A Genetic Approach for Reversible Database Watermarking using Fingerprint masking

J.Menaka Gandhi, T.J Shredha, S.Vishali, V.K. Vishnupriya

[1]*Assistant Professor, Dept. of Information Technology, Meenakshi College of Engineering, Tamil Nadu, India*
[2,3,4]*Student, Dept. of Information Technology, Meenakshi College of Engineering, Tamil Nadu, India*

---***---

Abstract - *Data mining involves the practice of examining large pre-existing raw databases to generate usable new information. Authenticity, privacy, and confidentiality of larger datasets are outsourced to the cloud and internet in the encrypted format. Although it has wide attention because of its large storage features, there are severe security and privacy challenges it brings. Watermarking is used to enforce proprietary rights on shared relational databases. There are various reversible watermarking methods used to protect the rights of owners and to recover original data. Generally, the methods introduced, compromise the original data to a large extent that cannot achieve a good balance between robustness, data quality degradation, malicious attacks, and data recovery. In this paper, we proposed a vigorous and reversible database watermarking technique, Genetic Fingerprinting Algorithm using Histogram shifting (GFAHS). The original database is encoded along with the fingerprint of the owner and watermarked together. The outcomes are in terms of robustness and prevention from illegal ownership violations.*

Key Words: Genetic Fingerprinting algorithm, Histogram shifting, Robustness, Watermarking, Ownership violation.

1. INTRODUCTION

Data mining examines large databases to generate usable new information. It can be used in a variety of ways, such as database marketing, credit risk management, fraud detection, spam Email filtering, or even to discern the sentiment or opinion of users. Data Mining techniques are used to predict the hidden patterns and behaviors. This allows the businesses to analyze data from different dimensions, then categorizing and summarizing it into meaningful information. Data is an important asset to any organization and thereby, it is essential to safeguard it from online frauds [1]. Data security is not just important for organizations, but also for the remote devices that connect to an organization gets attacked by attackers to tap sensitive information. Data breaches and cyber-attacks are anticipated to increase as the computer networks expand.[1] While there are technical methods for protecting the content of the websites, there are tech-savvy thieves who can probably still find their way around these protections. Here in our proposal, we have used the Watermarking technique where duplicate data is embedded with the original data to protect the original content and make it harder to copy the item.

Reversible watermarking techniques are also called as invertible that are applied where the authenticity of an image has to be granted and the original content is to be decoded. According to Risk Based Security research published in the year 2019, the first six months had seen more than 3,800 publicly disclosed breaches that exposed an incredible 4.1 billion compromised records. The most remarkable fact is that 3.2 billion of those records were exposed by just eight breaches. As for the exposed data itself, the report has an email that existed in 70% of breaches and passwords of 65% at the top of the pile.

2. PRELIMINARIES AND PREVIOUS WORK

2.1 Firefly optimization algorithm

This paper illustrates the reversible database watermarking that uses a firefly algorithm which is called as an optimization algorithm. This approach uses DEW (Difference Expansion watermarking) algorithm with FFA (Firefly Algorithm) called FFADEW [2]. This technique is used to embed watermark with relational databases and also reverse the watermarked database to original data after watermark extraction and verification. The drawbacks of this system includes that it does not use the notion of persistent watermarking which improves the integrity and does not discuss how to improve the existing techniques in terms of persistency.

2.2 A Novel watermarking

A reversible watermarking technique recovers the primary data from watermarked data and secure data quality. This technique used to recover large data even after the malicious attacks and security threats[3]. RRW analysis is done where the watermarking is spotted with an increased number of decoding accuracies in various modules. It provides ownership protection over a relational database. The drawbacks of this system include, this violates query processing in watermarking and also the distortion rate is relatively high.

2.3 Prediction errors by efficient histogram modification

A Reversible image watermarking algorithm advocates the efficiency of modifying a pair of histogram bins. Multiple sets of histograms are selected in sequence for data embedding and pre-process of pixel values is carried out to avoid the overflow and underflow [4]. Blind extraction and recovery are processed by merging a precomputed location and other data into a watermarked image. This algorithm produces high payload data Hiding. The drawbacks of the system include, it has a wide-open to security threats and malicious attacks. The system fails in pattern recognition and authentication.

2.4 Persistent Watermarking

This paper explains Persistent watermarking of the relational databases that combine both the Public and Private watermark technique which allows the owner to prove his ownership rights and allows the end-user to demonstrate the correctness and originality of the Information without the loss of data. The public watermarking is based on a part of database state which is invariant in the processing of queries and the private watermarking technique is based on the original database state called abstract database which remains invariant in the processing of the associated queries [3]. The system refuses the data coloring method of cloud watermarking to acknowledge and ensure the mutual reputation.

3. PROPOSED SYSTEM

The proposed system presents the method to secure the information using the GFAHS (Genetic Fingerprinting Algorithm using Histogram shifting). We consider a typical multimodal system, made up of a Biometric authentication that operates by providing physical access and logical access to the internal information. This system also includes watermarking methods. In the watermarking concept, the duplicate data is overwritten above the original context; here the entire information in the database is encrypted and watermarked where no one could view the original context. It also includes double encryption where the fingerprint authentication is also encrypted within the database. Group sharing in the cloud is a hot topic in recent years. Our proposed work focused on providing a secure and robust feature in the group sharing in an organization using watermarking technology [1]. The watermarking is followed by using a genetic algorithm. This genetic algorithm helps in partitioning the available database randomly into tuples and distributes these tuples unsystematically to the employees present in a particular team. These employees update the database on the completion of their tasks. The update is carried periodically where each employee is aware of only their particular tuples or tasks and updates their task details to their team leader individually. On completion of all the distributed tasks, the team leader finally combines the task and forwards it to the HR Manager of the organization. Here a robust biometric authentication is used by the manager to view the task details.

3.1 ARCHITECTURE

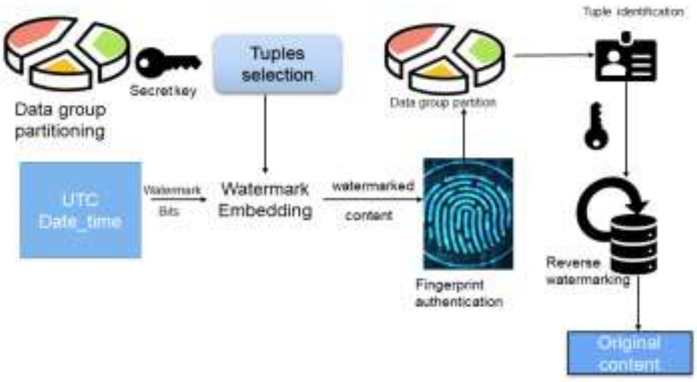


Chart -1: FINGERPRINT AUTHENTICATION

In the data group partitioning, the data gets partitioned with the help of a secret key. In the UTC Date time the alphanumerical strings are converted into the form of 0’s and 1’s (Chart-1) and then encode the fingerprint authentication with the watermark embedding. The data retrieval process is carried out with the following process which encompasses tuple identification followed by reverse watermarking[1] to retrieve the original content.

3.2 MODULE DESCRIPTION

The modules included in our proposed work are:

Database Partitioning
Daily Update
Captcha (Private Key) generation
Information Sharing and Accessing

3.2.1: Database Partitioning

Genetic Algorithm is used to partition the database randomly and distribute it to the team members [8]. This is the initial part of the group sharing where available information in the database is shared evenly and randomly to the team mates[Fig -3].

Step 1:
Generate tuples i.e.) tuples () from the group of data and select a set of info () from the generated tuple.

Step 2:
To Calculate the fitness function, select the random number from generated tuple which is compared with the total length of the generated info().
if the generated value is less than the info Len() then the value gets incremented.

Step 3:
If the fitness score is less than or equal to 0, we have reached the target.

Step 4:
Else generate the new set of data

3.2.2: Daily Update

This system holds a hierarchical structure where the HR Manager, Team Leaders, and their respective Team members are registered onto the organization's website [Fig -1]. The Manager and Team Leader are given the privileges to accept and deny the team members' requests [Fig -2]. Managers now hold their respective employee's login credentials [8]. The Team Leader assigns the tasks to their Team Members regularly and on submission of those tasks, a typical document is generated as the daily report of the ongoing project and delivered to the Team Leader [9]. Here the Team Members are not allowed the privilege to rewrite or view the document once submitted to the Team Leader.

3.2.3: Captcha (Private Key) generation

Captcha is the recently secured and developing technology that is used to generate random alphanumeric values to differentiate between the automated computer program and human logins. Here different captcha is generated [Fig-4] for each document randomly [9]. And when the Team Leader wants to access those documented files, they have to provide the password and captcha to view. A QR code is generated [Fig -5] at last that allows the HR manager to view the daily updates of the Team Leader and the Employees which can be scanned to gain access.

Step 1:

To generate a captcha, initially create a random string combination Random() in the String combination=0123456789ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz

Step 2:

The length of the captcha is six which is generated from the list of the string combinations, which follows with the next combination length

Step 3:
Once the captcha is generated, it displays in the image format "GenerateCaptcha.aspx?"

3.2.4: Information Sharing and Accessing

From the information that is shared by the Team Leader and Team Members the HR Manager can now keep track of the details of the documents updated daily[Fig -3]. The Manager can also navigate to the specific document of the Team Members using the Team Member's Identity[10]. Once the final documented file is updated to the Manager no one can access or view the document. The Manager is provided with dual authentication (fingerprints and password) for access[10]. The fingerprint authentication is carried out using the SHA Algorithm which involves pixel to pixel verification of the fingerprint image [Fig -6].

4. IMPLEMENTATION

EMPLOYEE REGISTRATION

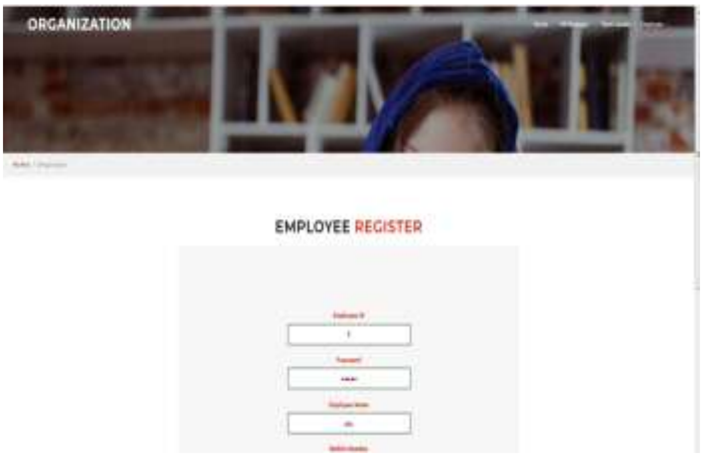


Fig -1: Employee registration

Each member of the organization has to register before communicating and sharing information among them.

TEAM LEADER LOGIN

Each member in the company has to log in before group sharing.

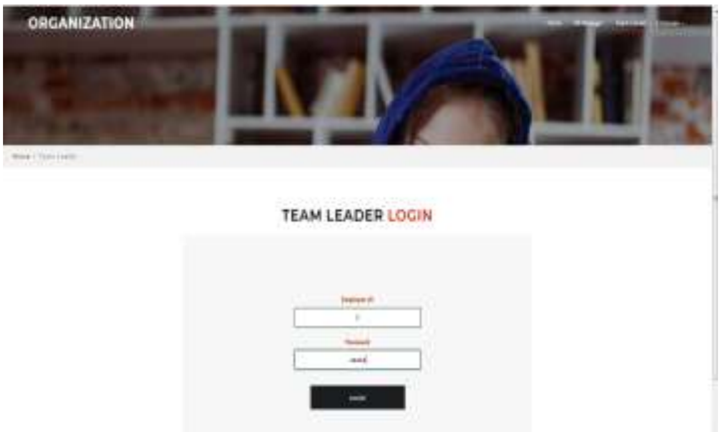
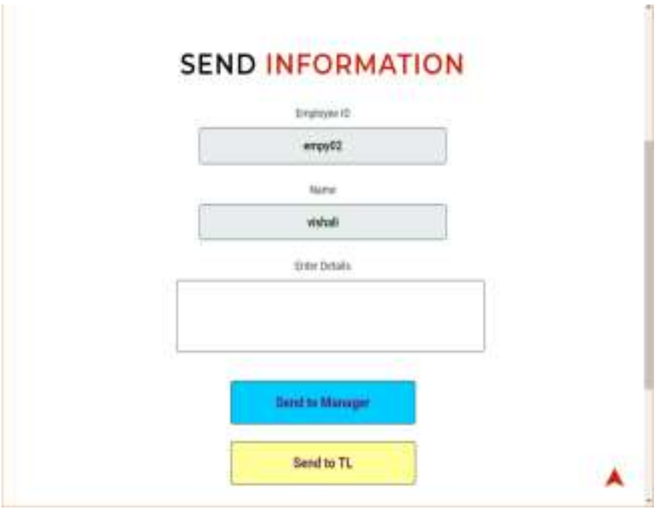


Fig -2: Team leader login

EMPLOYEE DAILY UPDATES

On this page, the employee has to include their daily updates which generate a pdf document that is sent to both the Manager and Team Leader.


SEND INFORMATION

Employee ID: empy02

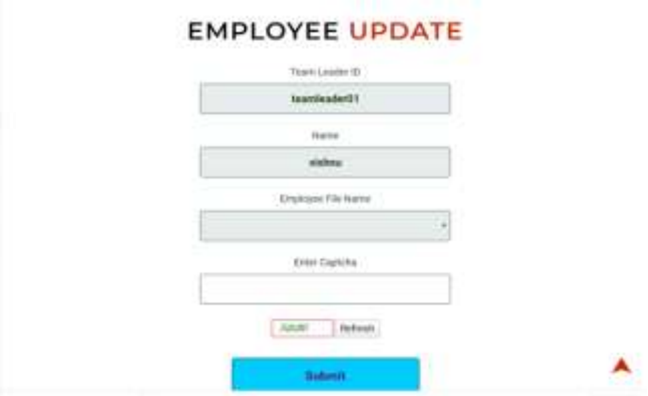
Name: vishali

Enter Details

Fig -3: send information

CAPTCHA VERIFICATION

To open and view the pdf document obtained from the employee, the Team leader has to provide the Captcha which is used to view the information shared.


EMPLOYEE UPDATE

Team Leader ID: teamleader01

Name: vishnu

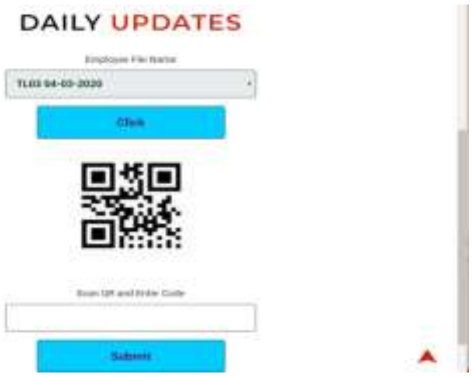
Employee File Name

Enter Captcha

Fig - 4: Employee Update

QR CODE GENERATION

A QR-Code is generated to the manager which helps him to view the document.


DAILY UPDATES

Employee File Name: TL03 04-03-2020

Scan QR and Enter Code

Fig -5: QR Generation

FINGERPRINT VERIFICATION

The final fingerprint authentication results in viewing the entire document shared by the Team leader.

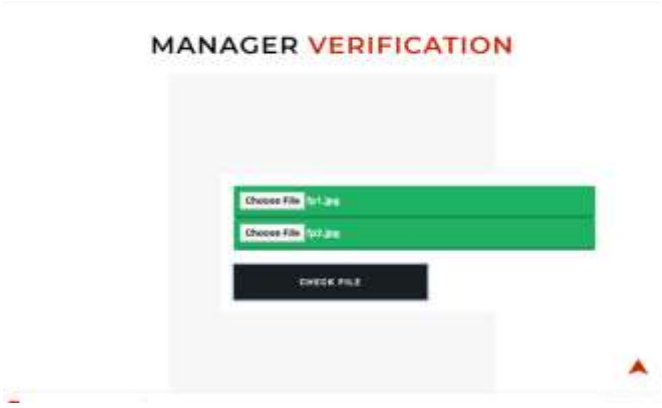

MANAGER VERIFICATION

Fig -6: Fingerprint Verification.

5. CONCLUSION

Thus, this paper encompasses around securing the information communicated among the project members and also provides a robust nature, when subjected to malicious attacks and also prevents illegal copying and malicious attacks. This prevents the organization's private information from data breeching which provides high security and protection of the database.

REFERENCES

[1] Donghui Hu, Dan Zhao and Shui Zheng "A New Robust Approach for reversible Database Watermarking with Distortion Control" Vol31, No.6, June 2019.

[2] McNickle, "Top 10 data security breaches in 2012," Apr. 17, 2013. [Online]. Available: http://www.healthcarefinancenews. com/news/top-10-data-security-breaches-2012

[3] R. Agrawal and J. Kiernan, Chapter 15—watermarking relational databases, in Proc. 28th Int. Conf. Very Large Databases, San Francisco: Morgan Kaufmann, 2002, pp. 155–166.

[4] R. Halder and A. Cortesi, "A persistent public watermarking of relational databases," in Proc. Int. Conf. Inf. Syst. Secur., 2010, pp. 216–230.

[5]D. Gross-Amblard, "Query-preserving watermarking of relational databases and XML documents," ACM Trans. Database Syst., vol.

[6] E. B. M. Shehab and A. Ghafoor, "Watermarking relational databa using optimization-based techniques," IEEE Trans. Knowl.

[7] R. Halder, S. Pal, and A. Cortesi, "Watermarking technique for relational databases: Survey, classification, and comparison, "J. Universal Comput. 2010

[8] M. Rabin, "Efficient dispersal of information for security," Journal of the ACM (JACM).
[9] J. G. et al. (2006) The expanding digital universe: A forecast of worldwide information

[10] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song,

[11] Y.Zheng, X.Yuan, X. Wang, and X. Gui, "Toward encrypted cloud media center with Secure deduplication," IEEE Trans. Multimedia, vol. 19, no. 2,pp. 251-265,2017.

[12] Y. Wu, Z. Wei, and R.H. Deng, "Attribute-based access to scalable media in the cloud- Assisted content sharing networks," IEEE Trans. Multimedia, vol. 15,no.4,pp. 788-788,2013

[13] L. Zhou, V. Varadharajan, and M. Hitchens, "Cryptographic role-based access control for secure cloud data storage systems," Information Forensics and Security IEEE Transactions on, vol. 10, no. 11, pp. 2381–2395, 2015.

[14] F. Chen, T. Xiang, Y. Yang, and S. S. M. Chow, "Secure cloud storage meets with secure network coding," in IEEE INFOCOM, 2014, pp. 673–681

[15] D. He, S. Zeadally, and L. Wu, "Certificateless public auditing scheme for cloud-assisted wireless body area networks," IEEE Systems Journal, pp. 1–10, 2015.