TILED BITMAP ALGORITHM AND FORENSIC ANALYSIS OF DATA TAMPERING (AN EVOLUTIONARY APPROACH)

Ashish Chavan[1], Sana Shaikh[2], Radhika Randhir[3]

[1, 2, 3]*Department Of Computer Science and Engineering, Zeal Education Society's DCOER, Pune, Maharashtra, India*

Abstract

Secure data storage is need of this era. Medical, financial, military and private authorities' legal information are all in need of secure storage. In the era of globalization and dynamic world data outsourcing is unavoidable. Security is highly concern in data outsourcing environment, as data is under the third party service provider. In present systems, third party can access & view data information even though they don't' have authority to do so or even when the data outsourced to the auditors and allows the employee of that organization to do modification in the database. This may lead to the serious data leakage, data tampering & even data theft causing severe business impact to the data owner. As in now there are many such cases occurred in finance & insurance sector where the data tampering done by the auditors or by the employees of the same organization itself. In this paper we have proposed a basic solution to overcome the problem of tamper detection by notarizing the original data. A heuristics approach is presented in our model where a validator always authenticate the data for original identity using strong one way hash key functions like MD5 hash key provided with protected valid notarizer. By providing distinct digital signatures for many data owners, the proposed system aims a strong notarization service & validation schemes to maintain high data security and high integrity requirements.

Keywords: *Security, hash key,MD5,SHA-1,DSA, Data outsourcing, Notarization, Efficient Notarizer, Digital Signature, Validator, Forensic Analysis, Data Set, Avalanche Effect, Secure Data Server.*

----------***----------

1. INTRODUCTION

Secure storage of data is an everyday need for public businesses, government sectors and many institutions. For many organizations, if data is unauthorisely modified, whether by an outsider or by an internal intruder, it could cause severe problems for the company. And even for their clients as well.
There can be many reasons why someone want to tamper the data. For example, an student who receives a "E" grade in his mathematics subject, in which he needed at least a "B+", could try to dishonestly change his grade to a "B+" in the database fo school. This would be an example of outside intrusion, unless the student somehow access to the database containing the grade. Outsourcing healthcare Insurance services are highly emerging today. However, there are many concerns raised about data security and its standard quality norms. The Health Insurance Portability and Accountability Act (HIPAA) are widely renown norm for healthcare facilities and Indian mnc's are well versed with this Act and other regulatory bodies. Other standards/acts relevant for data security are:

* ITA-2000: The Information Technology Act 2000
* PCI DSS: Payment Card Industry Data Security Standard
* ISO27001, ISO 27001: Information Security Standard

At present, very few technologies are used to protect critical data using Notarization services are shown

In this paper we have presented an approach where the System is providing Panel and platform to the data owner where he/she creates a Digital signature by using which digital signature can always notarize (authenticating client while performing each and every database transactions) the transaction process. This Operation will be performing by a Validator to check the integrity constraints of the data for every single transaction. Legitimate and authorized users will be checked by their given distinct digital signature by corresponding data owner by the notarizer service. similarly, the confidentiality, originality & integrity factors of information does not rely on an implicit assertions of trust on the server for legal protection provided by specific service contracts, but instead trust on the technical services assured provided by Notarizer and validator.

2. RELATED WORK

Many models are proposed to find the tamper detection of data tamper process like

2.1 Monochromatic Algorithm

The Monochromatic Algorithm uses only the progressive (black) hash chains that we have seen so far, and it is the simplest algorithm considering in terms of implementation.

2.2 RGB Algorithm

In the RGB Algorithm, three new types' chains were introduced, denoted by the colors red, green, and blue, to the original (black) chain in the so-called Monochromatic Algorithm. These hash chains can be calculated concurrently; all constitutes linked series of hash value of single transactions in committed order. While added hash values must be computed, and there's no such additional disk reads are needed. The additional processing is entirely done in main memory. The RGBY Algorithm contains the red, blue, green chains and adds on a yellow chain. This renders the new algorithm more basic and more powerful.

2.3 RGBY Algorithm

The RGBY algorithm is advanced as compared to RGB Algorithm and improvement of the previous RGB Algorithm. The main perception of the previously presented R-G-B forensic analysis algorithm during notarization events, additionally to restructuring the entire hash chain , Validator rehashes the portions of database and then notarize those values, both are done separately from full chain.

2.4 A3D Algorithm

The a3D Algorithm is the most advanced algorithm that does not lay repeatedly a "fixed" pattern of hash chains over the database. Instead, the lengths of the partial hash chains varies (decrease or increase) as the transaction time varies whether increment or decrement, in such a way so that at every point in time a complete binary tree of hash chains generated on the top of the database. This enables forensic analysis to be speed up significantly and magnificently.

2.5 Tiled Bitmap Algorithm

This introduces the concept of a candidate set (all possible locations of detected tampering(s)) and provides a full behavior of the candidate set and its cardinality. An efficient algorithm for computing the candidate set is also shown. At last, the implementation of the Tiled Bitmap Algorithm is described, along with a contrast to other forensic algorithms in terms of space complexity, time complexity and estimated cost.

Where candidate Set Method is to organize values of targeted binary array in the reverse sequence and renumbering method to rearrange values of targeted binary array in required exact order

In our proposed System the DBMS manipulates a cryptographically a strong one way hash key function for each records filled and its then notarized with the help of notarization service, which is done to determine the consistency of the data by correlating it to the values stored with the value of notarization service. In continuation with this process, algorithms were designed to further analyzing any intrusion of entry in database.

3. PROPOSED METHOD

In this part, we discuss our approach how Tamper Detected and Forensic Analysis accordance to the steps shown in figure 2.As shown in figure consists13 main steps.

Step 1, 2, 3: this approach have taken certain scenario whose architecture is shown in Fig-1 where the data admin who outsourced his data to the 3rd party server whose creating his digital signature by using any text file consists of any secret information/document of his own. This signature created by using SHA-1 algorithm combined with DSA (Digital Signature Algorithm) produces a digital certificate file which is totally encrypted. This also generates a private key for verification.

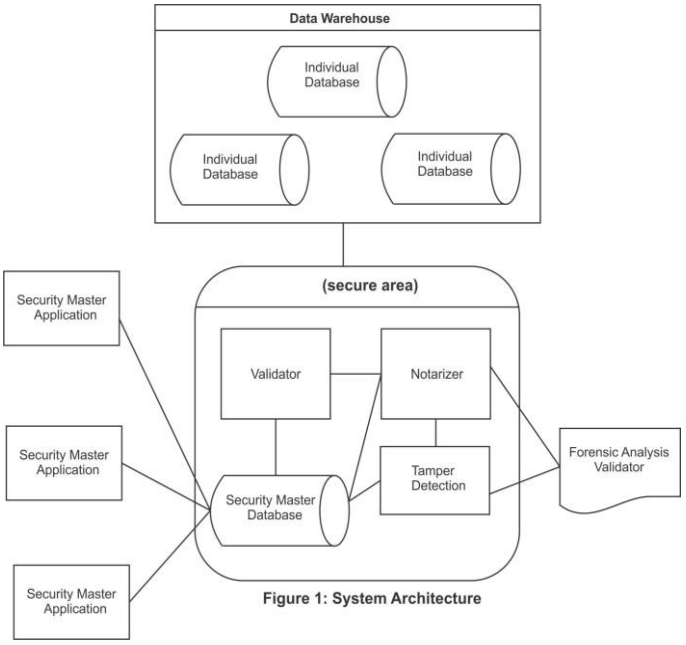

Figure 1: System Architecture

Step 4, 5, 6: In these three steps we upload both the master data on which process needed to be performed & also a digital signature. This digital signature is created by using SHA-1 hashing algorithm gives high security. This SHA-1 with DSA algorithm strengthen a strong digital signature which acts as a unique signature of the corresponding owner for the own data which is to be handled by the 3rd party server that is by the employees & auditors.

By using panel provided the data admin in our web application owner will use by uploading both master data and digital certificate which is stored in a specific location of the web server location of data ware house system.
Step 7: step focuses that whenever a transaction occurs to database then all the data attributes entering through the web application either by a employee of organization or outside

auditors by provided their panel is also stored temporally in the java bean classes.

Step 8: The digital signature which is in data ware house third party server by the data owner in step 4 behaves as a notarizer, element for every transaction done by the employees & auditors. Whenever a transaction occurs for each of those transactions the notarizer does authenticate for the private key. If this private key is similar as provided by the data admin then it notarizes the transaction positively. Else the transaction denied.

Step 9: Validator is the process which actually provides the transacting details to the notarizer to validate the data.

Step 10: If the data is authenticated by validator using a strong cryptographically one way hashing is performed by using MD5 algorithm which results to a sixteen byte hash value for the master data & also for the transaction log by the auditor or by inside the employee of the organization .

Step 11, 12: In this step the MasterDatabase5 hash value of the master data & the transaction data are both verified for avalanche effect. If the avalanche affect causes any positive modification then we consider there's tampering happened at the transaction data then we denote that data set to an infected one or a tampered one. In this way we are going to detect the data tampering for the various data owners for their respective data with the respective applications with their corresponding signatures. The proposed model provides an exact privacy for the data owner with their own private signatures so we can guarantee that in our taken situation both the data and the tamper detection process will last for a long.

Step 13:once the data record is considered as tampered then the forensic analysis is performed on this record to find who's the person who tampered the data, when did tampering happened and what are the exact fields or tuples where tampering occurred.

Here in our proposed algorithm it takes Master data set and transaction data set then we create a another set called data set Dset which actually is array of data field index whose value initially is "0". This shows that the fields are not yet tampered again for each master data set Mset and for each transaction data set Tset this algorithm takes each data attributes of these two sets and compare both. If unequal then that data record is considered as tampered and then di that belongs to Dset is set to value "1". This way the complete tuple is checked for the accurate tampered fields. Then this array of tampered fields is reorganized and also checking for more parts and modules to print the final result.

The name of the person who tampered the data can be recognized using servlet which actually set username as he/ she login into the system by using date time operation on the same instance we can manipulate exact time when data tampering happened.

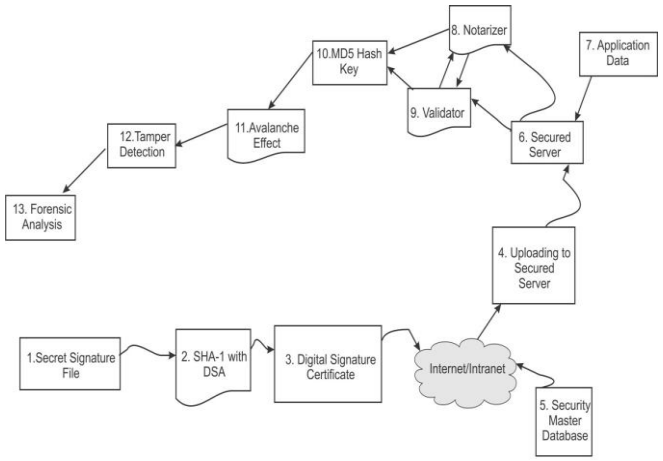

Fig-2 Overview of our Approach

We propose the above Described Tamper detection method by using following:

1. Text file is converted into a Digital Signature using SHA-1 with DSA
2. Digital Signature validated
3. if (Validated) then

MD5 is applied for Original and Transaction Data to obtain 16 byte hash keys

4. Avalanche Effect is checked for both Hash keys

if(Avalanche effect) then

Data tamper has occured

else

Data is Safe

We propose the above Discussed Forensic Analysis method by using following Algorithm:

// input: M_{set} is the set of master database
// T_{set} is the set of Transaction database
// D_{set} is the set of Data Field Index
// Un is Username
// td is date and time
// R_{set} is the set of Result
// output: R_{set} the set of Result

function forensicAnalysis ($M_{set,}$ $T_{set,}$ D_{set} ,Un, td, R_{set})
1: d_i =0 // index data field
2: $R_{set=}$ ""
3: for i= 1 to number of data fields
4: t_i data field of T_{set}
5: m_i data field of M_{set}
6: if ti and m_i are not equal then
7: d_i =1
8: end of for
9: for i= 1 to number of data fields
10: if d_i =1

11: $R_{set} = R_{set} + d_i$
12: Return R_{set}

Algorithm	Running Time	Cost	Space Complexity	Dynamic Performance Based on Hash Chains	Multiple Corruption Event
Monochromatic	Fast	High	More	No	No
RGB	Low	High	More	No	No
A3D	Low	Medium	Medium	No	Yes
Tiled Bitmap	Medium	Low	Less	No	Yes
Our Approach	Fast	Low	Less	Yes	Yes

Table No 3: Comparision of All Algorithm with our Approach

4. RESULT ANALYSIS

As suggested in Tiled Bitmap algorithm [1] database tampering can be done for tiles here in our approach we are using dynamic data.

A robust and well-organized database was created to maintain the system as a whole and provide the central point of interaction for all tools in the system.

This database is easily expandable for future versions of this project. Three applications were created, one for each role, to allow users to interface with the Database. They provide an organized and controlled way for employees and auditors to interact with the system. The Complete System is integrated and developed using Java classes for Apache tomcat server edition. At this point, there are many ideas for additions that can be made to these applications; some ideas are outlined in the next section.

All in all, a large step was made toward securing databases from intrusion and the maintenance of such intrusions. By utilizing a central security master database as part of enterprise architecture for auditing, as well as role-specific GUIs, it is possible to efficiently manage the auditing of databases across an enterprise.

4.1 Results for Intrusion and Forensic Analysis.

As we Discussed in section 3 where we have taken the scenario of outsourcing data storage to the third party server and if the tampering is happened then its detected on the instance and forensic analysis is also been done and print the results in a notepad file. This result can be shown in below figure 3.

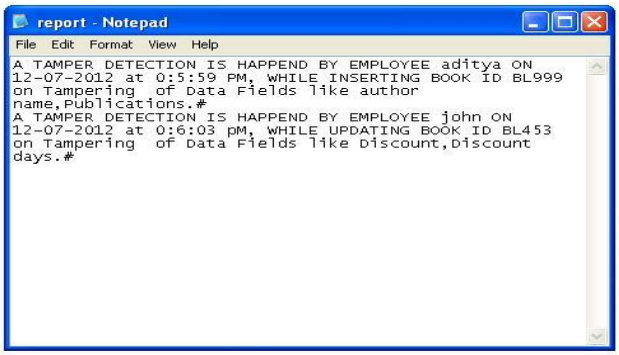

Fig 3: Results for Tamper detection and Forensic Analysis

CONCLUSIONS

Forensic analysis commences when a crime has been detected, in this case the tampering of a database. Such analysis attain to ascertain when the tampering occurs, and what data were updated. The present paper expands upon that work by presenting the Tiled Bitmap Algorithm, which is cheaper and more powerful than prior algorithms. This algorithm employs a logarithmic number of hash chains within each tile to narrow down the when and what. Checking the hash chain values produces a binary number; it is the task of the algorithm to compute the pre image of bitwise We also note that previous algorithms do not handle multiple corruption events well, whereas the Tiled Bitmap Algorithm can independently analyze corruption events occurring both in different tiles and multiple corruption events occurring within a single tile.

By creating a central database for all of the tools in the system to interact with it made it possible for the notarizer and validator to perform their operations successfully.

They can now store their data in this central database as well as use the information stored in it to schedule future executions. The three role-specific applications allow auditing or data updation to be started on individual databases and then be maintained. The necessary tools for auditing a database are in place. It is now possible for Doctor's offices, companies, and government agencies to protect their information from threats by implementing this Enriched System.

REFRENCES

[1]. "The tiled bitmap forensic analysis algorithm", K.E. Pavlou and R.T. Snodgrass, IEEE transaction on knowledge and data engineering, Vol. 22, pp no.590-601, April 2010

[2]. CSI/FBI, "Tenth Annual Computer Crime and Security Survey," http://www.cpppe.um.edu/Book store/Documents/2005CSISurvey.pdf, 2009.

[3]. "An Infrastructure for Database Tamper Detection and Forensic Analysis", M. Malmgren, honors thesis, Univ. of Arizona, http://www.cs.arizona.edu/projects/tau/tbdb/MelindaMalmgrenThesis.pdf, 2009.

[4]. U.S. Dept. of Health & Human Services, The Health Insurance Portability and Accountability Act (HIPAA), http://www.cms.hhs. Gov/HIPAAGenInfo/, 2009.
[5]. Investigative Data Mining for Security and Criminal Detection. J. Mena, Butterworth Heinemann, 2003.
[6]. "Indexing Information for Data Forensics", M.T. Goodrich, M.J. Atallahand, and R. Tamassia, Proc. Conf. Applied Cryptography and Network Security, pp. 206-221, 2005.
[7]. "Tamper Detection in Audit Logs", R.T. Snodgrass, S.S. Yao, and C. Collberg, Proc. Int'l Conf. Very Large Databases, pp. 504-515, Sept. 2004.
[8]. "Forensic Analysis of Database Tampering", K.E. Pavlou and R.T. Snodgrass, Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 109-120, June 2006.