A Robust and Reversible Watermarking Technique for Healthcare System

Shivkumar Mishra, R. R. Sedmakar

Abstract*: *Advancement in Information and technology results in use informational system on a large scale. This informational system consisting of the relational databases stored over the network and shared by different owners in a collaborative environment for decision-making and knowledge extraction purpose. Sharing these database results in the security threats such as data tampering and ownership rights. Medical data are also stored and shared by the different health organization over the cloud network. This medical data are also targeted by the attacker to manipulate or misuse the original data for their benefits. Thus, this data which are stored and shared over the cloud network needs to be taken care by implementing some powerful security mechanism. Watermarking technique are used as powerful security mechanism for ownership protection and data tampering from the malicious attackers. However, using the watermarking technique results in the modification of the original data that degrade the data quality and results in data distortation. Thus, reversible watermarking techniques are used for data protection along with data recovery. There are various reversible watermarking techniques available such as Reversible Watermarking Technique (RRW), Difference Expansion Watermarking Technique (DEW), Genetic Algorithm based on Difference Expansion (GADEW), Prediction Error Expansion Watermarking technique (PEW), and Quantization Index Modulation (QIM). The proposed system is the combination of the RRW technique and QIM technique along with distortion control for selecting the appropriate features for watermarking and data recovery. The proposed method will be more secured and robust against the attack with less data distortion and higher data recovery, implemented for relational datasets of medical healthcare system.

Index Terms*: *Data Recovery, Data Quality, Numerical Data, Reversible Watermarking, Security.

I. INTRODUCTION

In today's world of digital communication, large amount data is being used and generated due to the increasing use of the Internet and Cloud Computing. These data are stored in different digital formats such as images, audio, video, texts and relational data. Relational data are shared extensively between the different owners and with research communities and in virtual data storage locations over the Cloud in order to work in a collaborative environment and make data openly available for knowledge extraction and decision making. For example, Walmart- a large multinational retail corporation that has made its sales database available openly over the Internet so that it may be used for the purposes of identifying market trends through data mining techniques[3]. However, these openly available datasets becomes attractive targets for attackers to attack and tamper the datasets.

Revised Version Manuscript Received on 28 October 2018.

Shivkumar Mishra, M.E Student, Department of Computer Engineering, Thakur College of Engineering & Technology, Mumbai (Maharashtra), India.

Dr. R. R. Sedamkar, Professor & Dean, Academics, Thakur College of Engineering & Technology, Mumbai (Maharashtra), India.

For example, there are documented attack incidents which contain personal information related to customers using certain Wal-Mart video services was stolen[4]. According to a survey related to the security of outsourced customer data[4], it is reported that 46% of organizations do not consider security and privacy issues while sharing their confidential data. Therefore, 64% organizations have to face data loss repeatedly. Data breaches in the health care and medical domain also increasing alarmingly[5]. Therefore, it is important for the owner, that in shared environments such as Cloud, security threats that arise from untrusted parties and attacks for relational databases need to be taken care along with the enforcement of ownership rights on behalf of their owners.

Watermarking techniques are used for security purpose to protect ownership rights and data tampering for a wide variety of data formats such as images, audio, video and relational databases. But using watermarking technique also results certain modifications in the original data; as a result of which, the data quality gets degraded. Thus, reversible watermarking is used to overcome the problem of data quality degradation by allowing recovery of original data along with the embedded watermark information. Here we propose a reversible watermarking technique that tries to overcome the problem of data quality degradation by allowing recovery of original data along with the embedded watermark information that keeps the data useful for knowledge discovery. Data modifications are allowed to such extent that the quality of the data before embedding watermark information and after extracting is acceptable for knowledge extraction process. As well as, knowledge discovery becomes successful in decision support systems where high quality data recovery is possible.

Thus, for implementing the above method following two techniques are used:

(1)Genetic algorithm based reversible watermarking technique

(2) Quantization index watermarking (QIM) technique

This two technique are used embedding the data into appropriate watermark, which keeps the data useful for knowledge discovery along with Distortion control mechanism is used to recover the original data from watermark data

II. RELATED WORK

Several watermarking technique are available for security of ownership data and tampering of the data from the malicious attacks. Some of the techniques, which are used in proposed system, are mentioned here:

2.1. Reversible Watermarking Technique:

The first reversible watermarking technique proposed by the Zhang et al. [6]. Based on histogram expansion method. In this method of distribution of error between two evenly distributed variables and some selected initial nonzero digits of errors to form histograms, and this selected nonzero initial digits of errors is used for reversible watermarking. This technique is keeps track of overhead information to authenticate data quality. However, fails in case of heavy attacks.

Difference Expansion watermarking technique (DEW) uses the method of arithmetic operations along with transformation on the numeric features. The watermark information are embedded in LSB to minimize the distortion [7]. The Genetic Algorithm based on Difference Expansion Watermarking (GADEW) technique is robust and reversible method for relational database. It overcome the drawbacks of DEW by increasing the watermarking capacity and lowering the false positive ratio [8]. Prediction–error Expansion watermarking techniques (PEEW) uses predictor to in order to oppose the difference operator to select candidate pixel for watermark embedding. This technique is proposed by the Farfoura et al. And it is fragile against malicious attacks as it uses the fractional parts of the numeric feature [9]. Proposed method uses RRW (Robust reversible watermarking) technique which uses Genetic Algorithm in order to embed optimum value for the selected features of the datasets by preserving the data quality and minimizing the data distortion along with data recovery.

2.2. Quantization Index Modulation Water Marking Technique:

Quantization Index Modulation (QIM) is a robust technique used in the relational database. In this technique modulation is used to modulate the relative angle of the centre of mass of the circular histogram which is associated with the group of values on numerical attribute of the relational data [10]

III. TECHNIQUE USED

For Implementing the Proposed system the following two techniques are used:

(1) RRW- Robust Reversible watermarking technique.

(2) Quantization Index Modulation watermarking technique.

3.1. RRW- Robust Reversible Watermarking Technique based on the Genetic Algorithm:

RRW watermarking technique is used improve the data recovery along with robust security mechanism for the relational database. It includes the following phases: (1) watermark preprocessing (2) watermark encoding; (3) watermark decoding and (4) data recovery

1. Watermarking Preprocessing:

Watermark preprocessing phase computes the different parameter in order to generate optical watermark value. It includes (a) Suitable features selection for watermark embedding (b) Optimal watermark calculation using optimization technique.

A. Features Analysis and Selection:

All the features of the datasets are ranked according the importance in mutual extraction and mutual dependences on each other. For this mutual information (MI) is used to calculate the mutual dependences of two random variables.

$$MI(A,B) = \sum_a \sum_b PAB(a,b)\log\frac{PAB(a,b)}{PA(a)PB(b)} \quad (1.1)$$

Where MI(A,B) measures the degree of correlation of features by measuring the marginal probability distributions as PA(a), PB(b) and the joint probability distribution PAB(a,b).

Then MI of one feature with all other features is computed using the relation.

MIfi = (MIfij).

Where i, j = 1,2 ..., ft with i 6= j, and ft is the total number of features

B. Watermark Creation using Genetic Algorithm:

In order to create the optimal watermark original data is embedded using Genetic Algorithm (GA). GA is evolutionary algorithm used in population based computation by applying basic genetic operation that includes selection, crossover, mutation and replacement and evaluate the quality of each chromosomes by applying the fitness function. Thus, for Optimal watermark embedding GA is used.

Steps for the creation of the Optimal Watermark Embedding:

1. Initially random population of binary bits is selected as chromosomes.
2. Fitness of each chromosomes is calculated by using constrained optimization function.
3. Most suitable chromosomes chosen as a parent chromosome by applying tournament selection mechanism.
4. Genetic Operation such as crossover and mutation is performed on the parents chromosomes in order to get new offspring with new fitness function value along with diversity of small random change.
5. Elitism Strategy is applied to get two individual chromosomes with best fitness value.
6. Next Generation of population is created by replacing the fit chromosomes of the previous generation.
7. Step 2 to 6 is repeated until MI_O (Mutual Information of original data) and MI_W (Mutual Information of Watermarked data) reaches equal values after no repetition.
8. Optimal watermark string and best fitness value (β) is obtained at the termination of the steps.

This optimal fitness value β is used for embedding the original data which is needs to be watermarked.

2. Watermark Encoding:

The optimal value of β for each features is saved for using in encoding and decoding process. This β value is added to every tuple of selected features when bits is 0 and subtracted when bits is 1. After calculating the value of β, ηr parameter is calculated. Which represent the percentage change in watermark encoding.

$$\eta r = Dr * \zeta \quad (2.1)$$

Where ζ = Watermark encoder used for watermarking.

Dr = Recovered data for tuple r
r = tuple in the database

The Watermark encoding process starts with embedding of MSB bit (Most Significant bit) of watermark. If the MSB of watermark is 1 then new value of the Dr is denoted as DWr which is calculated by using following equation.

$$Dw_r = Dr - \beta \quad (2.2)$$

And, if the value of watermarked data is 0 then is calculated by the given the equation

$$Dwr = Dr + \beta \quad (2.3)$$

This procedure is applied on all the tuples of the datasets.

3. Watermark Decoding:

In Watermark decoding process, first we identify the features which have been watermarked. The watermark decoder ζ is used to decode the watermark on the signal bits at a time with compromising the data quality. then the value of ηdr, ηr and ηΔr is calculated using the value of tuple r. this value will be different for different tuple r.

$$\eta dr = D'w * \zeta \quad (2.4)$$

$$\eta \Delta r = \eta dr - \eta r \quad (2.5)$$

where

D'w = Watermarked database after attack
ηΔr = The difference in the value of features after encoding and decoding process
ζ = Watermark decoder

The watermark decoding consist of the following two steps:

1. For a every single feature all the tuple in D'_W the watermark bits are detected from the LSB (Least Significant bit) and move toward the MSB (Most Significant bits). This process is performed in the matrix ηr.
2. The bits as then decoded as per the percentage change in the data. If $\eta\Delta r \leq 0$, the detected watermark bit will be 1. If $\eta\Delta r > 0$ and $\eta\Delta r \leq 1$, the detected watermark bit will be 0.

4. Data Recovery:

In data recovery process some post process is performed for data correction and the data recovery. The optimized value of the β is used to recover the data which is given by the eqution;

$$Dr = D'Wr + \beta \quad (2.6)$$

$$Dr = D'Wr - \beta \quad (2.7)$$

Where Dr = Recoverd data

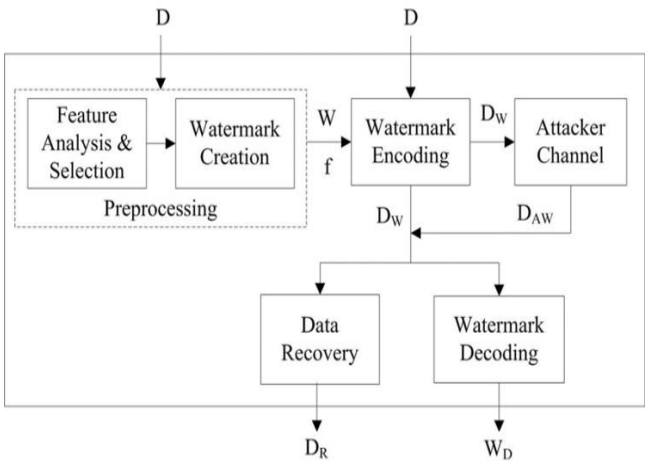

Figure 1: RRW Architecture

3.2 Quantization index Modulation based watermarking:

The quantization index modulation (QIM) embed the information by first modulating an index (m) or set of indices with the embedded information and then quantizing the host signal with the associated quantizer or set of quantizers.

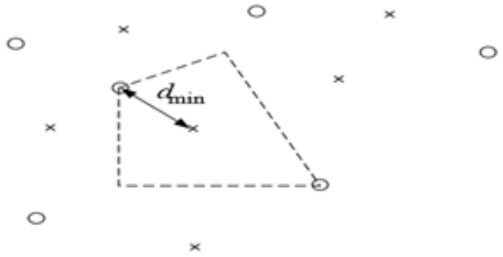

Fig 2: QIM for Message Embedding

Suppose there are two points X's and O's which belongs to two different quantizers with there associated index. For effective robustness perturbation is performed by the minimum distance dmin in given tuple of the data. For example shown in above determine the distortion, if m=1 the host signal belongs to the nearest X point, If m=2 host signal belongs to the nearest O point, Here in our system we divide the database in various domain each domain is encrypted by different key value for example if our dataset contains 800 tuple of row and column then we divide the this data into 4 groups(domain) each of 200 tuples and then each group will be encrypted by using different key.

IV. PROPOSED METHOD

The proposed methodology is combination of RRW technique and the QIM technique which will enhance the overall security of the system.

The basic functioning of the proposed system is as follows:

1. The relational data which is needed to be stored or send to another location will be selected for the watermarking
2. The features which needed to be watermarked will be selected and accordingly all the parameters such MI = Mutual Information, λ = usability constrained, β = Optimized value for watermark, etc are all calculated in feature analysis and selection face of watermarking technique
3. Watermark creation and Watermark encoding is done using genetic algorithm used in RRW method
4. Again using the same parameter which is selected in used in feature selection and analysis process Quantization Index modulation(QIM) technique is applied In this step the data which is watermarked by RRW technique is divided
5. Here for centroid is calculated for the features which are selected for the watermarking for RRW steps of features selection and analysis
6. After applying both watermarking techniques data is considered to be fully secured and stored or send to some other location for storing
7. Now, for recovery purpose first we need to decode the QIM based watermarking data by applying arithmetic operation on the centroid which is used for watermarking and after that reverse RRW technique for full data recovery

This the basic function of the proposed system which uses RRW and QIM techniques to build the application which is more secure and robust as well as the data recovery after the alteration of the data by attacker is more and data can be easily recovered.

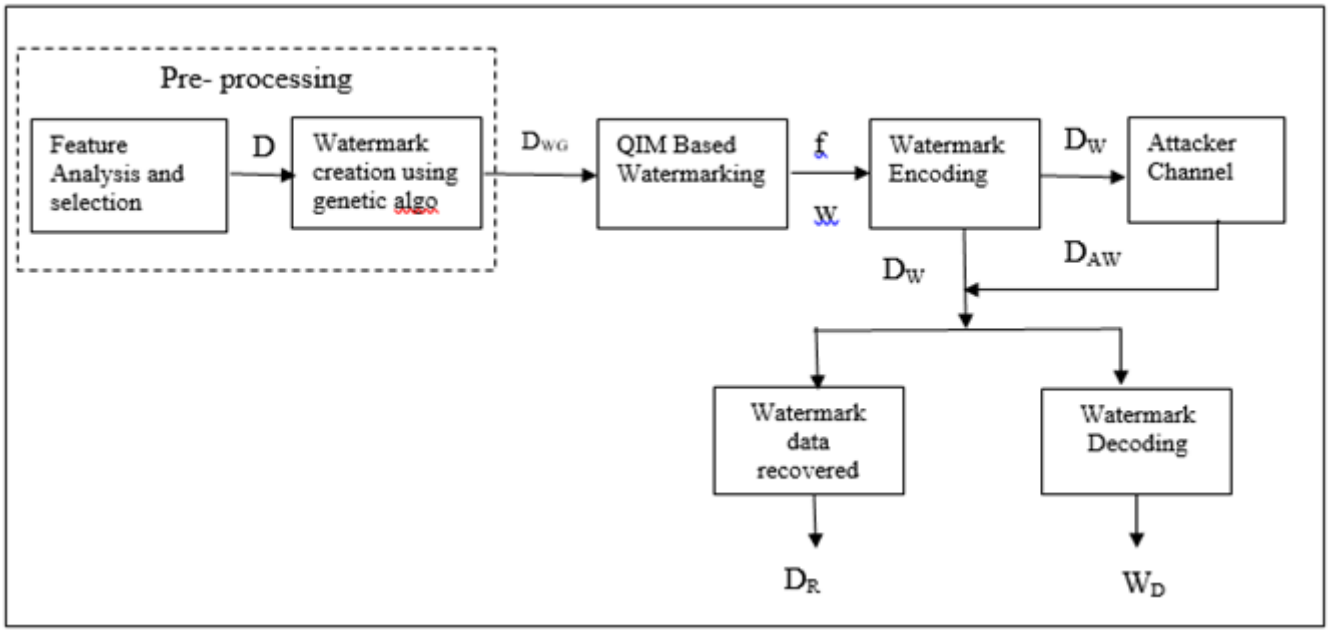

Fig 3: Architecture of Proposed System

D = Original database
D_{WG} = Watermark Creation using Genetic algorithm
D_{WGM} = Watermarking using QIM
w = Watermark bits
f = No of features in the database
D_W = Final Watermark Database
D_{AW} = Watermark data after malicious attack
Dr = Recovered watermark data
Wd= Decoded Watermark

V. RESULT

1.) Time Complexity:

The time complexity is the computational complexity that describes the amount of time it takes to run an algorithm. Time complexity is commonly estimated by counting the number of elementary operations performed by the algorithm, supposing that each elementary operation takes a fixed amount of time to perform. Thus, the amount of time taken and the number of elementary operations performed by the algorithm are taken to differ by at most a constant factor. Here the time complexity of the encrypted and decrypted algorithm is shown.

A. Encryption Algorithm:

The Encryption graph show the time taken by the algorithm to encrypt the data. The difference between the time taken to encrypt the data by advanced algorithm and original algorithm is shown in the graph

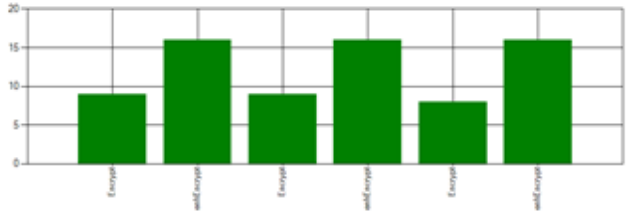

Fig. 3 Encryption Graph

B. Decryption Algorithm:

The Decryption graph show the time taken by the algorithm to decrypt the data. The difference between the time taken to decrypt the data by advanced algorithm and original algorithm is shown in the graph

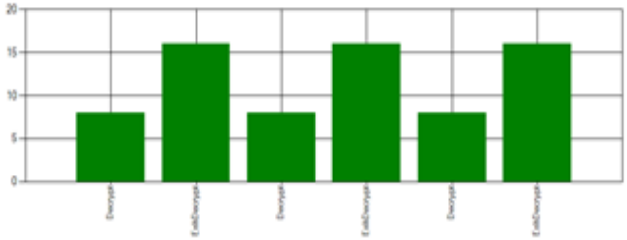

Fig 4. Decryption Graph

2.) Space Complexity:

Space complexity is a measure of the amount of working storage needed by an algorithm. That is how much memory, in the worst case, is needed at any point in the algorithm. As with time complexity, we're mostly concerned with how the space needs grow, in big-Oh terms, as the size N of the input problem grows

Here the input for previous and advanced algorithm is original data which is need to be encrypted and the space require for the algorithm execution in both case is same, therefore the space need is same for both algorithm.

VI. CONCLUSION

The proposed system is more robust and secure against the malicious attack. The data degradation is minimum and the data quality of the watermarked data is maintained after the recovery in case of attack by attacker. or vice versa.

We have also try to increase the size of the datasets with enhance security performance. Thus, we have focus on enhancement of the overall system security by minimum data distortion and high data recovery for large relational datasets of Medical HealthCare system.

ACKNOWLEDGEMENT

The proposed work on A Robust and Reversible Watermarking Technique for HealthCare System is possible because of the guidance given by my guide Professor Dr. R. R. Sedamkar without his guidance and support this work has not been possible.

REFERENCES

1. M.kamran iftikar, M. kamran, Zahid Anwar, " , IEE RRW –" A Robust and Reversible Watermarking Technique for Relational Data"",IIEE Transactions on Knowledge and Data Engineering, 201
2. Javier Franco-Contreras, Member, IEEE, and Gouenou Coatrieux, Senior Member, IEEE "Robust Watermarking of Relational Databases With OntologyGuided Distortion Control" , Information IEEE Tranction on Information forensics and security, VOL. 10, NO. 9, September 2015
3. "Walmart to start sharing its sales data," http://www.nypost.com/p/news/business/walmart-opensup, last updated: 09:55 AM on February 4, 2012, last accessed: July, 20 2013.
4. "Identity theft watch," http://www.scambook.com/blog/2013/ 04/identity-theft-watch-customer-passwords-stolen-fromwalmart-vud u-video-service/, last updated: April 11, 2013, last accessed: July, 20 2013.
5. "Securing outsourced consumer data," http://www.databreaches.net/securing-outsourced-consumerdata/, last updated: February 26, 2013, last accessed: July, 20 2013.
6. I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," Image Processing, IEEE Transactions on, vol. 6, no. 12, pp. 1673–1687, 1997.
7. G. Gupta and J. Pieprzyk, "Reversible and blind database watermarking using difference expansion," in Proceedings of the 1st international conference on Forensic applications and techniques in telecommunications, information, and multimedia and workshop. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2008, p. 24.
8. K. Jawad and A. Khan, "Genetic algorithm and difference expansion based reversible watermarking for relational databases," Journal of Systems and Software, 2013
9. M. E. Farfoura and S.-J. Horng, "A novel blind reversible method for watermarking relational databases," in Parallel and Distributed Processing with Applications (ISPA), 2010 International Symposium on. IEEE, 2010, pp. 563–569.
10. V. Pournaghshband, "A new watermarking approach for relational data," in Proc. 6th Annu. Conf. ACM-SE, 2008, pp. 127–131
11. B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," IEEE Trans. Inf. Theory, vol. 47, no. 4, pp. 1423–1443, May 2001.
12. M. Kamran and M. Farooq, "A formal usability constraints model for watermarking of outsourced datasets," IEEE Trans. Inf. Forensics Security, vol. 8, no. 6, pp. 1061–1072, Jun. 2013.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication