ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Semi-Blind Reversible Watermarking Technique for Relational Data

V.Harivardhan Reddy
M.Tech(SE) Student
Department of IT
Geethanjali College of Engineering & Technology.

K.Srinivas
Professor
Department of CSE
Geethanjali College of Engineering & Technology.

Y.Raju
Assistant Professor
Department of CSE
Geethanjali College of Engineering & Technology.

ABSTRACT

Advancement in information technology is playing an increasing role in the use of information systems comprising relational databases. These databases are used effectively in collaborative environments for information extraction; consequently, they are vulnerable to security threats concerning ownership rights and data tampering. Watermarking is advocated to enforce ownership rights over shared relational data and for providing a means for tackling data tampering. When ownership rights are enforced using watermarking, the underlying data undergoes certain modifications; as a result of which, the data quality gets compromised. Reversible watermarking is employed to ensure data quality along-with data recovery.

However, such techniques are usually not robust against malicious attacks and do not provide any mechanism to selectively watermark a particular attribute by taking into account its role in knowledge discovery. Therefore, reversible watermarking is required that ensures; (i) watermark encoding and decoding by accounting for the role of all the features in knowledge discovery; and, (ii) original data recovery in the presence of active malicious attacks. In this paper, a robust and semi-blind reversible watermarking (RRW) technique for numerical relational data has been proposed that addresses the above objectives. Experimental studies prove the effectiveness of RRW against malicious attacks and show that the proposed technique outperforms existing ones.

INTRODUCTION

In today's world data is generated due to the high use of cloud computing and internet. Usually data stored in different format like video, audio, images. Watermarking strategies have generally been utilized to ensure security regarding possession safeguard and sealing for information designs and also watermarking technology used to protect data from hackers. The watermarking program introduces small error into the object being watermarked. The watermarking programming brings little blunders into the item being watermarked. These deliberate mistakes are assembled marks and every one of the imprints constitutes the watermark. The imprints must not significantly affect the helpfulness of the information and they ought to be put in a manner that a malevolent client can't pulverize them without making the information less valuable.

A type of computerized watermarking procedures are getting utilized over the way of the last couple of numerous years for proprietorship shield of advanced media reminiscent of photographs, sound, video , and normal dialect handling application. With the expanding notoriety of irreversible and causes the coming about watermarked information to end up not quite the same as the unique substance. Subsequently, information investigation and choice making on the contorted rendition of information gets to be incomprehensible.. Digital watermarking manner has been successively applied to preserve the multimedia works and software merchandise. In a similar way, database watermarking has been proposed on massive database safetymanipulate. Reversible watermarking of social databases is a generally new and developing region. It is a decent competitor answer for guaranteeing information proprietorship assurance and its uprightness. Social information have a specific configuration and are not quite the same as that of other advanced information, for example, sound, video, programming, and pictures. The strategies for possession insurance and information recuperation of social information needs to consider new limitations taking into account actualities, for example, (i) a database comprises of tuple/records, which is the place the watermark should be installed; (ii) the requesting of the records in a database connection; and (iii) information operations such as insertion, cancellation, and adjustment that regularly happen in a database. Besides, contingent upon the assortment of data spared inside of the database.

Reversible water marking technique it recovers data with ownership protection. Some technique used for ownership protecting such as hash function, fingerprinting. Interactive media watermarking is extremely not quite the same as that used to watermark social databases due to a major distinction in the properties of the information. Interactive media information is exceptionally associated and ceaseless while social information is free and discrete. With the appearance of current copyright insurance and data concealing systems, database watermarking can be utilized to implement possession privileges of social information.. These all technique modify the data very large but they loss actual data quality. To overcome this problem reversible watermarking technique is used.

EXISTING SYSTEM

The first reversible watermarking scheme for relational databases was proposed by Zhang and Yang. In this technique, histogram expansion is used for reversible watermarking of relational database.

Histogram expansion technique is used to reversibly watermark the selected nonzero initial digits of errors. This technique is keeps track of overhead information to authenticate data quality. However, this technique is not robust against heavy attacks.

Difference expansion watermarking techniques (DEW), exploit methods of arithmetic operations on numeric features and perform transformations. The watermark information is normally embedded in the LSB of features of relational databases to minimize distortions.

Gupta and Pieprzyks' proposed reversible watermarking technique introduces distortions as a result of the embedding process. Changes in the data are controlled by placing certain bounds on LSB. On the contrary, to limit the distortions, the data outside the limited bounds is left unwatermarked. As a result, the watermark robustness gets compromised.

Sonnleitner proposed a robust, blind, resilient and reversible, image based watermarking scheme for large scale databases. The bit string of an image is used as a watermark where one bit from the bit string is embedded in all tuples of a single partition and the same process is repeated for the rest of the partitions. This technique demonstrates a remarkable decrease in watermark detection rate during various types of heavy attacks, and the database tuples get highly distorted.

PROPOSED SYSTEM

This paper proposed a robust and semi-blind reversible watermarking (RRW) technique for numerical relational data.

RRW mainly comprises a data preprocessing phase, watermark encoding phase, attacker channel, watermark decoding phase and data recovery phase.

In data preprocessing phase, secret parameters are defined and strategies are used to analyze and rank features to watermark. An optimum watermark string is created in this phase by employing GA—an optimization scheme—that ensures reversibility without data quality loss.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

In the watermark encoding phase, the watermark information is embedded in the selected feature(s). Two parameters, b the optimized value from the GA and hr a change matrix are used in the watermark encoding and decoding phases.

Finally, the watermarked data for intended recipients is generated. The attacker channel comprises subset alteration, subset deletion and subset insertion attacks generated by the adversary. These malicious attacks modify the original data and try to degrade its quality.

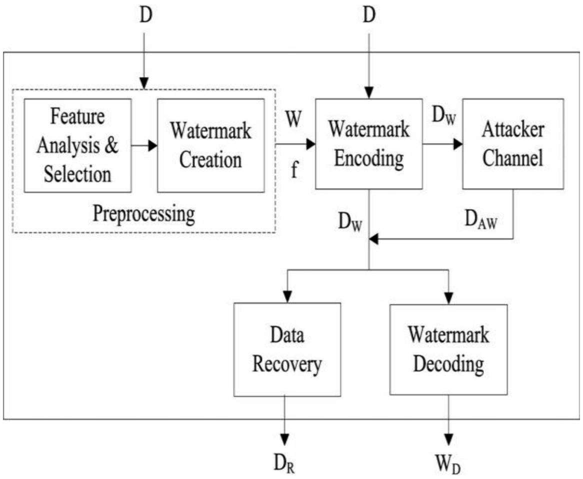
In the watermark decoding phase the embedded watermark is decoded from the suspicious data. In order to achieve this preprocessing step is performed again, and decoding strategies (feature selection on the basis of MI, b the optimized value from the GA and hr the change matrix) are used to recover the watermark.

Semi-blind nature of RRW is used mainly for data reversibility in case of heavy attacks (attacks that may target large number of tuples). Original data is recovered in data recovery phase, through post processing steps for error correction and recovery.

Advantage

1. RRW is robust
2. The data quality remains intact after watermarking.

SYSTEM ARCHITECTURE:



IMPLEMENTATION

Modules

- Watermark Preprocessing
- Watermark Encoding
- Watermark decoding
- Data Recovery

Module Description:

Watermark Preprocessing

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.

Feature Analysis and Selection

Mutual information is calculated for every feature and the lowest feature is selected for watermarking.
Mutual information of every feature with all other features is calculated by using

$$MI(A,B) = \sum_a \sum_b P_{AB}(a,b) log \frac{P_{AB}(a,b)}{P_A(a)P_B(b)}$$

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).

Watermark Creation through Genetic Algorithm (GA)

In this step the genetic algorithm is used to create the watermark bits. The optimal fitness value obtained through GA is basically the change—β—to be embedded in the original data that needs to be watermarked. The purpose of getting an optimum value of b is to justify the amount of change that a feature value can withhold without compromising the data quality.

Watermark Encoding

In this module the optimized value of β is embedded in the particular selected features.
A GA is used to create optimal watermark information that includes: (1) Optimal chromosomal string (watermark string of length *l*); and (2) β value.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

β is a parameter that is computed using GA and represents a tolerable amount of change to embed in the feature values. Once the optimum value of b for each candidate feature A is found, it is saved for use during watermark encoding and decoding. A watermark (bit string) of length l and an optimum value β is used to manipulate the data provided it satisfies the usability constraints λ. The value β is added into every tuple of the selected feature A when a given bit is 0; otherwise, its value is subtracted from the value of the feature. It is ensured that the mutual information of a feature remains unchanged, when the watermark is inserted into the database. The watermark is inserted into every tuple for the selected feature of the dataset.

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, ∇

Watermark decoding

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

The decoding phase mainly consists of two steps:
Step 1. For every candidate feature A of all the tuples in D'W, the watermark bits are detected starting from the least significant bit and moving towards the most significant bit. The bits are detected in the reverse order compared with the bits encoding order because it is easy to detect the effect of the last encoded bit of the watermark. This process is carried out using the change matrix ηr.

Step 2. The bits are then decoded according to the percentage change values of watermarked data. If ηΔr ≤ 0, the detected watermark bit will be 1. If ηΔr > 0 and ηΔr ≤ 1, the detected watermark bit will be 0.

The final watermark information is retrieved through a majority voting scheme using
$W_D \Leftarrow mode(dtW(1,2,\ldots,l))$.

Algorithm 2. Watermark Decoding

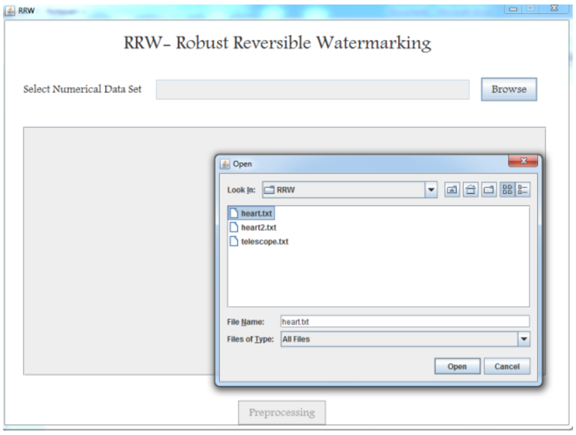
Input: D_W or D'_W, ∇, l
Output: W_D
  for $r = 1$ to R **do**
    //loop will iterate for all tuples of the data
    for $b = l$ to 1 **do**
      //loop will iterate for all watermark bits b from 1 to length l of the watermark
      $\eta_{d_r} \Leftarrow D'_{W(r)} * \zeta$
      $\eta_{\Delta_r} \Leftarrow \eta_{d_r} - \eta_r$
      if $\eta_{\Delta_r} \leq 0$ **then**
        detected watermark bit (dtW) is 1
      else if $\eta_{\Delta_r} > 0$ and $\eta_{\Delta_r} \leq 1$ **then**
        detected watermark bit (dtW) is 0
      end if
    end for
  end for
  $W_D \Leftarrow mode(dtW(1,2,\ldots,l))$
  return W_D

Data Recovery

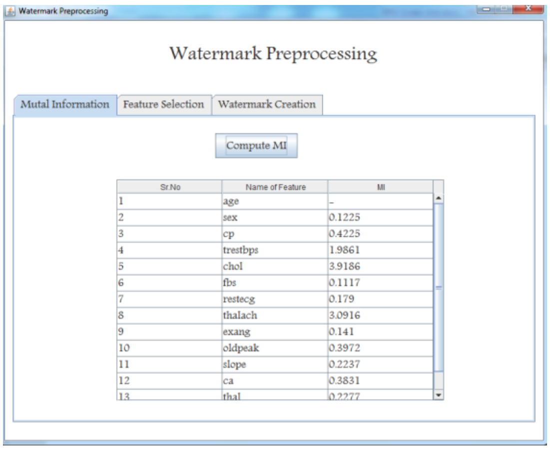
After detecting the watermark string, some post processing steps are carried out for error correction and data recovery. The optimized value of b computed through the GA is used for regeneration of original data. The value of a numeric feature is recovered using

$$D_r = D'_{Wr} + \beta$$
$$D_r = D'_{Wr} - \beta$$

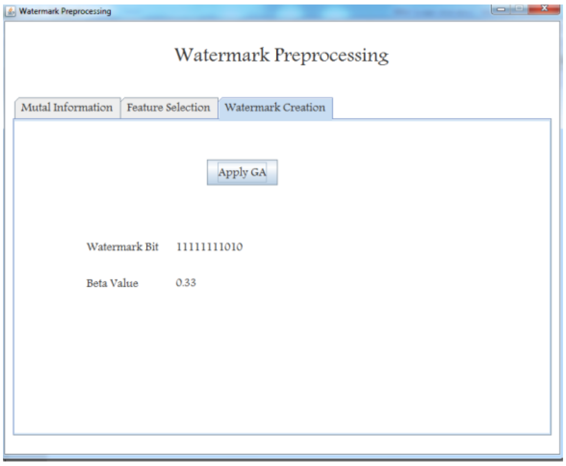
SCREEN SHOTS



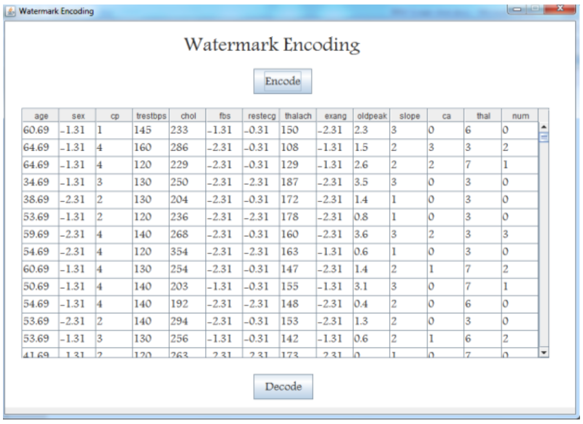
Mutual information:



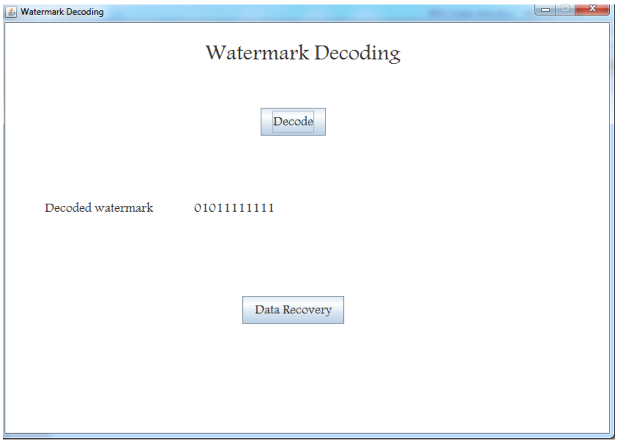
Watermark Creation:



Watermark Encoding:



Watermark Decoding:



CONCLUSION

Irreversible watermarking techniques make changes in the data to such an extent that data quality gets compromised. Reversible watermarking techniques are used to cater to such scenarios because they are able to recover original data from watermarked data and ensure data quality to some extent. However, these techniques are not robust against malicious attacks—particularly those techniques that target some selected tuples for watermarking. In this paper, a novel robust and reversible technique for watermarking numerical data of relational databases is presented. The main contribution of this work is that it allows recovery of a large portion of the data even after being subjected to malicious attacks. RRW is also evaluated through attack analysis where the watermark is detected with maximum decoding accuracy in different scenarios. A number of experiments have been conducted with different number of tuples attacked. The results of the experimental study show that, even if an intruder deletes, adds or alters up to 50 percent of tuples, RRW is able to recover both the embedded watermark and the original data. RRW is compared with recently proposed state-of-the-art techniques such as DEW, GADEW and PEEW to demonstrate that RRW outperforms all of them on different performance merits. One of our future concerns is to watermark shared databases in distributed environments where different members share their data in various proportions. We also plan to extend RRW for non-numeric data stores.

REFERENCES

[1] Y. Zhang, B. Yang, and X-M Niu, "Reversible watermarking for relational database authentication," J. Comput., vol. 17, no. 2, pp. 59–66, 2006

[2] 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

[3] 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

[4] M. E. Farfoura and S.-J. Horng, "A novel blind reversible method for watermarking relational databases," in Proc. IEEE Int. Symp. Parallel Distrib. Process. Appl., 2010, pp. 563–569.

[5] D. M. Thodi and J. J. Rodriguez, "Expansion embedding techniques for reversible watermarking," IEEE Trans. Image Process., vol. 16, no. 3, pp. 721–730, Feb. 2007

[6] X. Li, B. Yang, and T. Zeng, "Efficient reversible watermarking based on adaptive prediction-error expansion and pixel selection," IEEE Trans. Image Process., vol. 20, no. 12, pp. 3524– 3533, Dec. 2011.

[7] E. Sonnleitner, "A robust watermarking approach for large databases," in Proc. IEEE First AESS Eur. Conf. Satellite Telecommun., 2012, pp. 1–6.