International Journal of Engineering Technology, Management and Applied Sciences

IJETMAS

www.ijetmas.com May 2017, Volume 5, Issue 5, ISSN 2349-4476

Robust Watermarking of Relational Databases Based on Extension of Quantization Index Modulation to Circular Histograms

Miss. Priya G. Tale
Student: P. G. Dept. of Comp. Sci. &Engg.
SGBAU, Amravati
Maharashtra, India

Dr. R. V. Dharaskar
Director: MatoshriPratishthan Group of Institute
SRTMU, Nanded
Maharashtra, India

Dr. V. M. Thakare
HOD: P. G. Dept. of Comp. Sci. &Engg.
SGBAU, Amravati
Maharashtra, India

ABSTRACT

In the internet and cloud computing paradigms the use of digital data is increasing excessively, where such kind of data is stored in various digital formats and also in related nature as a relational data. Watermarking methods has been proposed for multimedia, digital documents, software and, more recently, for databases. Two categories, depending on application, distinguish watermarks: fragile watermarks for tamper detection and robust watermarks for ownership verification.In this paper, a new robust database watermarking scheme the origin of which is based on a semantic control of the data distortion and on the extension of quantization index modulation (QIM) to circular histograms of numeric and non-numeric attributes is proposed. It can be used for verifying database authentication as well as for traceability when identifying database origin after it has been modified. With feature region selection method, a non-overlapping feature region set is selected which has the greatest robustness against various attacks and can preserve data quality as much as possible after watermarked. This work is formulated by a multidimensional knapsack problem (MDKP) and solved by a genetic algorithm based approach. An experimental result of the proposed scheme indicates improved watermark capacity with less distortion and maintains original data quality.

Keywords

Robust Watermarking, Relational Database, Quantization Index Modulation, Circular Histograms.

1. INTRODUCTION

In the rapid growth of internet technology, the relational databases are more remotely accessed and shared not only because of their growing economic value but also because of the evolution of data-mining tools which turn them into a prime component in decision making [1]. However, such an access intensifies risks in terms of security: integrity, authenticity, confidentiality, traceability as well as of copyright protection concerns. Indeed, outsourced data can be rerouted from their final objectives or altered without permission. Every year, several information leaks are reported, even in domains of activity where data are very sensitive such as of defense, healthcare etc. [2][3]. Secure access and confidentiality of data are usually achieved by means of cryptographic mechanisms. Nevertheless, once these mechanisms bypassed or more simply when the access is granted, data are no longer protected [4]. In such a context, watermarking appears as an interesting security mechanism, a posteriori protection that leaves access to data while maintaining them protected in terms of integrity or traceability as example. It encodes the message within host data based on the principle of controlled distortion. The watermark should be imperceptible to users and is independent of the host data file format storage [5].

This paper presents, a new robust watermarking scheme where distortion control method is applied in conjunction with an adaptation of Quantization Index Modulation (QIM); robust modulation according to study up till now has never been considered in database watermarking. This modulation is used so as to modulate the relative angle of the center of mass of circular histograms associated to group of values of one attribute of the relation and is less or not at all sensitive to the rounding integer operation or dependent on the existence of attributes with fractional parts. Moreover, the proposed method is also not depending on the

www.ijetmas.com May 2017, Volume 5, Issue 5, ISSN 2349-4476

storing structure of the database, making it robust to tuple reordering in a relation. The proposed scheme theoretically proved that the use of QIM leads to a scheme that is robust to the most common attacks in the state of the art: tuple insertion and suppression. The proposed scheme also employs a feature region selection method which uses genetic algorithm (GA) based multi-dimensional knapsack problem (MDKP) to investigate two issues of: one is avoiding repeated selection of robust regions for watermarking to resist similar attacks, and the other is the difficulty of selecting the most robust and smallest feature region set to be watermarked.

2. BACKGROUND

A Robust Digital Watermarking scheme has employed a feature region selection method based on the idea of simulated attacking and multidimensional knapsack problem (MDKP) optimization techniques in [1]. This method can be integrated into the feature-based watermarking schemes to enhance their robustness against various types of attack. The experimental results which apply StirMark attacks to some benchmark images watermarked on the feature regions selected by the method in[1] exhibit better robustness in robust digital watermarking than existing methods[1].

The Reversible Data Hiding Scheme divides RDH mechanism into two types: Type I- The features can be formulated as a binary sequence and can be compressed by using a generic compression algorithm, Type II- The features are non-binary and compressed in some specific manners. The scheme in [2] has proved that the recursive code construction can reach the rate–distortion bound when the decompression/compression algorithms used in the code are optimal, which establishes equivalence between source coding and RDH for binary covers [2].

The Reversible Data Hiding and Lossless Data Compression Scheme provide the recursive code construction from binary signals to gray-scale signal, which modifies the histogram in a bin by bin manner according to the optimal transition probability. It also proved that this code can approach the rate-distortion bound as long as the entropy coder reaches entropy. In other words, this code is optimal for RDH if the entropy coder is optimal for lossless data compression (LDC) [3].

A Robust Lossless Database Watermarking Scheme integrates the robust lossless watermarking modulation which originally proposed for images within a common database watermarking scheme. The method employed in [4] is not depending on the storing structure of the database, making it robust to tuple reordering in a relation [4].

A Robust Database Watermarking Scheme with Ontology Guided Distortion Control method is applied in conjunction with an adaptation of Quantization Index Modulation (QIM) in[5]; robust modulation according to study up till now has never been considered in database watermarking. Moreover, the scheme proposed in[5] theoretically proved that the use of QIM leads to a scheme that is robust to the most common attacks in the state of the art: tuple insertion and suppression.

The subsequent structure of this paper is as follows: the brief introduction of proposed technique is given in Section 1. Section 2 discusses background. Section 3 discusses previous work. Section 4 discusses existing methodologies. Section 5discusses analysis and discussion. Section 6describes proposed methodology. Section 7 discusses the possible outcomes and result. Finally section 8 concludes this paper.

3. PREVIOUS WORK DONE

Jen-Sheng Tsai *et al.* (2011) [1] proposed scheme has presented a feature region selection method based on the idea of simulated attacking and multidimensional knapsack problem (MDKP) optimization techniques which investigate two issues of existing feature-based schemes: one is avoiding repeated selection of robust regions for watermarking to resist similar attacks, and the other is the difficulty of selecting the most robust and smallest feature region set to be watermarked.

Weiming Zhang *et al.*(2012) [2] has proposed the scheme which improves the recursive construction by using not only the joint encoding of message embedding and feature compression methods but also a joint decoding

of feature decompression and message extraction. The proposed code proved to be an optimal solution when the compression algorithm reaches entropy. The current codes are designed for binary covers.

Weiming Zhang *et al.* (2013) [3] proposed scheme, employed the recursive code construction from binary signals to gray-scale signal, which modifies the histogram in a bin by bin manner according to the optimal transition probability. This novel code can be easily implemented by recursively applying the decompression process and compression process of an entropy coder.

Javier Franco-Contreras *et al.* (2014) [4] has proposed the robust lossless watermarking modulation technique originally proposed for images by De Vleeschouwe, and integrates it within a common database watermarking scheme. The proposed technique is one that manipulates circular histograms of data and is less or not at all sensitive to the rounding integer operation or dependent on the existence of attributes with fractional parts.

J. Franco-Contreras *et al.* (2015) [5] has proposed the scheme which introduced a new semantic distortion control method which takes advantage of ontology over the database scheme. This scheme shown, that one ontology provides semantic knowledge or description of the database that can help to identify the allowable attribute distortion in a tuple.

4. EXISTING METHODOLOGY

4.1 Robust digital watermarking scheme with feature region selection method

A Robust Digital Watermarking scheme includes a novel feature region selection method for robust digital image watermarking. This method aims to select a non-overlapping feature region set, which has the greatest robustness against various attacks and can preserve image quality as much as possible after watermarked. Firstly it performs a simulated attacking procedure via some predefined attacks to evaluate the robustness of every candidate feature region. According to the evaluation results, it then adopts a track-with-pruning procedure to search a minimal primary feature set which can resist the most predefined attacks [1]. Using the scheme of [1] in the final phase, the most robust and smallest set of non-overlapping feature regions is selected according to the result of attack resistance analysis. This work is formulated as follows:

$$R_p^+ = \arg\max_{R_p} \left\{ \sum_{i=1}^{N_a} x_{R_p}^{R_{a_i}} \middle| \min |R_p|; \right.$$
$$\left. \forall r_k, r_j \in R_p, k \neq j \rightarrow r_k \cap r_j = \emptyset \right\}$$

Where, R_p is the set of selected feature regions in which any two regions r_k and r_j are not overlapped.

4.2 The reversible data hiding scheme

The Reversible Data Hiding Scheme theorize the method using a decompression algorithm as the coding scheme for embedding data and proves that the generalized codes can reach the rate–distortion bound as long as the compression algorithm reaches entropy. By the proposed binary codes, this scheme improve three RDH schemes that use binary feature sequence as covers, i.e., an RS scheme for spatial images, one scheme for JPEG images, and a pattern substitution scheme for binary images. The experimental result shows that the novel codes can notably reduce the embedding distortion. Furthermore, by modifying the histogram shift (HS) manner, the proposed scheme also apply this coding method to one scheme that uses HS, showing that this code can be also exploited to improve integer-operation based schemes [2]. Using the scheme in [2], the reversible embedding capacity ρ_{rev} for a memoryless binary source with p_0>=2 is, for 0<=Δ<=1/2, given by:

$$\rho_{rev}(p_0, \Delta) = H_2\left(\max(p_0 - \Delta, 1/2)\right) - H_2(p_0)$$

IJETMAS
www.ijetmas.com May 2017, Volume 5, Issue 5, ISSN 2349-4476

4.3The reversible data hiding and lossless data compression scheme

The Reversible Data Hiding and Lossless Data Compression Scheme usually consist of two steps: first construct a host sequence with a sharp histogram via prediction errors, and then embed messages by modifying the histogram with methods, such as difference expansion and histogram shift. In this scheme, the focus is on the second stage, and this scheme proposes a histogram modification method for RDH, which embeds the message by recursively utilizing the decompression and compression processes of an entropy coder. Experiments show that this coding method can be used to improve the performance of previous RDH schemes and the improvements are more significant for larger images [3]. Using the scheme in [3], the rate-distortion function, i.e., the upper bound of the embedding rate under a given distortion constraint can be calculated, as follows:

$$\rho_{rev}(\Delta) = maximize\{H(Y)\} - H(X)$$

Where, X and Y denote the random variables of host signal and stego signal respectively.

4.4The robust lossless database watermarking scheme

The Robust Lossless Database Watermarking Scheme has employed the robust lossless watermarking modulation technique; the method used in this scheme can be used for verifying the authenticity of the database and also for verifying its integrity even if the database has been altered by intruder. In addition, the proposed scheme have theoretically established and verified experimentally the performance of proposed method in terms of robustness and capacity against two common attacks: tuple insertion and suppression. The experimental results allow the user to correctly select the proposed scheme parameters under constraints of robustness, capacity and also distortion [4]. Using the scheme in [4], the use of a cryptographic hash function, such as the Secure Hash Algorithm (SHA), ensures the secure partitioning and the equal distribution of tuples into the groups as:

$$n_u = H(K_S|H(K_S|t_u.PK))mod N_g$$

Where, N_g denotes the non-intersecting group of tuples, Ks is the secret watermarking key and tu.PK is the tuple primary key.

4.5The robust database watermarking scheme with ontology guided distortion control method

The Robust Database Watermarking Scheme with Ontology Guided Distortion Control method and quantization index modulation (QIM) to circular histograms of numerical attributes is integrated here. This modulation is used so as to modulate the relative angle of the centre of mass of circular histograms associated to group of values of one numerical attribute of the relation. The semantic distortion control of the embedding process lies on the identification of existing semantic links in between values of attributes in a tuple by means of ontology. This scheme further verifies experimentally the theoretical limits within the framework of database [5]. Using the scheme in [3], the Kullback-Leibler divergence (DKL) and the mean absolute error (MAE) between histograms of the attribute before and after watermark embedding is given by:

$$D_{\mathrm{KL}}(h_{A_t} \| h_{A_t}^{wat}) = \sum_{l=0}^{L-1} \ln\left(\frac{h_{A_t}(l)}{h_{A_t^{wat}}(l)}\right) h_{A_t}(l)$$

$$\mathrm{MAE} = \frac{1}{L \cdot N} \sum_{l=0}^{L-1} \left| h_{A_t}(l) - h_{A_t^{wat}}(l) \right|$$

Where, D_{KL} denotes the Kullback-Leibler Divergence, MAE denotes the mean absolute error, h_{At} and h_{At}^{wat} denotes the histograms of the original attribute A_t and of its watermarked version A_t^{wat} respectively.

5. ANALYSIS AND DISCUSSION

A novel method which is based on the GA-based MDKP solving procedure and the simulated attacking approach is developed to select the most adequate feature regions for robust digital image watermarking under the constraint of protecting image quality. The experimental results which apply StirMark attacks to some benchmark images watermarked on the feature regions selected by this method exhibit better robustness than

existing methods.It may be considered that this method consumes too much computation time in measuring the robustness of feature regions due to the simulated attacking. But in practice, this is not the case of concern [1].

The RDH schemes uses a strategy with separate processes of feature compression and message embedding, the higher embedding rate under a given distortion constraint may be achieved by using joint encoding of feature compression and message embedding. This scheme provides the recursive construction by using not only the joint encoding but also a joint decoding of feature decompression and message extraction. The proposed code construction significantly outperforms previous codes and is proved to be optimal when the compression algorithm reaches entropy. This scheme used only two simple methods to modify HS, and therefore, the scope for other method is opens [2].

The Reversible Data Hiding and Lossless Data Compression Scheme employ the novel code, which can be easily implemented by recursively applying the decompression process and compression process of an entropy coder. This code construction is proved to be asymptotically optimal when the entropy coder is optimal, which establishes equivalency between RDH and lossless data compression. Experiment results shows that the improvements provided by this scheme will be more significant for larger cover images [3].

A robust lossless relational database watermarking scheme which makes use of circular histogram modulation, provides a majority voting mechanism to extract the watermark which is then compared with the original one by means of correlation.This scheme allow the user to correctly select required parameters under constraints of capacity, robustness and also distortion. During tuple deletion attack, the method used in this scheme performs worse under strong attack conditions, i.e. when more than 50% of tuples are removed [4].

A Novel Robust Database Watermarking Scheme , security based on the construction of groups of tuples, this is conducted by means of a cryptographic hash function, e.g., SHA, which takes the secret watermarking key Ks as an input. The use of ontology improves the watermark imperceptibility/ masking. Notice that after this ontology based semantic analysis, the number of watermarked tuples can be reduced, impacting the watermark robustness [5].

Table 1. Comparison Bbetween Existing Methodologies

Watermarking Techniques	**Advantages**	**Disadvantages**
A Robust Digital Watermarking scheme with feature region selection method	The experimental results which apply StirMark attacks to some benchmark images watermarked on the feature regions selected by the proposed method exhibit better robustness.	The proposed method consumes lots of computation time in measuring the robustness of feature regions due to the simulated attacking.
Reversible Data Hiding Scheme	The proposed code is proved to be optimal when the compression algorithm reaches entropy.	The proposed scheme use only two simple methods to modify HS, and there is a scope to find more effective modifying methods whether exist or not.
Reversible Data Hiding and Lossless Data Compression Scheme	The proposed method establishes the equivalency between reversible data hiding and lossless data compression.	The rate distortion bound is not clearly specified.
Robust Lossless Database Watermarking Scheme	In the proposed scheme a majority vote mechanism is used to extract the watermark which is then compared with the original one by means of correlation.	The proposed method is nearly twice slower. The reason may stand in the histogram calculation for each sub-group of tuples.
A Robust Database Watermarking Scheme with Ontology Guided Distortion Control method	The proposed scheme, depending on the embedding modulation can distinguish “attribute-distortion-free” methods that do not modify attributes values from “attribute-distortion-based” methods.	In the proposed scheme the computation time complexity increases along with the number of tuples in the relation.

IJETMAS
www.ijetmas.com May 2017, Volume 5, Issue 5, ISSN 2349-4476

6. PROPOSED METHODOLOGY

A new robust reversible database watermarking scheme using distortion control method is applied in conjunction with an adaptation of Quantization Index Modulation (QIM) in this paper. This modulation is used for modulation of relative angle of the center of mass of circular histograms related to the group of values of any numeric or non-numeric attribute of the relation. Also QIM modulation is robust against various malicious attacks. As the embedding stage includes the preprocessing, the purpose of which is to make the watermark insertion independent of the database relation ordering or the way in which it is stored. To do so, an original database is required, and before message insertion the group of tuples has to be created which gives a set of Ng non-intersecting groups of tuples. The feature region set has to be selected which cause not at all or less distortion to the original data, for this a novel method based on simulated attacking approach and the GA based MDKP solving procedure is used which then select the most adequate feature regions for robust digital relational data watermarking under the constraint of preserving data quality. With Ng groups, the inserted message corresponds to a sequence of Ng symbols. For synchronization insertion of two watermark messages S1 and S2 of different nature is considered, where S1 is made robust by correlation based detection at the reading stage, and S2 is fragile which contains the information required to ensure the reversibility of the scheme. The S1 is inserted in the first Nr groups of tuples, and S2 is inserted into the other (Ng-Nr) groups of tuples. It contains a sequence of bits which encodes the overhead Ov_{info}required for reconstructing the whole database. Ontology in the proposed scheme is useful for having additional semantic pieces about the database content. Also the ontology can be derived from the database with the help of data mining operations. Finally the watermarked data is outsourced in the collaborative environment which might undergoes through some malicious attacks at attacker channel; again while extracting the watermark data at owner side, the preprocessing and demodulation steps has to be followed with the help of which owner can recover the data. The whole procedure of proposed technique is depicted using following Fig. 1.

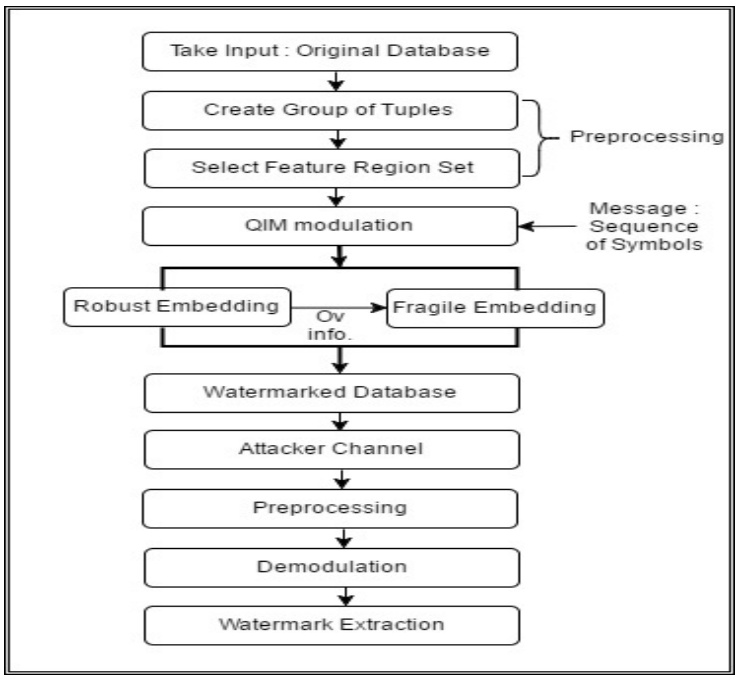

Fig 1: Flow diagram of proposed framework

7. POSSIBLE OUTCOMES AND RESULT

An experimental result of the proposed method indicates improved watermark capacity with less distortion. The experimental results allow the user to correctly select parameters of the proposed scheme under constraints of capacity, robustness and also distortion. The experiments have been conducted on a real life test

database. The use of ontology improves the watermark imperceptibility. The semantic distortion control aims at two main objectives: i) ensuring the correct interpretation of the information available in the database, by securing the semantic links in between attributes values. ii) Making the watermarking imperceptible to an attacker. Also, the QIM leads to a scheme that is robust to the most common attacks like tuple insertion and suppression.

8. CONCLUSION

In this paper, a new robust database watermarking scheme; the origin of which based on a novel semantic distortion control method and a QIM adapted to the modulation of attributes' circular histograms is proposed, where the GA-based MDKP solving procedure is developed to select the most adequate feature regions for robust digital relational data watermarking. The proposed scheme can be used for verifying the integrity of the database and also for verifying its authenticity even if the database has been modified. Message can be embedded by application of QIM to the centre of mass of circular histograms for a numerical and non-numerical attribute. The proposed scheme is robust against most common database attacks like tuple deletion and insertion as well as attributes' values alteration. This proposed scheme is appropriate for copyright protection.

FUTURE SCOPE

From Observation, the scope to be studied in future work, the propose method can be added more efficient methods that will develop faster robust measurement scheme and enhance the proposed method to design secure digital watermarking scheme. The proposed scheme also requires enhancing the performance under strong deletion attack conditions and also to improve the computation speed.

REFERENCES

[1] Jen-Sheng Tsai, Win-Bin Huang, and Yau-Hwang Kuo, "On the Selection of Optimal Feature Region Set for Robust Digital Image Watermarking ", IEEE transactions on image processing, Vol. 20, No. 3, Pg. No. 735-743, March 2011.

[2] Weiming Zhang, Biao Chen, and Nenghai Yu, "Improving Various Reversible Data Hiding Schemes Via Optimal Codes for Binary Covers", IEEE transactions on image processing, Vol. 21, No. 6, Pg. No. 2991-3003, June 2012.

[3] WeimingZhang,Xiaocheng Hu, Xiaolong Li, and Nenghai Yu, "Recursive Histogram Modification: Establishing Equivalency Between Reversible Data Hiding and Lossless Data Compression", IEEE transactions on image processing, Vol. 22, No. 7, Pg. No. 2775-2785, July 2013.

[4] Javier Franco-Contreras, GouenouCoatrieux, FrédericCuppens, Nora Cuppens- Boulahia, and Christian Roux, "Robust Lossless Watermarking of Relational Databases Based on Circular Histogram Modulation", IEEE transactions on information forensics and security, Vol. 9, No. 3, Pg. No. 397-410, March 2014.

[5] Javier Franco-Contreras and GouenouCoatrieux, "Robust Watermarking of Relational Databases With Ontology-Guided Distortion Control", IEEE transactions on information forensics and security, Vol. 10, No. 9, Pg. No. 1939-1952, September 2015.