An Automatically Detecting Integrity Violations in Database-Centric Applications

Mr. Wasif Khan, Mr. Pawan Tule, Mr. Saurabh Tiwari, Mr. Hrishikesh Karale, Guided by *Prof. Vaibhav Muddebihalkar*

Dr. Dy Patil Institute of Technology, Pimpri, Pune-411018

ABSTRACT

The task of maintaining database Integrity is always complex and hectic because there's tremendous amount of data in the database, the business analysts or database administrators can also make errors that lead to integrity violations, so the tools or applications which are handling these things should be empowered with high efficiency techniques. If there weren't any database integrity detection applications then it would be difficult for a client to trust the 3rd party service providers. Many methodologies and tools exists which are maintaining the integrity with proper accuracy but most of them do not reveal any details about the cause also these methodologies suffer from time and space complexities, so to enhance the process of maintaining database integrity proposed methodology put forwards an idea of detecting tampering attack on databases using tiled bitmap technique which is powered with avalanche effect of the hash codes of data tuples in regular intervals. System not only detects all the details of forensics and also it clearly restores the original database.

To enhance the process of maintaining database integrity proposed methodology put forwards an idea of detecting tampering attack on databases using tiled bitmap technique which is powered with avalanche effect of the hash codes of data tuples in regular intervals. System not only detects all the details of forensics and also it clearly restores the original database.

Keyword: - *Validation, Bilinear Pairing, Database Tampering and Avalanche Effect.*

1. INTRODUCTION

Database integrity violation detecting applications are widely used in various companies and organization for data analysis tasks. Database integrity violation detecting application are often written using some popular programming languages like JAVA, and they access and restore data from databases. Databases which contains various data are tremendous, they contain n number of tables consisting of n number of attributes. This tool is very helpful in detecting integrity violation if found in the system. BA and database administrator create and mention logical and physical schema of databases that has sophisticated checks,

Our system makes the noteworthy contribution

- We propose a solution for the problem in detecting probable integrity violations in database integrity violation detecting application
- We also generate forensic analysis of databases which signifies whether the database integrity violated? What is the time of violation? and also shows the ip address of the system through which violation occurs.
- Along with this service the data will be restored into its original form.

The section two of this paper is dedicated for analysis of the past work as literature survey segment. However section three is the propose technique is broadly elaborated whereas in fourth section the comparative analysis of the system is done. And finally in fifth section finally system is concluded with its future views and possibilities.

2. LITERATURE SURVEY

[1] Introduces Objects are categorized on the base of their distance, connectivity, density, so that similar objet can be found in the group and vice versa. We have used k-means and clarans algorithm for clustering."Matlab" is used for studying clustering and validation process, validation helps to select good quality cluster by using above two algorithms. With the limitations of being an expensive process as it include so many computations.[2] Narrates that we are retrieving the metadata from the database it will create identical mapping of the NOSQL database. Validation carried out in two phases 1 - bloom filter, 2 - cell validation engine in this we compare our source data with target data using bloom filter, the unmatched data is then forwarded to the cell validation engine which shows the corrupted cell.With the limit that false positive error may occur.[3] Explains a way to determine special character from the multi-lead ECG transmitted signals. We will be measuring this parameter in the QRS parts, PR seg, QT parts, waves are denoted by(P,Q,R,S,T)

1. Pre-processing :- It consist of eliminating noise and non-linear transformation to improve QRS detection.
2. Multi-lead QRS Checker: - It uses the multi lead QRS detector, considering the information of signal slope, it will consider the points whose distance is not more than 90ms from each other.
3. Fibrillation process :- No onset and offset is detected
4. Wave location:- It searches for the nearest QRS position the time and distance between the waves must be in the physiologically significant intervals must be 3 to 10%.

It has the limitation that it can be only used for cardiac diagnosis.

[4] FSV provide the improved analog of the properties of human visual system, it is available from 1D to 6D,its principle is to mirror the decision making process of expert, IQA database evaluate the viewable quality of the deformed picture with reference to original image. The quality of scale is then transformed into quantitative score.in this we are comparing the subjective score and metric score. We can improve FSV-2D by improving the parameter. Improved FSV-2D is then appealed to CEM data, resulting shows the decreasing accuracy with increase of frequency for the same mesh size. It has time complexity issues.[5] It introduces a new recent public key paradigm namely Signcryption which fulfils requirement of confidentiality as well as authenticity both messages between parties. Most efficient work is obtained with smaller cost significantly than the required signature encryption method. This method uses practical implementable method that is based on ID signcryption scheme which uses bilinear pairing that is presented. This is implemented within the scope of hardness of CDH i.e. Computational Diffie-Hellman hypothesis in a given principle form without any random model of oracle. Compares other relevant ID related signcryption schemes for evaluations and show satisfactory results. Thus it achieves confidentiality alongwith authenticity with the minimal computational cost. This methodology is limited to ID based signcryption makes it dependent and not introduced in practical real life.[6]It proposes a threshold key management method which is completely dynamic which is independent trusted party. It manages the MANET with functionalities of dynamically adding members, modifying and deleting a member also changing the values of threshold within scale of group. Adaption of distributed system creates a system key which is created with all the participants not by the most present likely in Key Generation Centers. In our system any of the participants is able to recover the system key however it can't reconstruct without any others involvement. Designated combiner verifies the correctness of secret shared at recovery level of system key. In similar fashion threshold polynomial based on ECC assures more secure and efficient method than traditional. With addition to this we also design new feasible bilinear pairing based signcryption method that provides more efficiency in both terms of computational complexity and communication load, also is aware of communication between two members. Limited to MANETs needs to be introduced in furthermore systems with more efficiency and enhanced security parameters.[7] Distributed key generation is a system which generates both public key and private key at a time in such a way that public key output is open while private key is passed amongst players in a secret way. It has been learnt for years and numerous protocols were introduced and majorly used in threshold cryptosystems and computing of distributed cryptographic. The given system majorly focuses on distributed generation of key in bilinear groups and introduces a rule on vector space access structures. A major issue in this system of DK Generation is it is completely based on vector space access structures is complex method.

[8] Introduces about Digital watermark, which is a currently unrealized ability of technology for protection of privacy and authentication of integrity in cloud computing domain. This study is specifically about Database authentication. It makes the use of order-preserving encryption scheme (OPES) coherently with discrete cos transformation (DCT), few watermarking technologies and cryptographic hash algorithms. With the limitations of Time complexity and cost.[9] Deduces that we will give an implementation and a sample of a prototype of the same architecture, which employs atotal general agreement algorithm on the first layer block chain. To improve

availability, integrity and scalability, we refine our solution by checking, Byzantine Fault Tolerant general agreement and a Hash Table that is distributed solution to generalize the first layer block chain ledger record among available nodes ensuring integrity. With the limitations as, its current performance limitations hinder actual exploitations.

[10] Narrates that Second-order SQL is comparatively much hard to break and also difficult than the first-order as the name itself suggests. Here we use a methodology based on ISR Instruction Set Randomization. This method randomize the trusted SQL keywords contained in Web applications to build new SQL instruction sets, and add a proxy server before DBMS, the proxy detects whether the received SQL instruction contains standard SQL keywords to find attack behaviour. This system can effectively detect 2^{nd}-order SQL injection attack and has a low processing cost compared to other techniques.[11] Signifies that, here we use the digital watermarking technology to guaranty the DB integrity where we get database content which is not misrepresented. Then we use SVR predictive method through which, obtained, is nothing but the characteristics of DB and then useof Huffman coding for encoding the same for compressing information of the functional part of the type of actual malicious software and not the part that spreads the malware to ensure the integrity. This is how tampering is detected. Here the limitations are , Time complexity and cost.[12]Introduces Avalanche Effect is a prudent property of Cryptographic hashing algorithm. This paper is about encryption in AES is 128bit by using Dynamic S-box, which relies on round keys. Dynamic S-box is more secured as compared to static S-box, as it makes challenging and more complicated for attacker to do any offline learning of an attack of one particular set of S-boxes. This algorithm leads to engender more secure block ciphers, solve the complications of the fixed structure S-boxes and will escalate the security level of AES block cipher system, with the limitation of every box is encrypted in the same way.

[13] Propose that, Hashing is the encryption- decryption algorithm. Hash value is a numeric value of a fixed length that uniquely identifies data. Hash values signifies large amount of data as reduced numeric value, so they are used with digital signatures. Hash values are also useful for validating the integrity of data directed through insecure channel. Hamming space is used to form the code or Hash value which symbolizes the data, which is transmitted through unshielded channel. The metric properties of the Hamming space is essential in determining basic concepts of coding theory such as error detecting and error correcting codes, with the limitation of time complexity.

3. PROPOSED METHODOLOGY

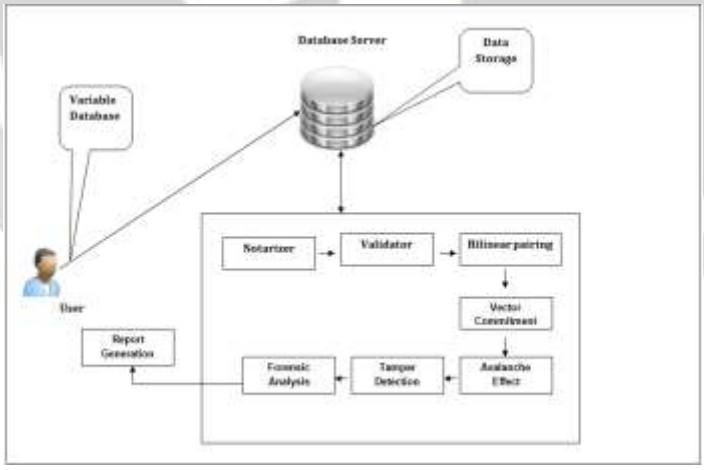

Figure 1: Proposed Sytem Overview

The proposed methodology for integrity violation of databases can be explained with the below mentioned steps.

*Step 1:*This is ths intial step of the system where clients are tranferring the data to store at the database storage sever where they are notarized with their authorization credentials like validation key formed due to MD5 hashing algorithm and random character selection.

Step 2: Here in this step a validation time is set , That is called as a Tile " T ".In this interval a iteration is been happening continuously to keep an vigil on database.

*Step 3:*On every iteration of validation period T, A two pairing hash tuple entities are created called previous and current.and Another pairing is created for the row data of the database tuple for past and current iterations.

*Step 4:*Here in this step a slight change in the hash keys are been measured for the past and present iteration pairings. Once if there is any change is find in the hash keys then the repsective tuple is considered as the tampered in between that instance and this process is known as alavanche effect.

As soon as the tamper is detected through avalanche effect row data tuples are subjected to the bitmap process. Where row data pairings are judged to find the some entities like

- ✓ Which id is been tampered
- ✓ What attributes are been tampered
- ✓ When did the tampered happened
- ✓ Who did the tampering

To find the culprit of the tampering systems sneaks into the database log file in regular interval to find the current login user status into the database.

Then in the end the system represented the whole forensic report in text file and stores in the desired location of the system by the admin.

The whole proposed system is expressed with the below Algorithm 1.

Algorithm 1: Tamper Detection using Tiled bitmap process
// Input: Data D

// Output: Successful identification of Tampering
Step 0: **Start**
Step1: Set T as tile (i.e. time)
Step 2: Initialize P_d as previous data
Step 3: Initialize C_d as current data
Step 4: Initialize P_{dh} as previous data hash
Step 5: Initialize C_{dh} as current data hash
Step 6: **WHILE TRUE**
Step 7: for each Tile T
Step 8: $C_{dh} \rightarrow C_d$ (C_{dh} is current data hash key from MD5)
Step 9: $P_{dh} \rightarrow C_{dh}$ (P_{dh} is previous data hash key from MD5)
Step 10: if C_{dh} != P_{dh}
Step 11: Generate Report for tampering
Step 12: Allocate $C_{dh} \rightarrow P_{dh}$
Step 13: **END WHILE**
Step 14: **Stop**

4. RESULTS AND DISCUSSIONS

The proposed methodology of tamper detection of database is implemented using java technnology with netbeans 6.9.1 as IDE and Mysql server 5.0 as database server in both web and stand alone paradigm. For hosting web application proposed model used apache tomcat as web and application server and dreamweaver 8.0 as web designing tool.

We performed an experiment to evaluate the time required for data tampering detection on increasing of several numbers of tuples as shown in table 1.

No of Tuples	Time in Milli seconds
50	175
100	258
150	302
200	398
250	470
300	552

Table 1. Time required for different number of tuples

On observing this output we plot a graph as shown in figure 2.

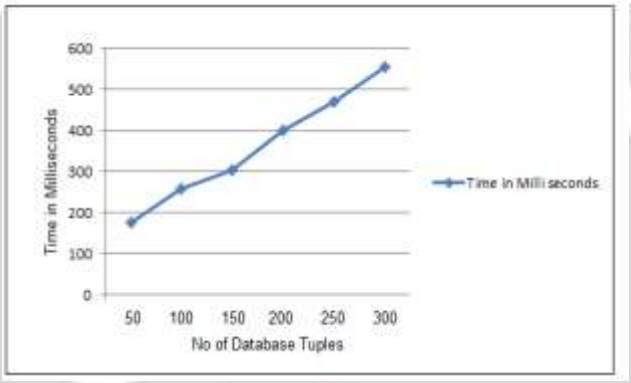

Figure 2: System performance based on increasing number of database tuples.

It is been observed that as the number of tuples in database increases the time taken for forensic analysis is not increasing by directly proportionate. So it clearly indicates that system performance is better as database tuple increases. This is exactly due to parallel computation of synchronized detection of tampering in database in object level.

5. CONCLUSIONS AND FUTURESCOPE

Due to increasing in outsourcing data storage at warehouse in drastic speed, So threat to the data is also increasing like anything from inside intruders. Our proposed method implements anti data tampering technique using tiled bitmap process which is powered with the avalanche effect concept.

On observing the performance of the system it clearly indicates that this process is very effective in handling database tampering and recovering system for huge size of the databases.

In future this system can be enhancing for even more accuracy using the distributed system for big data systems.

6. REFERENCES

[1] Momin, BashirahamadFardin. "Clustering and Validation for Very Large Databases (VLDB)." *Information and Automation, 2006.ICIA 2006.International Conference on*.IEEE, 2006.

[2] Goyal, A., Swaminathan, A., Pande, R., & Attar, V. (2016, April). Cross platform (RDBMS to NoSQL) database validation tool using bloom filter. In *Recent Trends in Information Technology (ICRTIT), 2016 International Conference on* (pp. 1-5).IEEE.

[3] Laguna, P., Vigo, D., Jane, R., &Caminal, P. (1992, October). Automatic wave onset and offset determination in ECG signals: validation with the CSE database. In *Computers in Cardiology 1992, Proceedings of* (pp. 167-170). IEEE.

[4] Zhang, G., Orlandi, A., & Duffy, A. P. (2018). Using image quality assessment (IQA) databases to provide an appraisal of the ability of the feature selective validation method (FSV) to compare two-dimensional datasets. *IEEE Transactions on Electromagnetic Compatibility*, *60*(4), 890-898.

[5] Karati, A., & Biswas, G. P. (2016, September). A practical identity based signcryption scheme from bilinear pairing. In *Advances in Computing, Communications and Informatics (ICACCI), 2016 International Conference on* (pp. 832-836). IEEE.

[6] Meng, X., & Li, Y. (2012, August). A verifiable dynamic threshold key management scheme based on bilinear pairing without a trusted party in mobile ad hoc network. In *Automation and Logistics (ICAL), 2012 IEEE International Conference on* (pp. 315-320). IEEE.

[7] Zhang, J., & Zhang, F. (2013, September). Secure Distributed Key Generation on Vector Space Access Structures in Bilinear Groups. In *Intelligent Networking and Collaborative Systems (INCoS), 2013 5th International Conference on* (pp. 803-808). IEEE.

[8] Xiang, S., & He, J. (2017). Database authentication watermarking scheme in encrypted domain. *IET Information Security*.

[9] Aniello, L., Baldoni, R., Gaetani, E., Lombardi, F., Margheri, A., &Sassone, V. (2017, September). A prototype evaluation of a tamper-resistant high performance blockchain-based transaction log for a distributed database. In *Dependable Computing Conference (EDCC), 2017 13th European* (pp. 151-154). IEEE.

[10]Ping, C. (2017, December). A second-order SQL injection detection method. In *Technology, Networking, Electronic and Automation Control Conference (ITNEC), 2017 IEEE 2nd Information* (pp. 1792-1796). IEEE.

[11] Wu, H. C., Hsu, F. Y., & Chen, H. Y. (2008, November). Tamper Detection of Relational Database Based on SVR Predictive Difference. In *Intelligent Systems Design and Applications, 2008. ISDA'08. Eighth International Conference on* (Vol. 3, pp. 403-408). IEEE.

[12] Balakirsky, V. B. (2005). Hashing of databases with the use of metric properties of the Hamming space. *The Computer Journal*, *48*(1), 4-16.

[13]Nejad, F. H., Sabah, S., & Jam, A. J. (2014, August). Analysis of avalanche effect on advance encryption standard by using dynamic S-Box depends on rounds keys. In *Computational Science and Technology (ICCST), 2014 International Conference on* (pp. 1-5). IEEE.