ISSN No:-2456-2165

Discovering Database Tampering through Blockchains

[1.]Dhanika Tannu, Computer Dept, G. H. Raisoni Institute of Engineering & Technology, Pune
[2.]Yadnya Bharambe , Computer Dept, G. H. Raisoni Institute of Engineering & Technology, Pune
[3.]Aachal Khambekar, Computer Dept,. G. H. Raisoni Institute of Engineering & Technology, Pune
[4.]Madhura Patil, Computer Dept, G.H. Raisoni Institute of Engineering & Technology, Pune
[5.]Snehal Kulkarni, Computer Dept. G. H. Raisoni Institute of Engineering & Technology, Pune
[6.]Anita Mahajan, Computer Dept. G. H. Raisoni Institute of Engineering & Technology, Pune

Abstract:- The volume of the data being stored across the world has been increasing consistently over the past few years. The data that is being received needs to be stored effectively to help understand and also manage the large amount of data. For this purpose, the Relational Databasesare being used to store this data in a row and column based format. This is one of the most common as well as one of the most effective forms of data storage. The RDBMS or Relational Database Management Systems have been one of the most efficient and widely used management systems for the management and storage of the data. The popularity of these systems make it a valuable target by the attackers to achieve unapproved and malicious access into this database. This attack on the RDBMS can cripple systems and lead to a largescale loss of data. Therefore, to provide a solution to this problem and effectively improve the security and forensic analysisof the RDBMS, the approach is being presented this research article which utilizes the blockchain approach to facilitate the forensic report generation using Bilinear Pairing and the detection of the avalanche effect.

Keywords:- *Database Integrity, Bilinear Pairing, Avalanche Effect, Validation, Notarization, Blockchain.*

I. INTRODUCTION

Data fulfils an integral role in computing across the world. The data is one of the most valuable resources that have been crucial for the development and advancement of thehuman race as a whole. The effective preservation of the datais one of the most critical tasks that are performed by the entirety of the human race over the course of our existence. This is essential due to the nature of the data as one the most effective forms, allowing for the communication and improvement in various aspect of the human life.

The data has been preserved in many forms and formats over the course of the human race. The earliest humans had limited resources and limited availability of materials or formats to share their knowledge and keep it preserved for long periods of time. The earliest data storage was in the form of word of mouth as the data was stored as information in the brains of the earliest humans. The human brain is powerful and has the ability to store and interface a large amount of databut it is limited with recall and has a high chance of losing the information over the course of time.

There have been several different data storage mechanisms, such as cave paintings, which have been utilized extensively but were limited in its storing capabilities and have not been that long lasting. This has been the case for storing large amounts of data, where human beings have been trying to find effective techniques for the data storage. With the introduction of books and their effective storages and the printing press allowing for copies being made consistently, the society could advance faster and effectively be contributing to the collective knowledge stores and enable a far better improvement in the living standards.

The introduction of the electronic media has enabled fora far better and data dense forms of storage that have been accessible and being used across the world. The data being generated has also been increasing exponentially every year. The data being generated across the world is usually being stored in the relational databases. These databases store the data in the form of tables in a row and column format. This makes it easier for the retrieval and searching of the effectively.

These relational database management have been highly popular in the recent years as it allow for the effective and precise management of the data. The RDBMS stores the data in the form of tables and containing some kind of relationship between the various attributes of the data. This is how most of the data is being stored by individuals as well as major corporations and organizations across the globe. This is the reason why these databases are targets of attacks and other intrusions as the data can be valuable to these attackers.

The attackers take undue advantage of the fact that all the important and confidential data is being stored in a single location that can be targeted to achieve their nefarious goals. This is one of the most effective and highly popular attacks that perform intrusion into these databases to gain access and steal and manipulate the data. This is one of the biggest problems faced by various organizations where the confidential and personal data of their clients is at risk that canput the whole company in imminent danger. Therefore, thereis a need for an effective approach that can provide protection and maintain the integrity of the RDBMS.

ISSN No:-2456-2165

The Literature Survey chapter of this research paper examines previous work. Section 3 delves into the approach indepth, while section 4 focuses on the outcomes evaluation. Finally, Section 5 brings this report to a close and gives some hints for future research.

II. LITERATURE SURVEY

The research of R. Awadallah et al. [1] focuses on cloud computing data breaches and the all-encompassing cloud service provider power over client data activities. The authors provide a method for enhancing the client's capacity to secure data. To offer data security and privacy during outsourced computations, the proposed technique uses homomorphic encryption. A novel solution depends on a distributed network of cloud service providers and Byzantine Fault Tolerance consensus is developed to secure data integrity and detect data tampering from the cloud service provider itself. There is no requirement for a direct connection between the various cloud service providers in the proposed technique. Cloud service providers must create master hash values of their databases and store them on blockchain networks, such as Bitcoin or Ethereum, to offer clients immutable verification data. To meet the needs of different clients, the authors supplied a quantitative study of overhead expenses based on multipletime possibilities.

G. Heo et al. upgraded DRM and digital fingerprinting tocombat unlawful digital content copying and leakage, and theyalso used blockchain to combat profit sharing, forgery, and falsification. However, blockchains are space constrained and cannot store enormous amounts of digital data. Because transactions, including their contents, are dispersed and held ineach peer's storage, wasted storage space and poor privacy protection are also crucial. To provide safe and dependable digital content trade, the proposed SBBC system incorporates off-chain and on-chain components [2]. The authors presented the WBFT consensus method, which calculates consensus weights depending on whether or not users are authorized, increasing dependability even further. Therefore, SBBC provides security and dependability by fixing and repairing current problems that have escalated into major concerns.

S. Rouhani et al. developed a permissioned blockchain-based end-to-end data trust system. The quality of input data isassessed using a unique trust model that includes the data owner's reputation, endorsements, and confidence in the given data [3]. Therefore, data consumers guarantee that the quality of accessible data sets has been evaluated and updated asneeded. Data owners benefit from safe, transparent, and automated access management managed by smart contracts in the proposed architecture. Data owners have total control over their data assets, and they are the only players in the system that can govern access rights without the help of third parties. By utilizing blockchain's provenance and audibility characteristics, data owners may additionally monitor and trace access regulations and alterations to their data assets.

M. I. Sarwar et al. presented a method that can let businesses save their financial data in Blockchain, or so-called Data Vaults, and chose the trial balance for implementation [4]. The suggested framework can assure data integrity in AIS or ERP systems, and if data is hacked at any point, breaches may be recognized quickly using the proposed approach. The proposed technology also verifies the accuracy of data submitted for commercial transactions or other comparable objectives. The authors looked at a stripped-down version of Blockchain that is several times lighter than the one used in Bitcoin. The suggested architecture is built on a mix of Blockchain and databases, and it may be used in a variety of fields.

R. Awadallah et al. proposed a client self-verification technique depending on one of two blockchain-based relational database systems: agile BC-based RDB or secure BC-based RDB. By adding more properties to chain records based on SHA-256, the suggested systems built the cloud relational database structure. Both also imitate decentralization by dispersing the created database over at least four cloud service providers. When a client submits a query, the cloud service providers must update their database and link the new record(s) to the prior chain of records [5]. When the customer asks for it, they create an RDB signature to deliver to confirm that they agree on the same outcome. The evaluations revealed that the flexible BC-based RDB system is affordable and losesjust a little amount of energy, up to 1 joule. Therefore, this technique has demonstrated its use in high-throughput databases.

For relational databases, J. Lian et al. presented TDRB, ablockchain-based tamper-proof detection middleware. The authors achieved the primary aim of tamper detection within the relational database by utilizing blockchain, hash checksum, symmetric encryption method, caching database, and other technologies. Experiments show that the TDRB middleware can concurrently balance performance and detection accuracy [6]. Overall, the TDRB middleware suggested in this research provides benefits such as platform crossover, extensibility, and minimal implementation costs. The middleware aids in the protection of relational databases, the prevention of internal manipulation, and the security of system-sensitive data.

The research by Y. Liu et al. addresses all of the technical elements of constructing a national health data centerwhile maintaining security. The suggested blockchain infrastructure connects traditional e-healthcare systems. The synchronization issues between blockchain and traditional EHS are addressed by the author's suggested Triple approach. The proposal's efficiency is demonstrated by several algebraic equations [9]. Furthermore, the success of the proposed strategy is demonstrated by the implementation outcomes and discussion. Furthermore, a protocol-independent, plug-and play interoperable blockchain-based EHS system can take thecontribution one step further.

ISSN No:-2456-2165

III PROPOSED METHODOLOGY

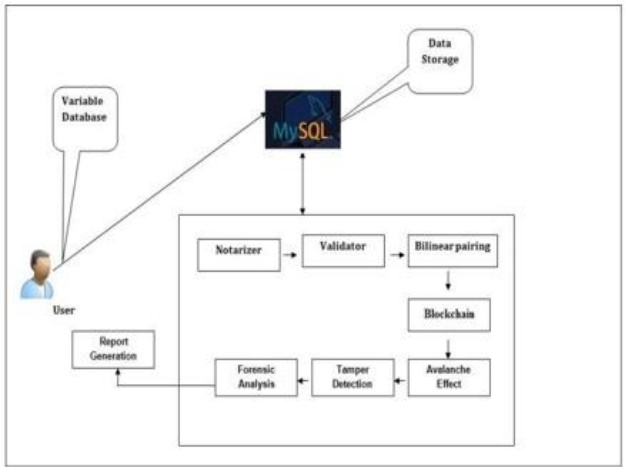

Fig 1: System Overview

The proposed methodology for the maintenance of the integrity of a Relational Database is displayed in figure 1 given above. There are a number of sequential steps for the fulfillment of the proposed approach which have been listed below.

Step 1: Notarizer – This is the initial phase, whereby each database client that desires to save their database with a third-party entity is given a unique notarization key based on their attributes. The MD5 hash key, which itself is created by the client's characteristics, generates this key by selecting seven characters at random. This key is used to let the client authenticate himself while uploading data to the cloud.

Step 2: Validator – After the data is saved on third-party servers, validation of the data begins for the specified duration, which can vary from 1 minute to 60 minutes. To acquire the Tamper Detection insights, the time, current, and prior data vectors are preserved in each validation.

Step 3: Blockchain Formation: The data is recorded in the database after it has been validated. Every characteristic ofthe Row is merged to generate a String during database storage to acquire a Hash key utilizing the MD5 cryptographichash function. The head key is the name given to such hash key. The block's subsequent Hash key is created by concatenating this Current hash key with another row. The hash keys preserve a chain of blocks created by this continual operation. The Terminal key is the final blockhead key acquired and should be saved as the distinct validation key. This Terminal Key will be used to execute the bilinear pairing approach described in the following step to discover manipulation in the database.

Step 3: Bilinear Pairing and Avalanche Effect – This is now the key component of the mechanism, where each database tuple is assigned a combination of hash keys for the particular validation period, which are described as the bilinear pairs. The integrity loss of the database tuples is assessed using these bilinear pairings. The Avalanche effect occurs when a single bit of the data tuples is changed, resulting in a significant variation in the hash key.

The Avalanche effect aids in determining whether or not database tuples have indeed been manipulated with. If the database tuples have been tampered with, the main key is retrieved and used as the tampered ID. The database intruder will ultimately use this ID to identify which tuples are being affected. The algorithm 1 below can be used to describe this process. Again when the tampering of the ID has been identified using the avalanche effect of the hash keys, the remainder attribute's integrity is determined by contrasting the original data tuples list of the previous and present thread of the bilinear pairs, which finally reveals the tampering operations specifics.

Step 4: Tamper Detection And Forensic Analysis – The accused will be apprehended in this stage by performing recursive monitoring on the database log file, which would be in the form of an XML, in which the database user's logs are tracked down out by string handling of the XML and corresponding his illicit activity with the present and former bilinear operations.

Finally, using this proposed methodology, we were able to get all of the information of database manipulation, such as who committed the tampering. When did the tampering take place? What qualities were tampered with? After all of these characteristics have been gathered, a complete report is prepared and delivered to the administrator.

The prior string of the bilinear pair is reinstated in the database for the compromised ID by changing all other characteristics to retrieve the original database tuples after this operation.

IV RESULT AND DISCUSSION

The proposed approach for maintaining the integrity of a Relational Database is written in Java, with NetBeans as the integrated development environment and MYSQL as the database. The proposed concept employs a Windows computer with a Core i5 CPU and 6GB of primary storage. Some observations, such as those listed below, are being done to examine the effectiveness of the proposed scheme.

RMSE is calculated as the error discrepancy between both the expected and the obtained values. This gives the optimum conclusions for the database tamper detection technique for eh evaluation of the systems performance. And RMSE may be calculated using the equation 1.

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

Where
Xp- Predicted number of tampered tuples Xo - Obtained number of tampered tuples

The RMSE result is absolutely astounding since the suggested model detects each one of the tuples that are tampered and provides a comprehensive report based on the experimental observations in Table 1. Table 1 shows that our system's RMSE is zero since it detects all manipulated data

ISSN No:-2456-2165

elements and produces accurate findings. This demonstrates the efficacy of the suggested methodology, which produces great results for the dynamic calculation of altered data in all tampering circumstances.

Experiment No	Total no of Database Tuples	Total no of Tampered Tuples	Total no. of Tampered Tuples Detected	RMSE
1	100	6	6	0
2	200	10	10	0
3	300	13	13	0
4	400	20	20	0
5	500	23	23	0

Table 1: RMSE Measurement Table

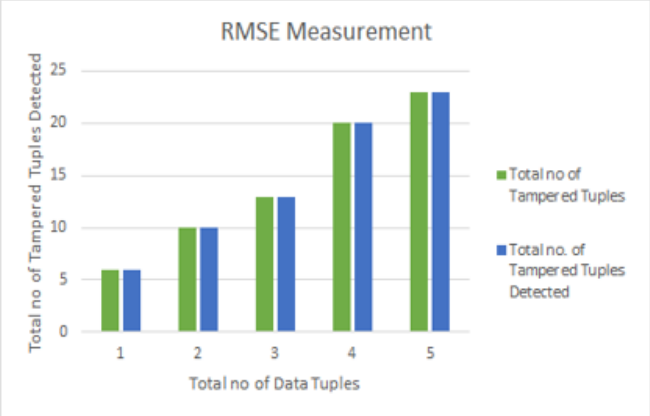

Fig 2: RMSE evaluation Results

V CONCLUSION AND FUTURE SCOPE

The presented approach for the realization of the forensicanalysis of any tampering being performed on the RDBMS has been stipulated in this research article. The most utilized forms of databases around world are primarily the RDBMS. The Relational Databases are easier to maintain and can be scaled indefinitely to achieve larger and larger storages of the data. This data is easier to query and retrieve which can be a blessing larger databases. This usability also put the RDBMS at risk as individuals with malicious intents would like to gain access into the database to fulfil their nefarious deeds of to perform data leaks and other crimes. Therefore, it is imperative to achieve an effective approach for safeguarding the data along with the realization of a forensic report todetermine the exact parameters of the tampering performed on the database. Therefore, this research approach has defined an effective methodology that deploys the Blockchain framework along with bilinear pairing and Avalanche effect detection, following which a report is generated for the user. The approach has been effectively evaluated for any error in the tamper detection which has resulted in highly positive outcomes.

For the purpose of future research the approach can be effectively implemented in a real time scenario on actual database servers through the implementation of distributed computing.

REFERENCES

[1]. R. Awadallah, A. Samsudin, J. S. Teh, and M. Almazrooie, "An Integrated Architecture for Maintaining Security in Cloud Computing Based on Blockchain," in IEEE Access, vol. 9, pp. 69513-69526, 2021, DOI: 10.1109/ACCESS.2021.3077123.

[2]. G. Heo, D. Yang, I. Doh and K. Chae, "Efficient and Secure Blockchain System for Digital Content Trading," in IEEE Access, vol. 9, pp. 77438-77450, 2021, DOI: 10.1109/ACCESS.2021.3082215.

[3]. S. Rouhani and R. Deters, "Data Trust Framework Using Blockchain Technology and Adaptive Transaction Validation," in IEEE Access, vol. 9, pp. 90379-90391, 2021, DOI: 10.1109/ACCESS.2021.3091327.

[4]. M. I. Sarwar et al., "Data Vaults for Blockchain-Empowered Accounting Information Systems," in IEEE Access, vol. 9, pp. 117306-117324, 2021, DOI: 10.1109/ACCESS.2021.3107484.

[5]. R. Awadallah and A. Samsudin, "Using Blockchain in Cloud Computing to Enhance Relational Database Security," in IEEE Access, vol. 9, pp. 137353-137366, 2021, DOI: 10.1109/ACCESS.2021.3117733.

[6]. J. Lian, S. Wang, and Y. Xie, "TDRB: An Efficient Tamper-Proof Detection Middleware for Relational Database Based on Blockchain Technology," in IEEE Access, vol. 9, pp. 66707-66722, 2021, DOI: 10.1109/ACCESS.2021.3076235.