ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 6, Special Issue 5, March 2017

National Conference on Advanced Computing, Communication and Electrical Systems - (NCACCES'17)

24th - 25th March 2017

Organized by

C. H. Mohammed Koya
KMEA Engineering College, Kerala- 683561, India

A Robust and Reversible Watermarking Technique for Numerical and Non-Numerical Relational Data

Pranav Prakash [1], Ciya.K.Paul[2], Saliha U.A [3], Vivek S [4]

U.G. Student, Department of Computer Engineering, KMEA Engineering College, Edathala, Kerala, India[1]

U.G. Student, Department of Computer Engineering, KMEA Engineering College, Edathala, Kerala, India[2]

U.G. Student, Department of Computer Engineering, KMEA Engineering College, Edathala, Kerala, India[3]

U.G. Student, Department of Computer Engineering, KMEA Engineering College, Edathala, Kerala, India[4]

ABSTRACT: Nowadays, information handled on network system is in digital form. The copying of digital content without quality loss is not so difficult. Watermarking is technique which provides the protection to the database by embedding information. Reversible watermark technology allows the distortion-free recovery of relational databases after the embedded watermark data are detected. Robust Watermarking ensures the recovery against the attack that causes false insertion, alternation, deletion of the relational databases. In this paper, a watermarking approach based on robust and reversible watermarking technique is used. This technique is used to watermark the numeric as well as the non-numeric features present in the relational database.

KEYWORDS: Robust watermarking, Reversible watermarking, non-numerical, numerical.

I. INTRODUCTION

In today's digital world due to rapid growth of the Internet, the extensive growth of digital contents, and the easier distribution exclusive rights fortification of owners is becoming more and more necessary [1]. Data is generated in the various formats like text, image and relational data. Sharing data online over the internet or virtual storage space is become vital activity for organization and, which may include importing and exporting of relational data [3].Watermarking technique provides the protection to the database by embedding information. In visible watermarking, the watermark information such as text or a logo which identifies the owner of the media, that is visible in the picture or video. Here the objective is to attach ownership or other descriptive information to the signal in a way that is difficult to remove. It is also possible to use hidden embedded information as a means of covert communication between individual. The watermarking technique which are irreversible may causes modification of data undergoes through it at the certain extent [8].

Watermarking can be threaten of malicious attack which may effect as alteration, deletion, or false insertion. The robust watermarking scheme possesses the exact recovery in the existence of active malicious attack [5]. In irreversible watermarking schemes, only the embedded watermark information can be extracted from the suspected data; however, in reversible watermarking schemes, the original objects can be recovered along with the embedded watermarks. Robust and reversible watermarking technique is used to watermark numeric as well as the non-numeric data in the relational database. This method tries to overcome the problem of data quality degradation by allowing recovery of original data along with the embedded watermark information. Watermarking scheme modify original database as a modulation of the watermark information, and causes impossible to prevent permanent distortion to the original database, and affects to meet the integrity requirement of many application [9].

II. LITERATURE SURVEY

The relational database watermarking is of various types as, Reversible watermarking, Irreversible watermarking technique based on various constraints like, Particle swarm optimization, Difference expansion, Difference expansion of triplets, Difference expansion based on SVR prediction. This technique selects the watermark from the features of the dataset which is too strong to attack. As the watermarking make some changes in the original data bits, this technique overcomes this problem by reversible watermarking [4].

Watermarking relational databases is a new research area that deals with the legal issue of copyright protection of relational databases. They focused on embedding short strings of binary bits only in numerical databases. Also they are able to recover original data from watermarked data and ensure data quality to some extent. However, these techniques that target some selected tuples for watermarking based on its mutual information [5]. Although, watermarking technology helps to prove the ownership through identifying data piracy, yet introduces permanent modifications into the data which are irreversible and the watermarked data is different from the original content. Reversible watermarking of relational databases is a candidate solution to ensure data ownership protection as well as data integrity. However a major drawback of these techniques is that they modify the data to a very large extent which often results in the loss of data. There is a strong need to preserve the data quality in watermarked data so that it is of sufficiently high quality and fit for use in decision making as well as in planning processes in different application domains. Ownership rights of the databases need to protect from malicious recipients; in the presence of data quality constraint. GA—an optimization algorithm is employed in the robust and reversible watermarking technique (RRW) to achieve an optimal solution that is feasible for the problem at hand and does not violate the defined constraints. An optimal watermark value is created through the GA and inserted into the selected feature of the relational database in such a way that the data quality remains intact [2].

Genetic Algorithm based on Difference Expansion watermarking (GADEW) technique is used in a proposed robust and reversible solution for relational databases [6]. GADEW improves upon the drawbacks mentioned above by minimizing distortions in the data, increasing watermark capacity and lowering false positive rate. To this end, a GA is employed to increase watermark capacity and minimize introduced distortion. This is because the watermark capacity increases with the increase in number of features and the GA runs on more features to search the optimum one for watermarking. However, watermark capacity decreases with the increase in watermarked tuples. GADEW used the distortion measures (AWD and TWD) to control distortions in the resultant data. In this context, the robustness of GADEW can be compromised when AWD and TWD are given high values. In blind reversible method for watermarking a reversible data embedding technique called Prediction-error Expansion (PE) on integers to achieve reversibility. This technique is fragile against malicious attacks as the as the watermark information is embedded in the fractional part of numeric features only [7].

III. PROPOSED SYSTEM

A. RRW ARCHITECTURE

Architecture section describes RRW for reversible watermarking of relational databases that improves data recovery ratio. The architecture of RRW is shown in Fig. 1. RRW includes the following four major phases: (1) watermark pre-processing; (2) watermark encoding; (3) watermark decoding; and (4) data recovery.

a. WatermarkPre-processing Phase

In the pre-processing phase, two important tasks are accomplished: (1) selection of a suitable feature for watermark embedding; (2) calculation of an optimal watermark.

Feature Analysis and Selection

For developing a decisive information model of various features of the dataset, all the features are ranked according to their importance in information extraction, subject to their mutual dependence on other features. For this purpose, mutual information (MI), is exploited, that is an important statistical measure for computation of mutual dependence of two random variables.

$$\text{MI(A,B)}=\sum_a \sum_b P_{AB}(a,b) log \frac{P_{AB}(a,b)}{P_A(a)P_B(b)} \quad (1)$$

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

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

$$MI_{fi} = (MI_{f_{ij}}) \quad (2)$$

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

The value of MI of each feature is then used to rank the features. The attacker can try and predict the feature with the lowest MI in an attempt to guess which feature has been watermarked. To deceive the attacker for this particular scenario, a secret threshold can be used for selecting the feature for watermark embedding. The feature(s) having lower MI (mutual information) than a particular secrete threshold specified by data owner can be selected for watermarking. With consideration the attacker can try and predict the feature with the lowest MI in an order to guess which feature has been watermarked. It keeps the important feature i.e. high MI feature protected.

Optimal watermark calculation

GA evolves a potential solution to an optimization problem by searching the possible solution space. In the search of optimal solution, the GA follows an iterative mechanism to evolve a population of chromosomes. The GA preserves essential information through the application of basic genetic operations to these chromosomes that include: selection, crossover, mutation and replacement. The GA evaluates the quality of each candidate chromosome by employing a fitness function. The evolutionary mechanism of the GA continues through a number of generations, until some termination criteria is met.

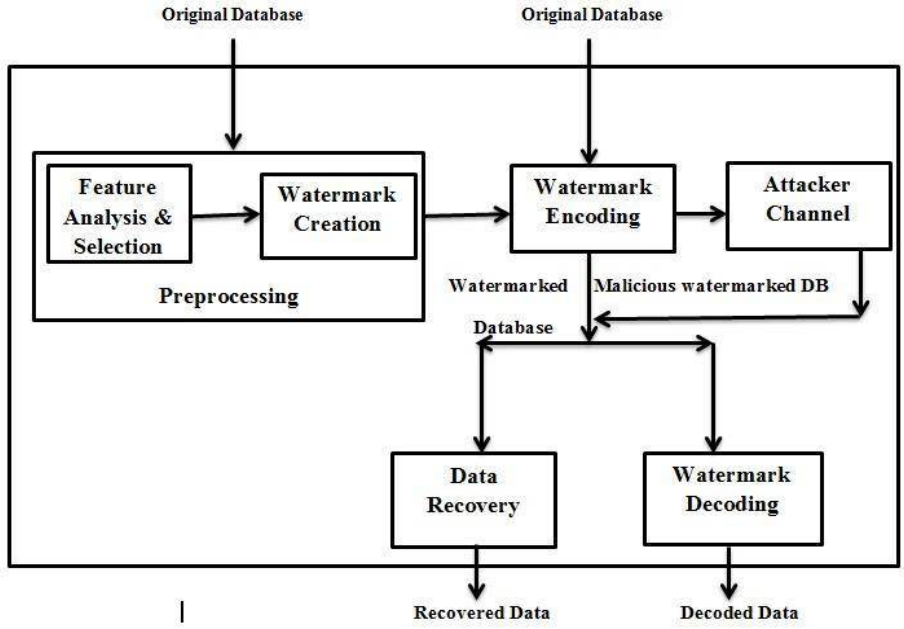

Fig 1. Architecture of RRW

During watermark creation phase, the following major steps of the GA are employed for getting optimal watermark information:

1) Initial random population of binary strings called chromosomes is generated. Gene values of each chromosome represents l-bit watermark string.

2) Fitness of each chromosome is evaluated by employing a constrained optimized fitness function.

3) Tournament selection mechanism is applied to get the most appropriate individuals as parent chromosomes.

4) Genetic operations of crossover and mutation are performed on parent chromosomes to create off springs. A single point crossover operator is applied to evolve high quality individuals, inheriting parental characteristics, by exchanging information between two or more chromosomes. A uniform mutation operator is applied to bring diversity in population through small random changes in gene values of binary chromosomes. The values of crossover fraction and mutation rate are set empirically.

5) Elitism strategy is applied to hire two individuals with best fitness value; as elites to the next generation without genetic changes.

6) Remaining population of the next generation is created by replacing less fit individuals of the previous generation with the most fit newly created off-springs.

7) Steps 2 to 6 are repeated until MIO and MIW reach approximately equal values for a certain number of generations.

8) Both, optimal watermark information string and best fitness value (b) is returned after the fulfillment of the termination criteria.

Elitism strategy is used to create optimal watermark information that need to be embedded in both numerical and non-numerical data. A non-numeric attribute in a relational database can be any string within the range of consonants, vowels or special characters, ie, for example in the case of an employee database, it would the data like such as address and city, whereas numeric data would be the ID values, phone numbers etc. In this selection strategy where a limited number of individuals with the best fitness values are chosen to pass to the next generation, avoiding the crossover and mutation operators. This strategy is known as elitist selection and guarantees that the solution quality obtained by the GA will not decrease from one generation to the next. The optimal fitness value obtained through elitism strategy is basically the change to be embedded in the original data that needs to be watermarked. The purpose of getting an optimum value is to justify the amount of change that a feature value can withhold without compromising the data quality. Digital numbers of each non-numeric attribute we are embedding the watermark information. Here we can use the random number generation algorithm for the creation of optimal watermark information.

b. Watermark Encoding

The main focus of watermark encoding phase is to embed watermark information in such a way that it does not affect the data quality. Watermark embedding phase embeds the watermark in accordance with the encoding strategies and data usability. During this phase, firstly the changes while embedding the watermark into the data are calculated using equation (1). For the non-numeric data, the ASCII value of each character of a string is generated. After the calculation of amount of change while watermark encoding, the data is watermarked. If the MSB of watermark string is 1, then data is watermarked using the equation (2). If the MSB is 0, then data is watermarked using the equation (3)

$$\eta_r = Dr * \zeta \quad (1)$$

ζ is the watermark decoder.

$$DWr = Dr - \beta \quad (2)$$

$$DWr = Dr + \beta \quad (3)$$

After watermarking, the data is released to the attacker channel. The attacker channel refers to all such possible attacks. The attacks can modify some contents of the data with an aim to corrupt the embedded watermark; making the data suspicious.

Watermark Encoding Algorithm:

Input: Database, watermark string, β

Output: Watermarked database, matrix

1. For all watermark bits 1 to length l
2. For all the tuples of the data
3. If watermark bit is 0, then compute changes using equation (1).
4. Watermark data using equation (3).
5. Insert η_r into the matrix.
6. End if.
7. If watermark bit is 1, then compute changes using equation (1).
8. Watermark data using equation (2).
9. Insert η_r into the matrix.
10. End if
11. End For
12. End For

c. Watermark Decoding

The Watermark decoding phase recovers watermark information effectively for detection of the embedded watermark. In this phase the percent change in the watermarked data is calculated using equation (4). After this, the difference between the original data change and the watermark detected change amount is calculated using equation (5). Final watermark information is retrieved through a majority voting scheme using Equation (6).

$$\eta_{dr} = D'W * \zeta \quad (4)$$

$$\eta_{\Delta r} = \eta_{dr} - \eta_r \quad (5)$$

$$WD \leftarrow mode(dtW(1,2,\ldots..,l)) \quad (6)$$

Watermark Decoding Algorithm:

Input: Watermarked Database, matrix containing change in the data values, length of watermark string
Output: Decoded Watermark

1) For all tuples of the watermarked data
2) For all watermark bits b from 1 to length l of the watermark
3) Compute percent change in the watermarked data using equation (4)
4) Compute the difference between original data change amount and the watermark detected change amount using equation (5)
5) If difference computed using equation (4) ≤ 0
6) Detected Watermark bit is 1
7) Else if difference is > 0 and ≤ 1
8) Detected Watermark bit is 0
9) End If
10) End For
11) End For
12) Compute final watermark string using equation (6)

d. Watermark Recovery

After detecting the watermark string, some post processing steps are carried out for error correction and data recovery. The main responsibility of post-processing is to use the decoded watermark bits, and convert these bits into the watermark information that was embedded as the watermark. If the detected watermark bit is 0, the data is recovered using equation (7). If detected watermark bit is 1, then data is recovered using equation (8).

$$Dr = D'Wr - \beta \quad (7)$$

$$Dr = D'Wr + \beta \quad (8)$$

IV. CONCLUSION AND FUTURE WORK

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, we presented a new approach to watermark a non- numeric attribute in the relational database. This algorithm can be used effectively where a huge amount of relational data is transferred between owner and authenticated users. One of our future concerns is to watermark shared databases in distributed environments wheredifferent members share their data in various proportions. A robust and distortion free watermarking technique has been proposed that is capable of recovering the original data. It allows recovery of large amount of the data and embedded watermark even after being subjected to malicious attacks. The existing watermarking technique is used to watermark only the numeric features of the database. But the proposed system allows watermarking of numeric as well as non-numeric features of the relational database. One of the future concerns is to regenerate the true data for all the tuples on deletion of data.

REFERENCES

[1] M.E. Farfour, S.-J. Horng, J.-L.Lai, R.-S Run, R.-JChen, and M. K. Khan, "A blind reversible method for watermarking relational databases based on a time-stamping protocol," Expert Syst. Appl., vol. 39, no. 3, pp. 3185–3196, 2012.
[2] Anitha.T, Bhuvaneshwari.N, Krithicka.S, Nathiya.C, Nithya.L, "A Robust and Reversible Watermarking (Rrw)Technique Based On Genetic Algorithm", International Journal on Applications in Engineering and TechnologyVolume 2: Issue 2: February 2016.
[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. Kamran, Sabah Suhail, and Muddassar Farooq, "A Robust, Distortion Minimizing Technique for Watermarking Relational Databases Using Once-for-All Usability Constraints" In IEEE Transactions on Knowledge and Data Engineering, VOL. 25, NO. 12, DECEMBER 2013.
[5] SamanIftikhar, M. Kamran, and Zahid Anwar, ―RRW—A Robust and Reversible Watermarking Technique forRelational Data,‖ In IEEE Trans. on knowledge and Data Engineering, VOL. 27, NO. 4, APRIL 2015.
[6] 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.
[7] 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. 563569.
[8] R. Agrawal and J. Kiernan, Watermarking relational databases, in Proc. 28th Int. Conf. Very Large Data Bases, 2002, pp. 155166
[9] Y. Zhang, B. Yang, and X.-M. Niu, Reversible watermarking for relational database authentication, J. Comput., vol.17, no. 2, pp. 5966, 2006.