ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 3, Issue 10, October 2015**

# Room Reservation and Encryption Based Reversible Watermarking in Relational Databases

Elizabeth Benny , Maria Kurian

M.Tech Student, Dept. of CSE, Ilahia College of Engineering & Technology, Muvattupuzha,Ernakulam, India.

Assistant Professor, Dept. of CSE, Ilahia College of Engineering & Technology, Muvattupuzha, Ernakulam, India.

**ABSTRACT:** Data is generated excessively due to the increasing use of internet and cloud computing. Relational data is shared extensively with research communities and in virtual data storage locations in cloud. The purpose is to work in a collaborative environment and to make the data openly available. Consequently, they are vulnerable to security threats concerning ownership rights and data tampering. Reversible Watermarking is applied to relational database for ownership protection and information hiding. This paper proposes a room reservation and encryption based reversible watermarking which ensures data quality along with data recovery. Encryption schemes are used to provide high security. Experimental results prove that the method is effective against malicious attacks and show that the proposed technique outperforms the existing ones.

**KEYWORDS**: Reversible watermarking, Genetic algorithm, Room reservation, Data quality, Data recovery.

## I. INTRODUCTION

The recent surge in the growth of technologies have resulted in the generation of large amount of data[1]. There are different sources to store the data such as images, video, audio, relational databases etc. Relational databases play an important role in the storage of data. It can be mainly used by the research communities to work in a collaborative environment and to make the data openly available. This helps them for knowledge extraction and decision making. However, to make the data openly available, it should be protected in some way. With the exponential increase of internet users and its technologies, the threats that arise from un-trusted parties also increased. A major weakness of digital technology is unauthorized and illegal distribution of digital objects is easily achieved and this activity is threatening to become the worst enemy of digital era[2].

Watermarking is a form of information hiding with a goal of preserving the copyright of the digital assets such as relational databases. It has been used to prevent tampering of variety of data formats and to ensure security in terms of ownership protection. Watermarking adds a level of protection by making deliberate changes in digital objects such as relational databases, providing that they can be detected in future. Reversible watermarking techniques can ensure data recovery along with ownership protection. Fingerprinting, data hashing, serial codes are some other techniques used for ownership protection[3]. Digital watermarking of multimedia content is more commonly known mainly image watermarking. However the basic process of multimedia watermarking is different from that used to watermark relational databases. This is because of the difference in the properties of data. Multimedia data is highly correlated and continuous whereas relational data is independent and discrete. Watermarking databases ensure security but the major drawback is that they modify the data to a large extent which results in the loss of data quality. The quality of the data cannot be compromised as it is used for knowledge extraction, information interchange and decision making. In this paper, we use a room reservation based reversible watermarking technique. It will overcome the problem of data quality degradation by allowing recovery of original data along with the embedded watermark information.

This paper focus on the development of an information model which identifies the features that do not have significant effect on the decision making process. The concept of Mutual Information [4],[5] which measures the amount of information that one feature contains about the other features in a database is used here. It will select the suitable features from the database for watermarking by ranking them. The process mainly comprises a (1) data preprocessing phase, (2) Watermark encoding phase, (3) attacker channel, (4) watermark decoding phase and (5) data

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 3, Issue 10, October 2015**

recovery phase. In data preprocessing phase, several strategies and secret parameters are used to analyze and rank the features to watermark. The position to which the watermark is embedded is determined with an optimization scheme called Genetic Algorithm. Room reservation for the watermark also happens in this phase. An encryption is also applied to the watermarked data for further security. In the watermark encoding phase, the watermark information is embedded in the selected feature(s). Finally, the watermarked data is generated. The attacker channel comprises of alterations in the databases which will degrade the quality of databases. In the decoding phase, the embedded watermark is decoded from the suspicious data giving the original database.

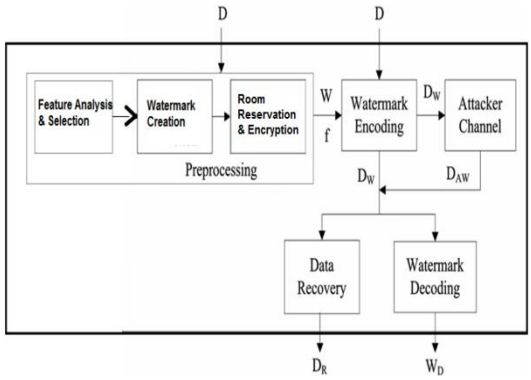


Fig.1. Architecture of the process.

## II. RELATED WORK

The need to provide security to the relational databases is an important issue as the use of internet is increasing day by day. The new techniques can produce the copies of the "original" without any quality loss in it. It can be easily distributed and widely used[6]. So, providing digital watermarking along with encryption to relational databases protect it from tampering and ensure copyright protection which maintains the integrity of relational databases.

The first reversible watermarking scheme was proposed in [7]. This schema guarantees a clear and exact tampered-or-not authentication without causing any permanent distortion to the database. A lossless and exact authentication of relational databases is achieved by reversible watermarking. It keeps track of overhead information to authenticate data quality. But this technique is not robust against heavy attacks. These attacks may target large number of tuples. Difference expansion watermarking techniques[9], exploit methods of arithmetic operations on numeric features and perform transformations. A watermarking scheme is proposed that is reversible as well as blind. Here, watermark is embedded in LSB of features to minimize distortions. This scheme provides reversibility to high quality original data set. It also provides rightful owner identification. There is no need to store original database at a secure secondary storage. Genetic algorithm and difference expansion based reversible watermarking (GADEW)[8] technique is a robust and reversible solution for relational databases. It is by minimizing distortions in the data and increases the watermark capacity. GADEW used the distortion measures AWD and TWD to control the distortions in the resultant data. In this context, the robustness of GADEW can be compromised when the values given for AWD and TWD are high. In [10], the author describes a robust, blind, resilient and reversible image based watermarking scheme for large scale databases. Here, we use the bit string of an image as the watermark. This technique demonstrates a remarkable decrease in watermark detection rate during various types of heavy attacks which results in the distortion of database tuples. It is suitable for databases of any size with reasonable performance on embedding and extraction. GA is a technique that controls data distortions and keep the data quality intact.

## III. PROPOSED ALGORITHM

The main architecture of the process is presented in Fig.1. It includes the following four major phases: (1) watermark preprocessing; (2) watermark encoding; (3) watermark decoding; and (4) data recovery. The watermark preprocessing phase computes parameters for calculation of an optimal watermark. These parameters are used for watermark encoding and decoding. The main focus of watermark encoding phase is to embed watermark such that it does not affect the data quality. After watermarking, the data is released to the recipients through a communication channel called attacker channel which is insecure. The data undergoes several attacks in the attacker channel. The watermark

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 3, Issue 10, October 2015**

decoding phase recovers watermark information for detection of embedded watermark. Data recovery phase mainly comprises the important task of successful recovery of the original data.

A. *Watermark Preprocessing Phase*

In the preprocessing phase, two important tasks are accomplished: (1) selection of a suitable feature for watermark embedding; (2) calculation of an optimal watermark with the help of an optimization technique; (3) Room reservation and Encryption of the database.

(1) Feature Analysis and Selection

All the features of the dataset are ranked according to their importance their information extraction. They use the concept of mutual information to determine the best position to embed the watermark. Mutual information of every feature with all other features is calculated by using the equation (1).

$$MI(A,B) = \sum_{a}\sum_{b} P_{AB}(a,b)\log\frac{P_{AB}(a,b)}{P_A(a)P_B(b)}. \qquad (1)$$

Where MI(A,B) measures the degree of correlation of features by measuring the marginal probability distributions and the joint probability distribution.

(2) Watermark Creation through Genetic Algorithm(GA)

Genetic Algorithm evolves a potential solution to an optimization problem. During watermark creation, it goes through a number of processing states like initialization, selection, mutation and crossover. Undergoing these processes yield the best position to embed the watermark.

(3) Room reservation and Encryption of Database

We will reserve the space for storing the watermark so that they can be decoded without the loss of data quality. The data is converted into number format and we add a zero at the right end of the number as reserving the space for embedding the watermark. The watermark is then converted to ASCII code and is added with the number to get the encrypted format. The encryption used is homomorphic encryption. Homomorphic encryption is a form of encryption which allows specific types of computations to be carried out on ciphertext and obtain an encrypted result which decrypted matches the result of operations performed on the plaintext. For instance, one person could add two encrypted numbers and then another person could decrypt the result, without either of them being able to find the value of the individual numbers.

**Definition**: Let the message space (M, o) be a finite (semi-) group, and let σ be the security parameter. A homomorphic public-key encryption on M is a quadruple (K, E, D, A) of probabilistic, expected polynomial time algorithms, satisfying the following functionalities:

- **Key Generation:** On input 1σ the algorithm K outputs an encryption/decryption key pair ( ke,kd) = k ϵ K where K denotes the key space.
- **Encryption**: On inputs 1σ, ke , and an element m ϵ M the encryption algorithm E outputs a ciphertext c ϵ C where C denotes the ciphertext space.
- **Decryption**:The decryption algorithm D is deterministic. On inputs 1σ, k, and an element c ϵ C it outputs an element in the message space M so that for all m ϵ M it holds if c = E(1σ,ke,m) then Prob[D(1σ,k,c) ≠ m] is negligible, i.e., it holds that Prob[D(1σ,k,c) ≠ m] ≤ 2-σ.
- **Homomorphic Property**: A is an algorithm that on inputs 1σ,ke, and elements c1,c2 ϵ C outputs an element c3 ϵ C so that for all m1,m2 ϵ M it holds: if m3 = m1 o m2 and c1= E(1σ,ke,m1) and c2 = E(1σ,ke,m2) then Prob[D(A(1σ,ke,c1,c2))] ≠ m3] is negligible.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 3, Issue 10, October 2015**

A homomorphic cryptosystem is a cryptosystem with the additional property that there exists an efficient algorithm to compute an encryption of the sum or the product, of two messages given the public key and the encryptions of the messages but not the messages themselves.

Another encryption scheme used is paillier cryptosystems. It consists of key generation and encryption. It uses public key encryption for providing security. Key generation is based on two randomly generated prime numbers. It provides high security such that if any of the tuple in a relational database is tampered, it is detected easily. So, it provides high detection rate. But if the size of the relational database is high, it takes time for encrypting and decrypting the entire database.

B. *Watermark Encoding Phase:*

Genetic algorithm is used to create the watermark. It includes two values (1) Watermark string of length l and (2) β value. B is a parameter which represents the tolerable amount of change to embed in the feature values. The β value is saved for watermark encoding and decoding. These two values, β and watermark length l is used to manipulate data if it satisfies the usability constraint λ. When the given bit is 0, the value of β is added into every tuple of the selected feature A. Otherwise, its value is subtracted from the value of feature.

**Algorithm 1.** Watermark Encoding

```
Input: D, w, β
Output: D_W, ∇
  for w = 1 to l do
    //loop will iterate for all watermark bits w from 1 to length
    l of the watermark
    for r = 1 to R do
      //loop will iterate for all tuples of the data
      if b_{r,w} == 0 then
        // the case when the watermark bit is 0
        changes are calculated by using Equation (6)
        data is watermarked by using Equation (8)
        insert η_r into ∇
      end if
      if b_{r,w} == 1 then
        // the case when the watermark bit is 1
        changes are calculated by using Equation (6)
        data is watermarked by using Equation (7)
        insert η_r into ∇
      end if
    end for
  end for
  return D_W, ∇
```

Fig. 2. Watermark Encoding Algorithm

**Algorithm 2.** Watermark Decoding

```
Input: D_W or D'_W, ∇, l
Output: W_D
  for r = 1 to R do
    //loop will iterate for all tuples of the data
    for b = l to 1 do
      //loop will iterate for all watermark bits b from 1 to
      length l of the watermark
      η_{d_r} ⇐ D'_{W(r)} * ζ
      η_{Δ_r} ⇐ η_{d_r} − η_r
      if η_{Δ_r} ≤ 0 then
        detected watermark bit (dtW) is 1
      else if η_{Δ_r} > 0 and η_{Δ_r} ≤ 1 then
        detected watermark bit (dtW) is 0
      end if
    end for
  end for
  W_D ⇐ mode(dtW(1, 2, ..., l))
  return W_D
```

Fig. 3. Watermark Decoding Algorithm

C. *Watermark Decoding Phase*

In the watermark decoding phase, the first process is to locate the features which have been marked. A watermark decoder is used which calculates the amount of change in the value of a feature such that it does not affect the data quality. The watermark decoder decodes the watermark by working with one bit at a time. The optimization process through GA is not required in this phase.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 3, Issue 10, October 2015**

D. *Watermark Recovery Phase*

After detecting the watermark string, some post processing steps are carried out for error correction and data recovery. The optimized value of β computed through the GA is used for regeneration of original data. The data recovery algorithm is presented in Algorithm 3.

## IV. RESULTS AND DISCUSSIONS

When watermarking of databases is done, they ensure ownership. But watermarking alone will not guarantee security and preserve data quality. The detection rate is very low. It will identify the tampering only if a number of changes are made in the database. When room reservation and encryption are done the detection rate of tampering is very high. Even a small change in the database can be easily detected with this. This ensures high security to the database. The graph in the fig. 4 shows the detection rate of watermarking with encryption as cent percent which ensures high security whereas watermarking alone does not detect the modifications in the database to that extend.

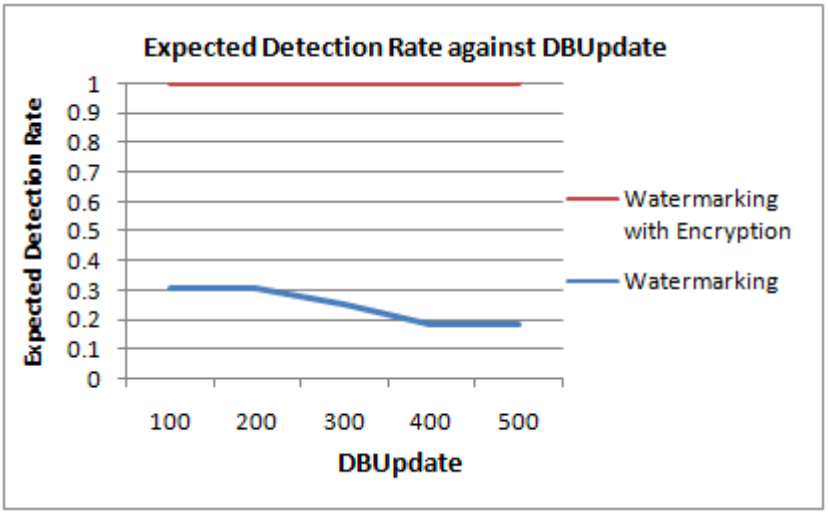


Fig. 4. Graph showing the expected detection rate against DBupdate

## V. CONCLUSION

A reversible watermarking technique is used in relational data for providing security to the data by reserving room for embedding the watermark. By using the concept of room allocation and encryption, it is possible for complete data recovery and higher security. The data quality will also remain intact.The watermark is detected with maximum decoding accuracy in different scenarios.

## REFERENCES

1. Y.-C. Liu, Y.-T. Ma, H.-S. Zhang, D.-Y. Li, and G.-S. Chen, "A method for trust management in cloud computing: Data coloring by cloud watermarking," Int. J. Autom. Comput., vol. 8, no. 3 pp. 280–285, 2011.
2. CBS News: "Digital piracy stronger than ever", Online link: http://goo.gl/Ws2rZ, valid as of October 2010.
3. S. Subramanya and B. K. Yi, "Digital rights management," IEEE Potentials, vol. 25, no. 2, pp. 31–34, Mar.-Apr. 2006.
4. T. M. Cover, J. A. Thomas, and J. Kieffer, "Elements of information theory," SIAM Rev., vol. 36, no. 3, pp. 509–510, 1994.
5. T. M. Cover and J. A. Thomas, Elements of Information Theory. New York, NY, USA: Wiley-Interscience, 2012.
6. R. Agrawal and J. Kiernan." Watermarking relational databases". In Proceedings of The 28th International Conference on Very Large Databases VLDB, 2002.
7. Y. Zhang, B. Yang, and X.-M. Niu, "Reversible watermarking for relational database authentication," J. Comput., vol. 17, no. 2, pp. 59–66, 2006.
8. K. Jawad and A. Khan, "Genetic algorithm and difference expansion based reversible watermarking for relational databases," J. Syst. Softw., vol. 86, no. 11, pp. 2742–2753, 2013.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 3, Issue 10, October 2015**

9. G. Gupta and J. Pieprzyk, "Reversible and blind database watermarking using difference expansion," in Proc. 1st Int. Conf. Forensic Appl. Tech. Telecommun., Inf., Multimedia Workshop, 2008, p. 24.
10. E. Sonnleitner, "A robust watermarking approach forlarge databases," in Proc. IEEE First AESS Eur. Conf. Satellite Telecommun.,2012, pp. 1–6.

## BIOGRAPHY

**Elizabeth Benny** is a final year Engineering Student pursuing Mtech(CSE) Degree from Ilahia College of Engineering and Technology, Affiliated to Mahatma Gandhi University, Kottayam. Her research interests are Data Mining and Security in Relational databases.

**Maria Kurian** is an Assistant Professor in the Computer Science Engineering Department, Ilahia College of Engineering and Technology, Affiliated to Mahatma Gandhi University, Kottayam. She received Master of Technology(M.Tech.) degree from Adi Shankara College of Engineering and Technology.