'. [CLIENT: 192.168.1.20]
2007-03-02 07:39:11.20 Logon        Error: 18456, Severity: 14, State: 8.
2007-03-02 07:39:11.20 Logon        Login failed for user 'sa'. [CLIENT: 192.168.1.20]
2007-03-02 07:53:07.39 Logon        Login succeeded for user 'sa'. Connection: non-trusted. [CLIENT: 192.168.1.20]
```

Figure 7. the error logs report for analyzing brute force attacks [15].



EventClass	LoginName	SPID	StartTime	TextData
Audit Server St...	sa	5	2008-08-30 07:22:08...	
Audit Login Failed	sa	53	2008-08-30 07:28:10...	Login failed for user 'sa'. [CLIENT...
Audit Login Failed	sa	53	2008-08-30 07:28:26...	Login failed for user 'sa'. [CLIENT...
Audit Login Failed	SA	53	2008-08-30 07:29:16...	Login failed for user 'SA'. [CLIENT...
Audit Login Failed	SA	53	2008-08-30 07:29:22...	Login failed for user 'SA'. [CLIENT...
Audit Server St...	sa	5	2008-08-30 07:30:04...	

Figure 8. failed login attempt captured within a default trace file [15].

Metadata can also be used to replay all events that took place on the system. Most of the current work explains how to recover deleted data fro m a system in one form or the other. It also helps to draw conclusions about events on the system fro m the data “Fig.9”. But the exact strategy for auditing will be different for different systems. They depend upon the policies in place for the system. Record ing everything we have mentioned on every system is not realistic. However, policies in some organizat ions may require recording a large portion of it. These may range fro m h igh-security computing systems, where even the access of certain files should be documented in its most complete form (who, where, when, how?) to home co mputers where maybe only the question of where certain files came fro m matters. Hence it depends on investigator what to consider for analysis.

VIII. CONCLUS ION

Literature survey reveals many challenges in the field of database forensics. The current database forensics tools are not capable of using all the relevant metadata fro m database for forensics purposes. So, the database forensics tool should be designed in such a way such that it will ext ract and analyze all the prioritized metadata sources in databases for capturing evidences by detecting major databases attacks and reconstructing the databases in case of anti-forensics attacks independent of the DBMS system used. Thus, informat ion accountability and integrity should be maintained in the context of relational databases by auditing all important artifacts fro m the databases. Also, this survey recognizes the need for a co mprehensive analysis framework wh ich can adopt and support interpretation of a variety of digital evidence sources. There is an abundance of metadata in today’s digital systems and survey recognizes its value to database forensics, particularly with regard to event reconstruction. Sequencing all events across multiple diverse digital evidence sources can also provide an investigator an overall view of all the events across all d igital evidence sources which can be very valuable during an investigation. Thus, we should attempt to integrate mu ltip le forensic and other analysis tools, primarily within a single framework to achieve this task. Thus in brief this survey aims to maintain the overall informat ion accountability in context of relational databases by auditing all logs and cache informat ion independent of the DBMS used.

REFERENCES

Journal Papers:

[1] Slim Rekhis and Noureddine Boudriga, A system for formal digital forensic investigation aware of anti-forensic attacks, IEEE transactions on Information Forensics and Security, vol. 7, no. 2 April 2012.

[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] 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.

[4] Martin S. Olivier, On metadata context in database forensics, Science Direct Digital investigation 5(2009) 115 – 123.

[5] Ali Reza Arasteh, Mourad Debbabi, Assaad Sakha, Mohamed Saleh, Analyzing mult iple logs for forensic evidence, Science Direct Digital investigation4S(2000)s82 – s91.

[6] Nitin Agrawal, William J. Bolosky, John R. Douceur, and Jacob R. Lorch, A five-year study of file-system metadata, ACM Trans Storage 3(3):9:1-9:32 2007.

[7] Florian Buchholz, Eugene Spafford, On the role of file system metadata in digital forensics, Digital Investigation (2004) 1, 298e309 Elsevier.com

[8] Harmeet Kaur Khanuja, D.S. Adane, Database security threats and challenges in database forensic: a survey, 2011 International Conference on Advancements in Information technology with workshop of ICBMG 2011, Singapore IPCSIT vol.20 (2011).

[9] Harmeet Kaur Khanuja, D.S. Adane, A framework for database forensic analysis, Computer Science & Engineering: an International Journal (CSEIJ), vol.2, no.3, June 2012 .

[10] Kevvie Fowler, Forensic analysis of a sql server 2005 database server" April 1, 2007 GCFA Gold Certification.

[11] Kyriacos e. Pavlou and Richard T. Snodgrass, Temporal implications of database information accountability, 19th International Symposium on Temporal Representation and Reasoning 2012.

[12] Kailash Kumar, Sanjeev Sofat, S.K.Jain, Naveen Aggarwal, Significance of hash value generation in digital forensic: a case study, International Journal of Engineering Research and Development e-issn: 2278-067x, p-issn: 2278-800x, www.ijerd.com volume 2, issue 5 (July 2012), pp. 64 -70.

[13] Ryan Harris, Arriving at an anti-forensics consensus: examining how to define and control the anti-forensics problem, Science Direct Digital investigation 3 s (2006) s 44 – s49.

[14] 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.

Books:

[15] Kevvie Fowler, Sql Server Forensic Analysis (ISBN: 9780321533203, 2009)