International Journal of Advance Research in Science and Engineering
Vol. No.4, Special Issue (01), Spetember 2015
www.ijarse.com
IJARSE
ISSN 2319 - 8354

NOVEL ARCHITECTURE FOR DATABASE FORENSIC ANALYSIS USING METADATA

Mrs. Reena R. Chaudhari[1], Dr. Jagdish.W.Bakal[2]

[1,2]*Department of Computer Engineering, ARMIET, Shahapur, Thane (E).(India)*

ABSTRACT

With the Internet being widely used for commercial transactions, cybercrime has also increased. Since not all digital crimes can be prevented, using only a proactive approach to security is insufficient and digital forensics has become an important field of study. Databases hold organization's most sensitive digital information and Database Forensics is a very new field with little literature. Today's forensic tools are challenged by databases because of the huge amount of data that they have to not only preserve, but also inspect after it is preserved. Meanwhile, currently available forensic tools tend to present the information of specific database only without providing any analysis or interpretation functionality. Thus, an open source forensic tool independent of DBMS is recommended. But little research has been done. Moreover, tools alter the database in a significant way. Although it may be acceptable that the evidence has been altered in some way. Hence, the significant use of forensic tool independent of DBMS is applied on multiple databases. Moreover, metadata files generated from databases plays an important role in database forensic analysis and analyze what kind of information is desirable for forensic investigations. With the help of metadata the malicious activities of DBA can be easily traced, as well as database attacks can be effectively detected and blocked. The evidences in the form of W's are determined to present in judiciary and data recovery is done in case of data wiping. There is a huge amount of literature and resources for Database Backup, Restore and Recovery. Thus, a framework for database forensic analysis is proposed.

Keywords*: *Database Forensics, Metadata, Linked Hash Technique, DOS, SQL Injection, Brute Force, Anti-Forensics Attacks.

I. INTRODUCTION

Forensics is a science dedicated to the methodical gathering and analysis of evidence to establish facts. Digital Forensics involves the investigation of digital devices recovered from the crime scene. It includes collection, preserving and analyzing the digital devices and presenting the evidence in the court of law. This Digital forensics has a wide variety of branches of forensics depending upon the different digital devices used as a means of committing a particular crime.

With the rapid development in the area of internet, technology and communication, the use of internet in day-today activities like online payment e-banking, e-shopping, etc. also increase in a vibrant manner. So the invent of internet have made our routine much easier. But at the same time the usage of internet leads to the problems. Many fraudulent can occur by use of internet. Moreover, many database attacks are increasing in the world of

internet. Database plays a key role in forensic analysis. Analyzing large amount of data that is needed for investigation is a major challenge while carrying out database forensics. In order to increase the availability of data, several redundant copies of data will be available in different logs such as audit logs, server cache, server artifacts etc. which increase the amount of information in which forensic investigation is to be done.

Forensic investigation is normally carried out after the crime or intrusion has been occurred. But in most of the cyber-crimes, criminals are not punished due to the lack of evidences. So, by using metadata of database the information regarding crimes like who does the crime, when it occurs, how it happened etc. is retrieved and it is produced in the form of evidences. The database attacks are detected and blocked along with tracing the malicious activity of DBA by forensic investigator.

II. FORMAL PROCESS OF DATABASE FORENSICS

Digital forensics research has directed to the development of various techniques and process models. However, due to certain characteristics of databases, many of these techniques are not completely transferable to database forensics which requires them to be adapted for handling database forensics. Thus, the database forensics as multi-staged process [2][3] is illustrated as follow.

2.1 Data Acquisition and Preservation in Database Forensics

The acquisition of data from a modified database that has not been damaged or compromised is often done by simply querying the database. However, when investigating a database that has been compromised and/or damaged, there are three data collection methods that can be used: Live acquisition, Dead acquisition or Hybrid acquisition (Fowler, 2008). As with digital forensics in general, a live data acquisition occurs when the system being analyzed is still running while the analysis is being performed. The dead acquisition method involves copying of data from the system being investigated without using the system itself while the hybrid data acquisition method combines the key elements of both live and dead acquisition methods. Regardless of the method used, it is important to ensure that digital evidence is preserved and data is not unintentionally altered or destroyed.

2.2 Collection and Analysis of Artifacts in Database Forensics

Forensic data often exists in several places on a database. It is important to know which data are important and prioritize evidence collection due to the volatility of some data. Apart from the data which can be collected from a modified database by executing queries, data can also be found in Transaction logs, Execution plan cache, Database log files and Data files, Trace files and tools normally used by database systems.

The analysis stage should take into account the dimensions involved in any particular investigation and where related information can be found. Another important criterion of the analysis phase of database forensics, is that previously deleted of modified data should be recovered and the actions performed by an intruder must be determined, particularly when investigating compromised or damages databases.

2.3 Database Forensics Investigation Process

Currently, commercial database forensics tools such as Oracle log miner, SQL trace are available for database forensics processes which are dependent on specific DBMS. Wong and Edwards (2005) presented a patent method for the forensics analysis of an Oracle database. 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. But, all of them have their own shortcomings. Hence, a system is designed such that it overcomes all the specified challenges fulfilling the given objectives.

III. CHALLENGES IN DATABASE FORENSICS

The related work concludes that many aspects of database forensics, which can be listed as the different challenges involved in database forensics investigations, are insufficient[1][2][3][4][5][9][10]. Database forensic is an important area in the field of digital forensic. The fact that the technologies used in digital forensic cannot be copied while conducting database forensics. Beginning from the metadata, there are different resources for collecting data from independent databases. First, is how one can knew whether modifications occurred or not in the database and if knew, then how to overcome it and also due to multi-dimensional nature of database it is challenging to guess the starting point of investigation. There is no exploratory on where to start an investigation in this case. Another challenge often encountered in database forensics is analyzing large amount of data which is a tiresome job and is major challenging for various areas like Engineering, Medicine, attack detection, etc. In order to decrease the size of metadata, an investigator must determine which data are relevant to an investigation. Also, process of elimination may prove to be a challenge as some valuable data might be damaged due to the number of different file formats used in databases. The data recovery may prove cumbersome due to the complexity of the database, in case of deleted data. Due to the scientific advancements, the increasing capacity of the digital evidences on daily basis can be the big challenge as there are multiple sources of digital evidences. Thus, there is a need for a broad analysis framework which can approve and support a variety of digital evidence sources. Metadata is significant in case of event reconstruction. Hence sequencing these events across multiple diverse sources can be very valuable during an investigation. Also, the integrity of generated metadata and digital evidence is a challenging task.

IV. EXISTING SYSTEM

The database forensics tools are mostly commercial and depend upon the specific DBMS used. Also, none of them have the ability to analyze all the digital evidence sources from databases at once. This reduces their efficiency in auditing the database with all respects. It is very cumbersome for them to detect all the 5W's of corruption event simultaneously. Hence an open source, generalized database forensics tool which will be independent of the specific DBMS has been recommended.

4.1 Limitations

1. Recovery of data in case, the data is modified/ deleted by DBA.

2. No crosschecking of sensitive data
3. Integrity of metadata and digital evidence is not preserved.

V. PROPOSED SYSTEM

The framework of the system as shown in fig.1 is carried out in two stages as indicated in fig.3 and fig.4:

The database forensics investigation process should in general include the following steps (fig.1):

1. Determining whether a database has been compromised or modified.
2. If yes, reconstructing the database.
3. Then, collection of volatile and non-volatile artifacts
4. Preservation and authentication of metadata.
5. Forensics Analysis of metadata for detecting corruption events in database i.e. who, when, where, how and what was done to database.
6. Preserving and analyzing the generated evidence.
7. End.

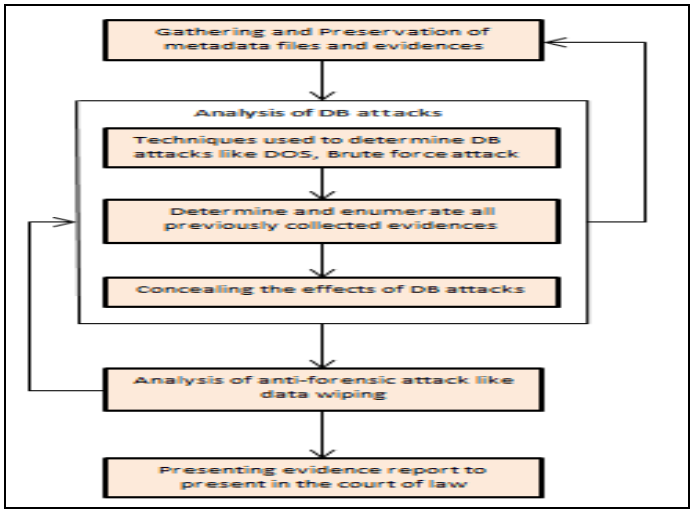

Figure 1. System Architecture for Database Forensic Analysis

5.1 Metadata File

In a database, each and every action performed on it are recorded in different logs like audit log, cache, trace files, web server logs etc. (fig 2). It also contains metadata [11] that describes the data residing in these logs. So the term metadata refers to the "data about the data". The metadata is retrieved from the log files [13] using the corresponding utility program of the DBMS used and is represented as an XML file which is then used for detecting database as well as anti-forensic attacks

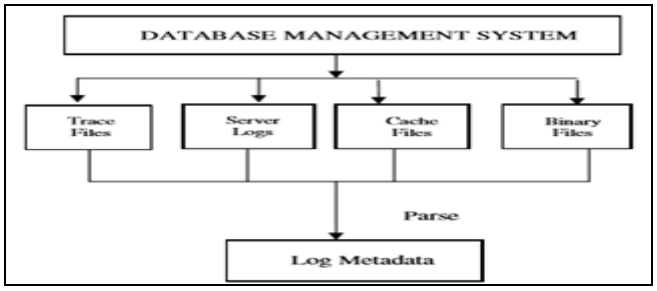

Figure 2. Log Metadata File.

5.2 Proposed Algorithm

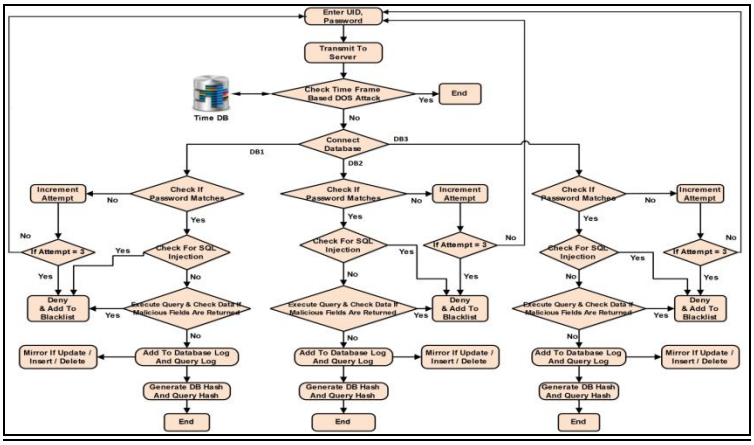

Figure 3. Client-Server Module.

1. User authentication.
2. Authentication request to server.
3. Check timeframe based DOS attack, if yes block IP address and end the service.
4. Connects to databases and detect database attacks like brute force &SQL injection attack. If any attack detected, then deny the service and add to blacklist database else execute the query and check for sensitive data is returned, if yes then deny the service and add to blacklist database else go to step.
5. Maintain Database and Query log and generate Mirror database in case of insert, delete and update of database.
6. Generate Database Hash and Query hash.

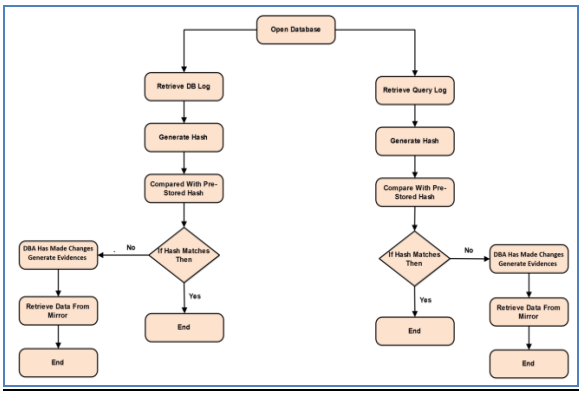

Figure 4. Forensic Module.

1. Apply forensic check.
2. Retrieve Database and Query log and generate hash.
3. Compare with pre stored hash, if hash doesn't match then go to Step 4, else Step 6.
4. DBA has made the changes in the database; generate the evidences to present in the court of law.
5. Retrieve data from mirror database in case data found has been changed
6. End.

VI. APPLICATIONS

1. Generated metadata file of the DBMS system used helps us to generate the digital evidence or proof against criminals.
2. This digital evidence can be presented in front of judiciary so that the suspect gets punished for his crime in definite time interval.
3. Also, the violation of different database security threats can be overcome through database forensics.
4. Major database attacks like SQL injection, brute force and DOS attacks can be detected.
5. Database forensics helps in identifying and blocking and Detection cybercrimes on internet.
6. In case of live analysis, fraud monetary transactions can be detected and blocked.
7. In case, the database data is deleted by the criminal, it can be recovered with the help of mirror database.

VII. RESULTS AND DISCUSSIONS

The proposed system can be revealed in following three different parts

7.1 Input Sets

Primarily from MS SQL server as well as My SQL Database (Fig 5) and Oracle Database server (Fig 6) all the significant log files are to be collected as inputs for the system with the help of audit files of respective databases. The XML output of the MS SQL server audit files, relevant output of Oracle audit table and of My SQL audit is shown in (Fig 7, Fig 8 and Fig 9).

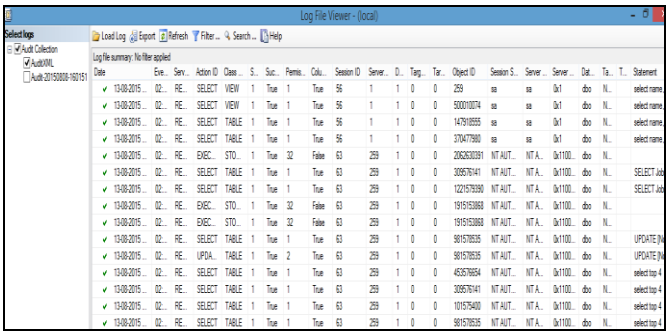

Figure 5. Snapshot Representing List of Logs in My SQL & SQL Server

IJARSE
ISSN 2319 - 8354

rishab, AUDIT_TEST, REENA, unknown, 2015-08-01 20:28:12.0, AUDIT_TEST, BLACKLISTIP, 7, DELETE, null, null, null, null, null, null, null, null, null, null, null, null, null, null, 3789, 2, 2, 0, null, null, null, null, 2015-8-1 20.28.12.300000000 +5:30, null, null, 0, 1940:5352, 0800240015010000, 1047199, null

, rishab, AUDIT_TEST, REENA, unknown, 2015-08-01 20:28:12.0, AUDIT_TEST, BLACKLISTIPR, 7, DELETE, null, null, null, null, null, null, null, null, null, null, null, null, null, null, 3790, 2, 2, 0, null, null, null, null, 2015-8-1 20.28.12.456000000 +5:30, null, null, 0, 1940:5016, null, 1047204, null

, REENA\rishab, AUDIT_TEST1, WORKGROUP\REENA, REENA, 2015-08-01 21:55:53.0, AUDIT_TEST, USERINFO, 7, DELETE, null, null, null, null, null, null, null, null, null, null, null, null, null, null, 3878, 10, 9, 0, null, null, null, null, 2015-8-1 21.55.53.136000000 +5:30, null, null, 0, 1940:3436, 0A0006001F010000, 1049957, null

, REENA\rishab, XYZ, WORKGROUP\REENA, REENA, 2015-08-01 22:17:25.0, AUDIT_TEST, USERINFO, 7, DELETE, null, null, null, null, null, null, null, null, null, null, null, null, null, null, 3888, 9, 8, 0, null, null, null, null, 2015-8-1 22.17.25.380000000 +5:30, null, null, 0, 1940:6492, 0200240006010000, 1050514, null

, REENA\rishab, PQR, WORKGROUP\REENA, REENA, 2015-08-02 11:50:15.0, AUDIT_TEST, USERINFO, 7, DELETE, null, null, null, null, null, null, null, null, null, null, null, null, null, null, 3894, 11, 8, 0, null, null, null, null, 2015-8-2 11.50.15.362000000 +5:30, null, null, 0, 1940:7280, 06000D0010010000, 1051940, null

, rishab, AUDIT_TEST, REENA, unknown, 2015-08-16 11:20:03.0, AUDIT_TEST, BLACKLISTIP, 7, DELETE, null, null, null, null, null, null, null, null, null, null, null, null, null, null, 4041, 2, 2, 0, null, null,

Figure 6. Snapshot Representing List of fields in Oracle Audit

7.2 Intermediate Steps

For all three Databases, outputs of audit files found also contain some extra information which is not needed for analysis. Thus, as in (Fig.5 & Fig.6) only the relevant parameters like Server principle name of table, action ID, time, object name and query executed are to be drawn by writing a separate programs to analyze the XML outputs of different log files obtained from three Databases for investigation.

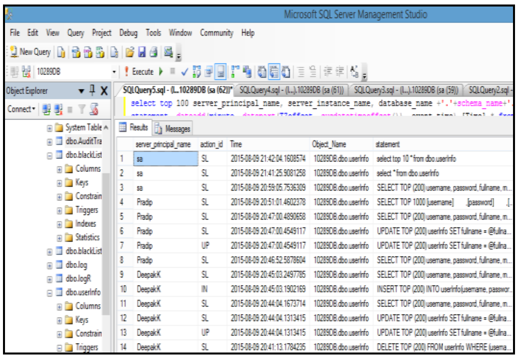

Figure 7. XML output of SQL Server.

USERNAME	TIMESTAMP	ACTION_NAME	OBJ_NAME
AUDIT_TEST1	2015-08-01 21:55:10.0	SELECT	SQLPLUS_PRODUCT_...
AUDIT_TEST1	2015-08-01 21:55:10.0	SELECT	PRODUCT_PRIVS
AUDIT_TEST1	2015-08-01 21:55:10.0	SELECT	SQLPLUS_PRODUCT_...
AUDIT_TEST1	2015-08-01 21:55:10.0	SELECT	PRODUCT_PRIVS
AUDIT_TEST1	2015-08-01 21:55:10.0	SELECT	DUAL
AUDIT_TEST1	2015-08-01 21:55:26.0	SELECT	USERINFO
AUDIT_TEST1	2015-08-01 21:55:35.0	SELECT	USERINFO
AUDIT_TEST1	2015-08-01 21:55:53.0	DELETE	USERINFO
XYZ	2015-08-01 22:14:03.0	LOGOFF	
XYZ	2015-08-01 22:14:03.0	SELECT	DUAL
XYZ	2015-08-01 22:14:03.0	SELECT	SQLPLUS_PRODUCT_...
XYZ	2015-08-01 22:14:04.0	SELECT	PRODUCT_PRIVS
XYZ	2015-08-01 22:14:04.0	SELECT	SQLPLUS_PRODUCT_...
XYZ	2015-08-01 22:14:04.0	SELECT	PRODUCT_PRIVS
XYZ	2015-08-01 22:14:04.0	SELECT	DUAL
XYZ	2015-08-01 22:14:29.0	SELECT	USERINFO
XYZ	2015-08-01 22:14:49.0	SELECT	USERINFO
AUDIT_TEST	2015-08-01 22:15:43.0	LOGON	
AUDIT_TEST	2015-08-01 22:15:53.0	LOGOFF	
AUDIT_TEST	2015-08-01 22:15:53.0	SELECT	DUAL
AUDIT_TEST	2015-08-01 22:15:53.0	SELECT	SQLPLUS_PRODUCT_...
AUDIT_TEST	2015-08-01 22:15:53.0	SELECT	PRODUCT_PRIVS
AUDIT_TEST	2015-08-01 22:15:53.0	SELECT	SQLPLUS_PRODUCT_...
AUDIT_TEST	2015-08-01 22:15:53.0	SELECT	PRODUCT_PRIVS
AUDIT_TEST	2015-08-01 22:15:53.0	EXECUTE PROCEDURE	DBMS_APPLICATION_I...
AUDIT_TEST	2015-08-01 22:15:53.0	EXECUTE PROCEDURE	DBMS_APPLICATION_I...
AUDIT_TEST	2015-08-01 22:15:53.0	SELECT	DUAL
XYZ	2015-08-01 22:16:20.0	LOGOFF BY CLEANUP	

Figure 8. Relevant Output of Oracle Database Server

dbUser	dateChanged	tableName	task
root@localhost	2015-08-16 18:44:10	userinfo	UPDATE
root@localhost	2015-08-16 18:44:28	userinfo	UPDATE
root@localhost	2015-08-16 18:44:28	userinfo	UPDATE
root@localhost	2015-08-16 18:44:28	userinfo	UPDATE
pradeep@localhost	2015-08-16 18:51:40	userinfo	UPDATE
pradeep@localhost	2015-08-16 18:51:40	userinfo	UPDATE
pradeep@localhost	2015-08-16 18:51:40	userinfo	UPDATE
pradeep@localhost	2015-08-16 18:51:40	userinfo	UPDATE
aaa@localhost	2015-08-16 18:53:32	userinfo	UPDATE
aaa@localhost	2015-08-16 18:53:32	userinfo	UPDATE
aaa@localhost	2015-08-16 18:53:32	userinfo	UPDATE
aaa@localhost	2015-08-16 18:53:32	userinfo	UPDATE
aaa@localhost	2015-08-16 18:57:17	userinfo	DELETE
aaa@localhost	2015-08-16 18:57:17	userinfo	DELETE

Figure 9.Relevant output of MySQL

7.3 Final Steps

After pre-processing/parsing the XML outputs, the estimated schema of metadata file (Fig.10) will contain all the required fields for investigation.

Figure 10. XML Schema of Metadata Files

Thus, result analysis reveals that by parsing only the relevant parameters from the log files and investigating only the identified data repositories the time complexity and space complexity can be reduced as compared to current system where time for analysis increases because the entire files needs to be analyzed for forensics investigation. Also, the idea of creating metadata files from multiple databases used in the system helps to increase the scalability factor.

VIII. CONCLUSION

From Research work reveals that database forensics is an important field of digital forensic but requires more research in this area. There are many tools and technologies in this field but all of them have their own limitations. Hence, to take these challenges as opportunities to make some innovations in the field of database forensics, there is a need for an open source methodologies and approaches. The new technical progresses have significantly increased the capacities of digital evidence being acquired and analyzed in a digital investigation. As a result, the process is becoming tedious and impossible. Moreover, metadata which plays an important role is abundant in today's digital systems and survey recognizes its value to database forensics, particularly with regard to event reconstruction. Analyzing all events across multiple diverse digital evidence sources can also help an investigator to view all the events which can be very valuable during an investigation. Thus, a framework is proposed to integrate multiple forensic and other analysis tools to achieve this task. Also, the integrity of metadata is preserved using strong hash technique like SHA-1 along with preventing and blocking database attacks. This will help to present the evidences in the court of law so that criminals are punished .Thus aims to maintain the overall information accountability by auditing all significant system generated log information, independent of the DBMS used.

IX. ACKNOWLEDGMENT

I would like to thanks Dr. Jagdish W. Bakal for his motivating support and useful advice about this paper.

REFERENCES

[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] Harmeet Khanuja And S.S.Suratkar, "ROLE OF METADATA IN FORENSIC ANALYSIS OF DATABASE ATTACK". 978-1-4799-2572-8/14/S31 2014 IEEE Conference.

[3] O.M.Fasan And M.S. Olivier, "ON DIMENSIONS OF RECONSTRUCTION IN DATABASE FORENSICS" Seventh International Workshop On Digital Forensics & Incident Analysis (WDFIA) 2012.

[4] 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.

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

[6] Ali Reza Arasteh, Mourad Debbabi, Assaad Sakha, Mohamed Saleh, "ANALYZING MULTIPLE LOGS FOR FORENSIC EVIDENCE" Science Direct Digital Investigation4s(2000)S82 – S91.

[7] 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.

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

[9] 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).

[10] Harmeet Kaur Khanuja, D.S. Adane, "A FRAMEWORK FOR DATABASE FORENSIC NALYSIS" Computer Science & Engineering: An International Journal (CSEIJ), Vol.2, No.3, June 2012.

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

[12] Kyriacos E. Pavlou And Richard T. Snodgrass, "TEMPORAL IMPLICATIONS OF DATABASE INFORMATION ACCOUNTABILITY" 19th International Symposium On Temporal Representation And Reasoning 2012.

[13] 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.

[14] Dr.M.Amutha Prabakar, M.Karthikeyan, Prof.K. Marimuthu, "AN EFFICIENT TECHNIQUE FOR PREVENTING SQL INJECTION ATTACK USING PATTERN MATCHING ALGORITHM" 2013 IEEE International Conference On Emerging Trends In Computing, Communication And Nanotechnology (ICECCN 2013).