A VARIABLE DATABASE INTEGRITY MAINTENANCE SYSTEM

Shruti Yadav[1], Kajal Zite[2], Vaibhavi Pande[3], Vaishali Bhorde[4]

[1,2,3] *Student, Computer Engineering, ICOER, Maharashtra, India*

ABSTRACT

Maintaining database at the application developer's end needs too many parameters like lease line, servers, licensed software's and many more. And this is a costlier affair that's why most of the application developers never rely on that. So, to overcome this financial and technical glitches application developers outsource their database storage headache to a third party. Where third parties always have high end infrastructure and sophisticated settings to store the data. The integrity of the data has always been a headache, most of the time this integrity is facing danger from the internal employees of the organization at the third party's end. Some methodologies are existed to tackle this, where they are handling the tampered data tuples and restore them. But they are never going in depth to analyze the facts to find the real culprit, and also, they often suffered with the greater time complexity. So proposed methodology put forwards an idea of handling the variable databases using the recursive bilinear paring and Avalanche Effect. And this process is catalyzed by the techniques of notarization and validation to detect the parameters like who, when and what.

Keyword: *- Database Integrity, Bilinear Pairing, Avalanche Effect, Validation, Notarization*

1. INTRODUCTION

Database integrity is the integrity of a particular database as maintaining the integrity of the database enables the users to utilize the benefits the database offers in a much better way. This leads to a better user experience, as the maintaining the integrity would retain the valuable data that is being stored in the database. Any loss in the integrity of the database would lead to a significant loss of data.

Database integrity should be managed in both the types of databases, such as relational databases and hierarchical Databases. More importantly it is used with relational databases as they are a lot more abundant and prone to integrity issues as they are used more often and that would lead to loss in the integrity of the data. there are different types of integrities that are responsible for the overall health of the database. The various types of integrities in a relational database are as follows.

Entity Integrity is useful for maintaining the overall integrity of a relational database, this means that the entities that utilize the database must be managed efficiently. Entity integrity is maintained when the keys that are used to perform the operation must always be distinct as duplication of data would happen and would lead to a loss of integrity. The primary key cannot be therefore, null to avoid damaging the integrity.

Relational integrity is one of the integrities that are essential for the maintenance of the total integrity of the database. This is essential as the primary key that is used to maintain the database should be unique in the sense that I could be used to defined the relationship between the elements of a database uniquely. Therefore, the foreign key cannot be null as it would lead to the value of the primary key to be null and this cannot be sustained.

Domain Integrity is also an essential aspect in maintaining the integrity of the database as it ensures that the columns and rows of the database are well connected and have a proper relationship between them. The domain

integrity ensures that all the columns in the database are explicitly defined in a particular domain. All the columns when they are defined in a specific domain, the relational database is said to have a very high degree of integrity.

As there have been increasingly more security issues that are surfacing every day, it is imperative to address the growing concerns. Encryption is the standard for maintaining the security since a long time. People have utilized the ciphertext even before the advent of computers to hide the information from everyone except for the intended receiver. Therefore, Encryption is the preferred choice for securing information and data on a computer.

Encryption is of varying degrees and its strength ascertains its reliability. There have been various encryption techniques that have been developed to safeguard the data, some of them more powerful than the others. Therefore, we need an actual quantifier that can present the strength of the Encryption standard in use.

The Avalanche effect is one of the techniques that can be used to determine the strength of the encryption. The Avalanche effect refers to the corresponding exponential change in the cipher text that is introduced with a minor change in the plaintext. The amount of change in the ciphertext determines the how powerful the encryption algorithm is, as changing one bit would usually result in change of a significant number of bits in the ciphertext. This effect is effectively known as the Avalanche Effect and it can be used to determine the strength of the encryption algorithm.

This research paper dedicates section 2 for analysis of past work as literature survey, section 3 deeply elaborates the proposed technique and whereas section 4 evaluates the performance of the system and finally section 5 concludes the paper with traces of future enhancement.

2. LITERATURE SURVEY

A. Alsharif [1] Database is one of the important and valuable assets. Usually, the database is very complex for developers. In this paper, DOMINO is a self-regulating technique that creates test data according to coverage criteria for integrity constraint. DOMINO use the trimmer to domain-specific operators to easily generate test data for relational database schemas. In this paper 34 relational database schemas hosted by three database management systems. DOMINO method is the best and fastest method, then the state-of-the-art search-based method, it also detects more faults.

T. Chen [2] To maintain the data is a very hectic task, then after the emergence of the cloud storage and big data, data task has come more complex and hectic too. The modern system, cloud computing big application such as e.g. Amazon, Google is depending on the database management system for database integrity. The Security Database is an important factor nowadays. In this paper, database security is maintained by the ORM i.e. Object Relational Mapping. They have proposed several techniques to detect the ORM code, configure ORM code efficiently and they monitor the execution of the system continuously and if there is any problem in a transaction it will report.

T. Salma [3] It has been a challenge to handle large data in the database management system. When there is a database the security of database arises. The data management system uses black boxes develop by the developers which provide the best adoption technique and abstraction for different database technologies. This paper presents a reliable and productive explicit dynamic support, including block update, delete and append. They have used token in distribution and verification of coded data, this technique achieves data error localization wherever the data corruption has been detected. Third party audits are also done where the user can check the integrity task by the third-party auditors.

S. Sundari [4] secure multiparty computation (SMC) is used on a larger scale nowadays in which data is distributed between different parties. Most of the companies tried to unite with different organizations for their mutual benefits and SMC help them to gain information from the large data. In this technique,two-phase validation is used for the authentication purpose where someone can cause unauthorized access to stored data. Data can crash due to any malfunctions. The proposed system is to check the integrity from different data sets and to protect them from different Distributed Data Integrity.

C. Wu [5] In this paper multi-level tamper detection technique is used for tamper detection such adjacent block based, intra-block detection and inter block-based detection. In earlier researches, they have used one parity check watermark, but in this paper, the author as used two copies of feature bit watermark for authentication purpose. Simple self-recovery fragile water is also imported in this paper. The Tamper detection performance using the watermark technique is better than the other technologies.

V. Mall [6] It is not possible every time to observe a structural change in images with eyes. Image editing software is easily available in the market so it has made the digital image processing very popular. This software is easily available on the internet in such case we cannot take risk regarding the integrity of the image. They have used the hash generation technique using singular value decomposition. The proposed method is very sensitive it keeps on examine every structural tampering.

Z. Shuxu [7] RFID is a "radio frequency identification" where the digital data is encoded in RFID tags or smart labels. RFID works on different technologies such as asset tracking ID, Badging, supply chain management. But RFID tags can tamper very easily. In this paper integer, piecewise linear chaotic tent mapping is used to improve watermarking generating and embedding algorithm with the safety of data. Researchers have come forward with lot method regarding tampering detection, but this method assigns to be a better one then the others.

C. TUNG [8] Camera Tamper detection Model is used for camera tampering process. In recent years the Video Surveillance has become one the important sector to be researched Nowadays the camera is installed everywhere for reason of security and in most of the place the camera eye is replaced with the human eye. In cameras, brightness and the crime are the two things which adjust the pixel in the background and handle the problems, light, and shadow. Thus, the CDTM reactive and creates the clusters and classifies the pixels.

W. Zhang [9] Key agreement is used for creating the key generation in the process of cryptography and for the communication purpose. Key agreement isn't safe when it comes to security.

To solve this security problem the key agreement method is merged with bilinear pairing to make it strong. These two techniques of key agreement one entity that creates the key and sends to another end safely i.e. are called a key sharing. The other one is a public key and the private key which passed publicly. Both methods are safe.

G. Xu [10] Digital Signature means a blind signature where the person gets to sign the part of the message without revealing the message or the message of the party. There are numerous applications in the blind signature. One of the famous applications of the blind signatures is e-cash where the transaction is done online, but the bankhas no idea about the user and the blind signature is also used in e-votes too. For making this blind signature is joined to bilinear pairing for the security reason.

S. Weimin [11] Signcryption it is a technique in which the message is encrypted in a single logical step. In this proposed method of secure multi-party multiplication, they have used Bilinear pairing. Signcryption is based on the bilinear pairing which makes the security strong and also the public verifiability. In stage 2 of signcryption three multiplications at the pointare done and one-time bilinear pairing operation is performed. Then the two-party multiplication is done using the signcryption protocol to secure multi-party multiplication

J. Gavade [12] In recent years there is a transformation of video through social media has become easy, anyone can tamper with video or edit the video and society believes in it blindly. Sometimes video is edited or tampered intentionally to convey the wrong message in public. In this proposed paper the two-step algorithm used to detect the tampered video. First, the video as to be detected which is tempered and then it is compared to the others frame to find the duplicate frame. In the later part, the work is extended to various attacks such as spatial and spatiotemporal.

Vinod Mall [13] In recent years and from the past few years there is much work done in the image processing community. The images are very important when it comes to biometric images, evidence in court or in the police investigation. In this paper hashing technique is used to check or to detect multiple image tampering is done or not. The previous researches were made on the single tampering means whether the image is a tampered not, but in this, it is improved to the multiple level tampering whether the image tampers more than one time.

3. PROPOSED METHODOLOGY

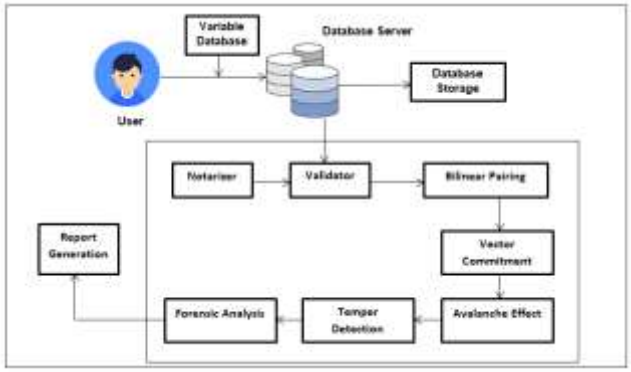

Fig 1: System Overview

The proposed methodology of variable database integrity maintenance system can be depicted with its step-in figure 1. And these steps are deeply narrated in the below mentioned steps.

Step1: Notarizer - This is the very first step, where each and every database client who is willing to store the database at the third-party organization are getting a unique notarization key based on their attributes. This key is being generated by the random selection of seven characters through MD5 hash key, which is generated by the client's attributes. And this key is helping to authenticate the client while he is uploading the data to the cloud.

Step 2: Validator - Once the data is stored in the third-party servers, then a validation of the data is started for the stated period, like it may be 1,2,3, or 60-minute period. Here in each validation time a current and previous data vectors are maintained to get the Tamper Detection results.

Step 3: Bilinear Pairing and Avalanche Effect - This is the Core part of the system where, for the given validation time a pair of hash keys are generated for the each and every database tuple and they are referred as the bilinear pairs. These bilinear pairs are compared for the integrity loss of the database tuples. A change in the single bit of the data tuples yields the massive change in the hash key, which is referred as the Avalanche effect.

This Avalanche effect helps to identify whether the database tuples are tampered or not. If the data base tuples are tampered then the primary key of the tuples are extracted as tampered ID. This ID eventually represents the which tuples are being targeted by the database attacker. This Process can be depicted with the below shown algorithm 1. Once the ID is detected for the tampering through the avalanche effect of the hash keys then, the remaining attribute's integrity is being measured through comparing the original data tuples list of the past and current thread of the bilinear pairs, this eventually yields the details of the tampering processes.

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

Algorithm 1: Tamper Detection using Tiled bitmap process

// Input : Database D_B

// Output: Tamper set T_{SET}

Function: tamper detection (D_B)

Step 0: Start

Step 1: PD_{SET} =Ø, CD_{SET} =Ø

[PD_{SET} : Previous DB Set , CD_{SET} : Current DB Set]

Step 2: PDH_{SET} =Ø, CDH_{SET} =Ø

[PDH_{SET} : Previous DB Hash Set , CDH_{SET} : Current DB Hash Set]

Step 3: CDSET → getDatabaseList

Step 4: CDH_{SET}→ Hashset of CD_{SET}

Step 5: ***while TRUE***

Step 6: WAIT FOR **T** [Tile: Time]

Step 7: $PD_{SET} \rightarrow CD_{SET}$

Step 8: $PDH_{SET} \rightarrow CDH_{SET}$

Step 9: ***for*** i=0 **to** size of CDH_{SET}

Step 10: **IF** $CDH_{SET} \neq PDH_{SET}$ **THEN**

Step 11: check CD_{SET} and PD_{SET} for Details

Step 12: Generate Report $\mathbf{G_R}$

Step 13: $T_{SET} = T_{SET} + \mathbf{G_R}$

Step 14: **End *for***

Step 15: **End *While***

Step 16: ***return*** T_{SET}

Step 17: Stop

Step 4: Tamper Detection And Forensic Analysis- Here in this step the culprit is going to nab using the recursive surveillance on the database log file, Which is in the form of an XML where logs of the database user is traced out by string handling of the XML and correlating his illegal activity with the current and the previous bilinear activities.

Eventually by doing this proposed model successfully got the all the details of database tampering like Who did the Tampering? When did the Tampering? On what attributes tampering was happened? Once all these parameters are collected, then a proper report is generated to deliver to the admin.

After this process the previous string of the bilinear pair is restored in the database for the tampered ID by updating its all other attributes to get the original database tuples.

4. RESULT AND DISCUSSION

The proposed methodology of Variable database integrity maintenance system is developed using the Java programming language by using NetBeans as IDE and MYSQL as the database. Proposed model uses the windows machine with a processor of Core i5 and primary memory of 6GB. Some experiments are being conducted to measure the impact of the proposed model with some experiments as mentioned below.

RMSE is measured as the error difference between the expected and the obtained values. Which eventually provides the best results for the database tamper detection system. And RMSE can be measure with the following equation 1.

RMSE= $\sqrt{(xp - xo)}^{2}$

Where

Xp- Predicted number of tampered tuples

Xo - Obtained number of tampered tuples

The Result of the RMSE is completely astonished as the proposed model identifies each of the tuples which are tampered and gives a complete report as the experimental data is accumulated in the below given table 1. Table 1 represents the RMSE of our system is zero as it identifies all the possible tampered data details and provides proper results. This shows the effectiveness of the proposed model where it yields excellent results for the variable computation of the tampered data in all scenarios of tampering.

Experiment No	Total No of Data in Database	Total No of Tampered Data	Total No of Correctly Detected Tampered	RMSE
1	100	5	5	0
2	200	9	9	0
3	300	12	12	0
4	400	19	19	0
5	500	22	22	0

Table -1: RMSE Measurement Table

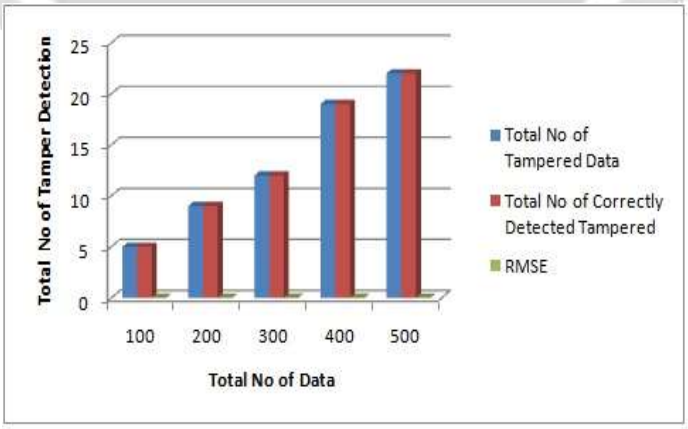

Fig -2: RMSE evaluation Results

5.CONCLUSION

This research article studies most of the existed database tamper detection models and analyzes their working pattern and finds some of the models only identify the tampered tuples, but never gone deep enough to analyze all the facts of the tampering. So, this paper uses some methods like Bilinear pairing along with the avalanche effect to effectively perform the forensic analysis of tampering. The result of this can be shown in previous section of Results and discussions where RMSE is measured for the expected outcome v/s obtained outcome. In this RMSE the proposed model performs excellently to yield RMSE of zero, which is again the best outcome of the proposed model.

This proposed model can be enhanced in the future for the big data to perform on the real time database servers using distributed computing.

6. REFERENCES

[1] Alagukumar, S and Lawrance, R, "Classification of Microarray Gene Expression Datausing Associative Classification", International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16), 2016.

[2] Liankuan Zhang, Chunlei Xia, Deqin Xiao et al, "A leaf vein detection scheme for locating individualplant leaves", International Conference on Information and Communication Technology Robotics, 2018.

[3] Ye Cao, Chenjun Yan, Jingyue Li,and Hong Zhou, "Leaf vein extraction and angle measurement using hue information and line detection", 9th International Conference on Intelligent Human-Machine Systems and Cybernetics, 2017.

[4] MelikeSardogan, AdemTuncer and YunusOzen, "Plant Leaf Disease Detection and Classification basedonCNN withLVQAlgorithm", 3rd International Conference on Computer Science and Engineering, 2018.

[5] KentaroFukuta, Tomomasa Nagashima and Yoshifumi Okada, "LEAF: leave-one-out forward selection methodfor cancer classification using gene expression data", IEEE/ACIS International Conference on Computer and Information Science, 2010.

[6] AzeilLouisseCodizar and Geoffrey Solano, "Plant Leaf Recognition byVenation and ShapeUsing Artificial Neural Networks", 7th International Conference on Information, Intelligence, Systems & Applications (IISA), 2016.

[7] MuhdSafarudinChek Mat, Jezan Md Diah, Mokhtar Azizi Mohd Din & Abd. Manan Samad, "Data Acquisition and Representation ofLeaves using Digital Close-RangePhotogrammetry for Species Identification", IEEE 5th Control and System Graduate Research Colloquium, Aug. 11 - 12, UiTM, Shah Alam, Malaysia, 2014.

[8] Manojkumar P., Surya C. M. and Varun P. Gopi, "Identification of Ayurvedic Medicinal Plants byImage Processing of Leaf Samples", Third International Conference on Research in Computational Intelligence and Communication Networks, 2017.

[9] Xiaohan Yu, ShengwuXiong, Yongsheng Gao, Yang Zhao,andXiaohui Yuan, "Multiscale Crossing Representation Using CombinedFeature of Contour and Venation for Leaf ImageIdentification", International Conference on Digital Image Computing: Techniques and Applications (DICTA), 2016.

[10] Karthik. G and Praburam. N, "Detection and Prevention of Banana LeafDiseases From Banana Plant Using EmbeddedLinux Board" Online International Conference on Green Engineering and Technologies, 2016.

[11] Arya M S, Anjali K,andMrs.DivyaUnni, "Detection of Unhealthy Plant Leaves UsingImage Processing and Genetic Algorithmwith Arduino", International Conference on Power, Signals, Control and Computation (EPSCICON), 2018.

[12] M. Merchant, V. Paradkar, M. Khanna and S. Gokhale, "Mango Leaf Deficiency Detection Using DigitalImage Processing and Machine Learning", 3rd International Conference for Convergence in Technology (I2CT)The Gateway Hotel, XION Complex, Wakad Road, Pune, India. Apr 06-08, 2018.

[13] Zulkifli Bin Husin, Abdul Hallis Bin Abdul Aziz, Ali Yeon Bin Md Shakaff and Rohani Binti S Mohamed Farook, "Feasibility Study on Plant Chili Disease Detection Using Image ProcessingTechniques", Third International Conference on Intelligent Systems Modelling and Simulation, 2012.