ISSN:0975-9646

An Improved Framework for Tamper Detection in Databases

Manisha K. Kambire[#1], Pooja H. Gaikwad[#2],Sujata Y. Gadilkar[#3] ,Yogesh A. Funde[#4]

[#1234]*Fellow Internship*
[#] *Innovatus Technologies, Pune,*
Maharashtra, India

***Abstract*— The Corporate world, government sector and many organizations has a requirement of secure database system. So due to huge amount of data most of the organizations prefers outsourcing of databases to trusted third parties. In many of the scenarios database tampering is happen due to inside employee of the organizations, so a need of highly efficient tamper detection system is required which answers all the queries of the post crime. Most of the existing methodologies and algorithms are helping to determine *when* and *what* of the tampered database. These methodologies generally not restore the pre-tampered original data. The proposed method presents an idea of improved framework for tamper detection in databases where it not only identifies *which* and *what* of tampering and also detects *who* and *when* parameters of the crime. System effectively uses the xml logs of database to identify the culprit. This boost the idea of forensic analysis for database tampering in huge setups.**

Keywords: Tamper Detection, Notarizer, Validator, MD5,Investigator, culprit.

I. INTRODUCTION

As the internet is available over the every corner of the world it is crowded by the users like never before. So, the increasing number of online users alarmingly raises the data in the database through their web applications. In the many domains the data is becoming so huge in every minute and it is also important to protect it from the inside attackers of the storage organizations. Domains like telecom, banking and online shopping are heavily relay on systems which are protecting the data.

There are some methodologies through which we can see the evaluation of modern database tamper detection system. The prior methodologies are monochromatic algorithm, RGB algorithm, RGBY algorithm, A3D algorithm, and polychromatic algorithm.

- **Monochromatic Algorithm :**

In Monochromatic algorithm it uses only a single hash chain for the given instant to identify the tampering data.

- **RGB Algorithm :**

In RGB algorithm the new types of hash chains are added which are indicating three colours like red, green and blue. All these three chains are parallely synchronized to commit the transactions in database.

- **RGBY Algorithm :**

RGBY algorithm improves RGB algorithm by adding another hash chain called 'Y' which denotes yellow colour. This extra hash chain is taken care of notarization service of transactions.

- **A3D Algorithm :**

This is one of the most advanced algorithms as it does not lay continuously for a fixed pattern of hash chains over the database. Instead of this the length of partial chains keep increasing as the transaction time increases.

- **Polychromatic Algorithm :**

This algorithm creates 'N' number of hash chains over the databases all the hash chains are sync parallely to commit the transactions and to identify tampering.

There are many different techniques to detect the tamper in data. Paper discuss some of the methodologies which are really contributed in data forensic analysis system.

DRAGOON [1] is an information accountability system for high performance databases where it is implemented to determine *when* tampering occurred and *what* data were tampered with. DRAGOON is scalable customizable and initiative and it provide a guarantee again its inside threats for the database. This prototype work on cryptographic hash function which are expanded to include forensic analysis.

This prototype work on the basis of capturing timestamp and compute a cryptographically strong one way hash function of tuple data and the time start. The hash value obtain from the different transactions accumulatively link from a hash chain. Here in DRAGOON a system called notarizer periodically send a hash value as digital document to external agents which actually authentic the data and it returns smaller notary ID's which are stored in MySQL managed database to form a secure master database. Then a system called validator initiates the traversing process for the whole database and hashes the data along the timestamp of each tuple. And then it compares with previously stored hash chain, if any change occurs in hash chain then it is considers as valid tampering.

[2]Demonstrate that the fuzzy method can also be used to perform forensic analysis in the data by using the three steps like clustering, extracting membership function from the data and fuzzy inference system.

Clustering data method actually clusters the similar data and forms the clusters with the unique ID. These clusters are then use to extract membership function depend on the similarity measure using Gaussian membership function. The extract membership function are then use for the fuzzy inference engine, where it contain three processes like fuzzyfier, defuzzyfier and inference engine.

Fuzzyfier takes the input of crisp values from extracted membership functions then it process with defuzzzyfier to get classified values. Then these values are proceed with inference engine to perform if-then rules to yield classified data of tampering.

The rest of the paper is organized as follows: Section 2 discusses some literature survey and section 3 presents the proposed methodology our approach. The details of the results and some discussions on this approach are presented in section 4 as Results and Discussions. Section 5 elaborates hint of some extension of the approach as future work and conclusion.

II. LITERATURE SERVEY

The proposed idea is based on morality of protecting the client's data by the third party. So the third party is always ensures the right source of the data using authentication of signatures.

[3] Propose a message authentication scheme based on cryptographic hash function. This uses NMAC and HMAC hash function to strength the cryptographic hashing technique.

Many systems are developed to identify the changes in data in the network by changing some top cryptographic hash functions where it scrutinizes the integrity of the data on gateway of data transferring [4]. [5] Invents an idea to protect the documents from tampering by checking the audit logs of previous access time. This is one of the most usable techniques in many of the tampered detection system where always the initial data is maintain unaltered and then it is referred as previous one.

The concept proposed in [5] can enhance more accurately for huge data by empowering several fundamental data structures with a novel techniques for organizing index structure as mention in [6]. A proper indexing is always eases the time of processing by construction of several data structures entities like arrays, linked list, trees and hash tables.

InnoDB is famous database engine which stores the data for MySQL, [7] represents the forensic system of InnoDB by representing the practical demonstration of reconstructing a data of any SQL table from the available dataset files. The advantage of this is to recover the data by the inconsistent activity on the database.

Many forensic systems are inactively working on log based tampered evident system. [8] Presents a framework where it identifies the tampering while entering the data through the applications by the user. By checking the integrity of original data, information is stored on server logs.

[9] Identifies tampered detection by considering the case of untrusted loggers who serves many clients. This system considers auditors as honest and identifies tampering based semantic evident in logs available while auditing process.

QUERIFIER is a tool used in [10] which actually analyse offline static logs. The behaviour of QUERIFIER is purely depend on flexible pattern matching language which is used to describe QUERIFIER . This makes the QUERIFIER to take extra clock time for identifying evident of tampering in logs.

Some forensic systems are also proposed like [11] to identify the tamper evidence even in multimedia information. [12] Expresses an idea to identify activity of any suspicious behaviour within database. Here system identify, collect, analyse, validate and interpret the process of forensic analysis.

For efficient tampered detection system implementation in database the machine performance is also most considerable fact, this is broadly discussed in [13].

[14] Shows a novel hiding technique to counter the detection of manipulations in the images through forensic analysis. Here it describes a technique of resizing and rotating bitmap images without living any periodic traces. This kind of method can vastly implement on digital image authentication. Ahead of this process [15] identifies a system of forensic analysis of Windows NT file system. The advantage of this system lies on NTFS boot sector fragmentation methodology which is most accurate to provide the best forensic analysis system for images.

Many of the audit logs system in forensic analysis widely depend on the audit log. So [16] provides security for the audit logs, So that even on accessing these log by the anonymous user it ensures that he will gain very little or no information from the logs.

Conducting forensic analysis on digital images is bit difficult task as it requires careful experimental analysis. So, to overcome this [17] Introduce an idea of resampling of digital images. Here author uses deeper-than-usual analysis for its performance resampling.

[18] Discuses the forensic analysis challenges in mobile phone. As now a days the mobile phone become an inseparable part of human life, so it is a easy target to any offender to do the tampering in contacts, sms's, images of mobile device . [18] Also classifies as levels of forensic analysis tools in mobile phones like manual, logical, hex-dump, chip-off and micro read.

Snodgrass in [19] describes the way where tamper happens illegally with a database mostly by a insider. Here it is described achieving accountability in cloud database by using replication service notarizer and audit logs. Audit logs are always playing an important role in forensic analysis as they do have the traces of transactions in unstructured format. So, in this case mining proper data from the logs plays a crucial role.

[20] Introduces another way of performing a digital forensic analysis on images in the absence of any digital watermark or signature. Here author quantified the nature of statistical correlations that results from specific form of digital tampering and they have devised detection schemes to reveal these correlation.

[21] Introduces an idea of self protecting system which is powered to detect malicious activity at run time and prevent the execution. The system also argues the presence of false positiveness due to lack of information in run time. Here the forensic analysis is done on the basis of kernel level auditing of system activities.

III. PROPOSED METHODOLOGY

In this section paper narrating the proposed framework for tamper detection in database. For efficient understanding of proposed system we need to understand the procedure for working of third party data handling scenario. A third party service provider is an organization which actually stores and protects many clients data upon same mutual agreement. So, there will be a huge threat for the data from the internal employees of third party. So, third party service provider need to enhance the system of his security to protect the data from tampering and algorithm need to identify the culprits who succeed in tampering . So this completes scenario is perfectly coupled with our system and which can be shown in fig. 1

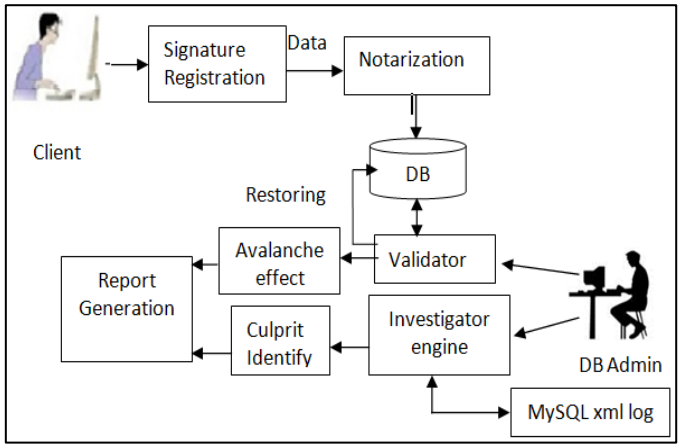

Figure 1. Overview of tamper detection system in database

Our proposed method can be easily describe with following steps:

Step 1: Here all the clients who are interested to outsource their database for restoring purpose need register at third party. This registration is empowered signatures creation and registration along with the profile creation . This signature creation will be done using strong one way hashing algorithm like MD5.On saving every record of data on third party side by the client system ensures the right source of the clients using the unique signature which is provided to client by the system.

Step 2: Here a service called notarization is actually participate in finding the right source of data on its arrival from the client . This is done by comparing signatures of the client with its original signature which store on profile creation. If the signatures are match then only data from the client are allowed to store third party database server. So, now the data which is store at third party can be considered as original and generated from right source.

Step 3: A special subsystem is deployed to identify the tampered id in the database called a validator. Here validator is assigned a time for visiting the database in regular intervals. For each visit of the database all the records are fetched and processed in the form of hash keys formed by MD5 algorithm. These hash key sets are compared with the previous visit set for the intrusion for the assigned time of validator. Then the changes in a single bit of hash key reflects a greater change in the database. This is represents as an "Avalanche effect".

Once the "Avalanche effect" is been identify in the database then every record in the database are linearly compared with the previous one to identify the exact attributes of tampered data. This process is enforced with recursive multithreading that efficiently handles the tampered detection process even in smaller time of validation. This can be shown in algorithm 1

Algorithm 1

//Input:- T_b as Table name
T as Time interval
//Output:- Tampered id

0: Start
1: Get database as DB and table name as T
2: Visit table T_b at interval T
3: T_b contain in double dimension m_v (master vector)
4: **for** i= 0 to m_v size at T
5: **for** j=0 to m_v size at T-1
6: Get DB tuple as D_T and D_{T-1}
7: Apply MD5 hash on D_T and D_{T-1}
8: **if** $(D_T != D_{T-1})$
9: **then** detect tamper
10: Stop inner for
11: Stop outer for
12: Stop

Step 4: After detection of *which* and *what* now our system is keen to identify *who* and *when* of the tampering. This is done by a heuristic approach where another a validation is triggered simultaneously for the assigned time. This validator actually keeps an eye on MySQL xml log to identify the culprit name on detection of tampering in database by the step 3. Once the culprit name is extracted from xml log immediately system extracted the tamper date and time and report all in a well structured manner.

Step 5: Once the system successfully detects *what*, *who* and *when* of tampering then the main issue is remind the system is about the loss of the data. To recover this loss the data which is carried by the validator in step 3 in its previous visit is actually the original data. This is been restore again in the database for the tampered id to make up the loss. This feature of our system always makes the client to keep the data at third party without any doubts.

IV. Result And Discussion

To evaluate the performance of our system the series of experiment are conducted on two dummy databases belongs to a bank and hospital. Each of the databases contains the maximum of the 2000 records and the experiment was conducted on MySQL Server 5.5 on java based windows platform machines. On conducting the experiment for different number of record in database for a minute of validation time we observe the following outcomes as shown below in Table 1.

Number Of Records	Time in Seconds for Dummy Bank DB	Time in Seconds for Dummy Hospital DB
1250	1.72	1.72
1500	1.64	1.66
1750	2.11	2.09
2000	2.96	2.19

Table 1. Evaluation of summary factor

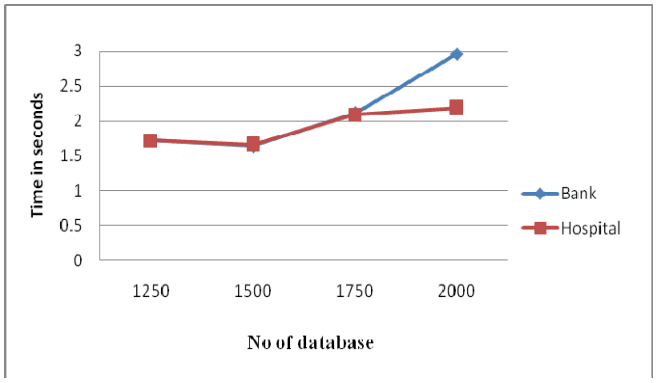

Figure2. Forensic analyser performance

The plot in figure 2 indicates not much more deflection in performance time. It is been observed that as the number of records increased the time taken for forensic analysis is not increasing by direct proportionate. So it clearly indicates system performance is better as a data increases. This is exactly due to parallel computation of synchronised detection of tampering in databases in object level.

V. Conclusion And Future Work

Propose system successfully identifies the tampered records for the given interval of the time. This is mainly due to the high performance of extensive multithreads in program. As a step ahead system identifies the culprit of tampering by constantly keeping an eye on MySQL user logs with accurate crime time. Systems greatly compensate the tampered data by restoring it with its original content.

The tampered detection and forensic system can enhance to perform on huge databases which are in cloud system by using distributed parallel computing.

References

[1]. Kyriacos E. Pavlou 1 and Richard T. Snodgrass, "DRAGOON: An Information Accountability System for High-Performance Databases", *International Conference on Data Engineering (ICDE)*, April 2012.

[2]. Kilian Stoffel, Paul Cotofrei, Dong Han, "Fuzzy Methods for Forensic Data Analysis", Soft Computing and Pattern Recognition (SoCPaR), International Conference 2010.

[3]. Mihir Bellare Ran Canetti Hugo Krawczyk, "Keying Hash Functions for Message Authentication", Advances in Cryptology - Crypto 96 Proceedings,**Lecture Notes in Computer Science Vol. 1109, N. Koblitz ed., Springer-Verlag, 1996.**

[4] John Edward Silva, "An Overview of Cryptographic Hash Functions and Their Uses", GIAC Security Essentials Practical Version 1.4b Option 1 January 15, 2003.

[5]. Kyriacos E. Pavlou and Richard T. Snodgrass, "Forensic Analysis of Database Tampering", ACM Transactions on Database Systems, Vol. V, No. N, September 2008.

[6]. Mikhail J. Atallah, "Indexing Information for Data Forensics" , CERIAS Tech Report 2005-131

[7]. Peter Frühwirt, Markus Huber, Martin Mulazzani, Edgar R. Weippl, "InnoDB Database Forensics", ARES 2012 - 2012 Seventh International Conference on Availability, Reliability and Security

[8]. Pallavi D Abhonkar, Ashok Kanthe, "Enriching Forensic Analysis process for Tampered Data in Database", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (5) , 2012,5078-5085

[9]. Scott A. Crosby Dan S. Wallach, "Efficient Data Structures for Tamper-Evident Logging",USENIX, Aug 2009.

[10]. Daniel Sandler, Kyle Derr, Scott Crosby, Dan S. Wallach, "Finding the Evidence in Tamper-Evident Logs",CS publication 22 May 2008.

[11]. Wenjun Lu, Avinash L. Varna and Min Wu, "Forensic Hash for Multimedia Information", SPIE Media Forensics and Security, 2010 .

[12]. Harmeet Kaur Khanuja1 and D.S.Adane, "A FRAMEWORK FOR DATABASE FORENSIC ANALYSIS", Computer Science & Engineering: An International Journal (CSEIJ), Vol.2, No.3, June 2012.

[13] Daniel Ayers, "A second generation computer forensic analysis system", digital investigation 6 (2009).

[14] Matthias Kirchner and Rainer B¨ohme, "Tamper Hiding: Defeating Image Forensics", Information Hiding Lecture Notes in Computer Science Volume 4567, 2007, pp 326-341.

[15] Mamoun Alazab, Sitalakshmi Venkatraman, Paul Watters, "EFFECTIVE DIGITAL FORENSIC ANALYSIS OF THE NTFS DISK IMAGE", ICIT 2009 Conference

[16] Bruce Schneier John Kelsey, "Secure Audit Logs to Support Computer Forensics", ACM Transactions on Information and System Security , v. 1, n. 3, 1999,

[17] *F. Uccheddu +, A. De Rosa +, A. Piva +, M. Barni,* "DETECTION OF RESAMPLED IMAGES:PERFORMANCE ANALYSIS AND PRACTICAL CHALLENGES", 18th European Signal Processing Conference (EUSIPCO-2010)

[18] Amjad Zareen, "Mobile Phone Forensics Challenges, Analysis and Tools Classification", 2010 Fifth International Workshop on Systematic Approaches to Digital Forensic Engineering.

[19] Kyriacos E. Pavlou and Richard T. Snodgrass, "Achieving Database Information Accountability in the Cloud".

[20] Alin C. Popescu and Hany Farid , "Statistical Tools for Digital Forensics",

[21] Ashvin Goel , Wu-chang Feng , Wu-chi Feng , David Maier, "Automatic high-performance reconstruction and recovery" , Computer Networks 51 (2007) 1361–1377