Research Article

Literati Journals
Empowering Researchers

Securing Numerical Relational Dataset using Robust and Reversible Watermarking Approach based on Genetic Algorithm

Nicy Jose[1]*

Abstract: In this paper, we basically concern for watermark embedding by taking into account of selectively watermark appropriate attribute for the processes of knowledge detection and the reconstruction of the original data in the presence of powerful cyber attacks. Here, robust and the reversible watermarking approach to numerical relational data based on the genetic algorithm has been proposed. Due to the accelerated growth in the management of the information systems that consists of relational databases, these datasets are highly exposed to the cyber threats. Watermarking is advocated to attain ownership claims over commonly shared relational datasets and for delivering a medium for tackling alterations. When we accomplish ownership rights using this approach the basic data, is subjected to some alterations as an aftereffect of which, the feature of the data gets mediated. Thus the reversible watermarking approach is employed to assure quality and restoration of the data. The analysis of the experiments shows the efficiency of this approach against malicious attacks and display that these techniques outrun the existing ones.

Asian Journal of Engineering and Technology Innovation
Volume 2016, Issue 2
Published on: 7/05/2016

Cite this article as: Nicy Jose. Securing Numerical Relational Dataset using Robust and Reversible Watermarking Approach based on Genetic Algorithm. Asian Journal of Enginnering Technology and Innovation, Vol 2016(2):170-174, 2016.

INTRODUCTION

In this today's digital era, due to the emerging use of cyberspace and cloud computing environment data is getting multiplied in a very huge amount. Storage of data is carried out in versatile forms such as images, audio, video and relational data. In order to protect these different forms digital watermarking approaches are being used for the ownership claims. Due to the circulation of the databases across the internet and cloud environment, the worry for the ownership protection of the relational datasets has arrived. For the purpose of extraction of knowledge and the planning processes, the owners of the data make data publicly available. As a result, these publicly available data's are more exposed to cyber intrusions. Even though there exists numerous watermarking approaches for protecting the ownership claims, it causes changes in the original data, such a way that the data after embedding the watermark is totally different from the initial content.

Reversible watermarking is an approach that looks at following new considerations such as (1) embedding of the watermark into the relational datasets comprising tuples/records (2) the organization of records (3) operations of the data in relational dataset. As we embed the watermark, data within the relational databases can be modified according to acceptable bandwidth. Nevertheless the bandwidth should not be too large to interrupt the data kind. The range to which alteration of data is allowed is in such a way, the data kind before and after watermarking is agreeable. It is the data holder who decides the amount of alteration that could be carried out within the relational databases without negotiating the data kind.

The basic watermarking technique is depicted above. During watermark embedding there includes a private key K which is known only to the owner and that key is used to embed watermark bits into the database which is original. Then that watermarked database is made available as public. In order to validate the suspicious database the validation process is carried out. Here by taking the private key which is used during watermark embedding, the watermark which is embedded is extracted from the doubtful database and then compared with the original database.

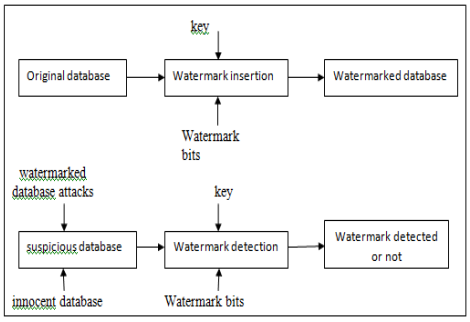

Figure 1: Basic Watermarking Technique

[1]Shree Devi Institute of Technology Kenjar Airport Road, Mangalore – 574 142 Karnataka State, india.
E-mail: ashwin@revainstitution.org
*Corresponding author

The four phases included in the reversible watermarking are (1) preprocessing (2) watermark encoding (3) watermark decoding (4) data recovery. This technique needs a proper mechanism to embed the watermark, so that the original data also the embedded watermark is able to recover. Based on the alterations brought-in to the original data, digital watermarking approaches can be categorized into distortion-free and distortion based watermarking. In distortion-based approach, they alter the data during the embedding process for safeguarding the ownership rights. These techniques should reduce the falsification into an adequate level so that data will be beneficial for the knowledge extraction methods. At the same time in case of distortion-free approach, they do not alter the original content during embedding. In those techniques, the embedding process is carried out in permuted or hashed values of the original content. Here the extraction of watermark is carried out from hashed values to allegate ownership.

The watermarking approach that is used for validating data uprightness and detection of data tampering are fragile. The embedding processes in these approaches are carried out with a goal that whenever the original data undergoes any alteration purposefully or malignantly by the attackers, the watermark should get easily breached. Therefore, these fragile approaches are not advisable for ownership claims. Therefore, for safeguarding the ownership and copyright claim there is a need of powerful watermarking in such a way that embedding technique should be strong against malicious attacks.

This paper proposes robust and reversible watermarking methods for the numerical data in order to safeguard the ownership claims of a database owner based on genetic algorithm. This technique facilitates the data to get validated and then reconstructed in the original form by taking-off the watermark and then recovering the original data.

RELATED WORK

Reversible Watermarking for Relational Database Authentication

The approach proposed in this [1] is based on histogram expansion. Histogram expansion technique is used to reversibly watermark the selected nonzero digits of errors to form histograms[1]. Here they maintain records of overhead information to validate the data kind. This scheme obtains the capability of exact reconstruction of the original data from the untampered watermarked relational datasets. In this approach, only the data owner has the capability to recover the original database state. Here they provide excellent and lossless design to validate the relational databases which are mainly suitable for sensitive data that does not have any durable falsification. This approach takes benefit of the irregular circulation of errors between randomly generated neighboring values to get reversibly watermarking and display the capability to restrict the watermarking distortion to the necessity of useful utilization by taking incomplete real values to calculate error using initial digit requirement. Nonetheless, this technique is not powerful against malicious attacks that are subjected to a large number of tuples.

Reversible and Blind Database Watermarking Using Difference Expansion

This paper [2] proposes a blind and a reversible watermarking technique based on difference expansion. A blind watermarking approach needs only the embedded watermark object and secret key to detect watermarks whereas the non-blind watermarking approach needs an unmarked multimedia object in addition to the secret key and the embedded watermark object. The drawbacks of non-blind watermarking scheme is that the object that is unmarked need to be stored in a secondary storage location and should be given back to the detection algorithm later. In this scheme they make use of the difference expansion technique of integers to attain reversibility. The important advantages supplied by these particular schemes are reversed to huge condition of original data content, legitimate owner recognition, resistance against subordinate attacks, and storing of the original dataset at a protected secondary storage is not at all required. In order to minimize the distortions watermark is embedded into the LSB of the features of relational datasets. This distortion is reduced to a great extent in this approach.

Reversible Fragile Database Watermarking Technology Using Difference Expansion Based on SVR Prediction

This paper [3] proposes a method to safeguard the database from being tampered. The intention behind the design of these techniques is to provide ownership proof [3]. But these methods are more prone to alteration attacks such as embedded watermark and the original data will not be able to detect if there is any change in the expanded value.

Genetic Algorithm and Difference Expansion Based Reversible Watermarking For Relational Databases (GAEDW)

The approach proposed in this [4] is the robust and the reversible solution for relational databases. In this paper they use the concept of both difference expansion and genetic algorithm (GA) in order to make better both watermark capacity and to reduce distortion. Since the proposed method is reversible, alterations made after watermark embedding can be fully recovered. Here data distortions are reduced by the increasing watermark capacity and by decreasing false positive rate. However, as the increase in the watermarked tuples decreases the watermark capacity. This is because the

watermark capacity increases with the increase in the number of features and the GA runs on more features to search the optimum one for watermarking [4]. Rather than selecting less effective features for a peculiar attribute, different attributes are explored using GA. Also distortion tolerance of different attributes is explored because analysis of distortion of only two attributes for a particular tuple may not be useful for watermark capacity as well as distortion. Therefore, they included both tuple and attribute wise distortion within the fitness function of the GA, making hard for an attacker to guess watermarked attribute. The technique proposed in this paper is highly robust against a broad range of attacks such as insertion, alteration and deletion attacks also various tuple-wise and attributes wise attacks

A Novel Blind Reversible Method For Watermarking Relational Databases

In this paper [5] they offer a novel blind reversible watermarking approach that assures protection of ownership claims in the Relational Databases. The previous approaches were mainly dealing with the introduction of long-lasting errors in the original content. In this particular approach, 100 percent reconstruction of the original database content after the watermark has been recognized and validated. Reversible watermarking is a technique in which one is able to recover the original data after validation of the contents. This property is mainly needed in case of sensitive applications such as medical and military data. Alteration of the data permanently is one of the main disadvantages of the entire irreversible relational database watermarking schemes. Here they use an embedding technique which is reversible called Prediction-error Expansion (PEE) on integers. Even though seventy percent of the watermarked relation tuples are deleted, the watermark detection can be completed successfully. Here are some of the desired properties of this particular approach proposed in this paper includes imperceptible, robustness, reversibility, blind detection and prevention of legitimate embedding of a watermark and authentication.

Expansion Embedding Techniques For Reversible Watermarking

In this paper [6] they propose a shifting method in the histogram expansion as an alternative method for embedding a location map. Reversible watermarking facilitates the useful information embedding in a host signal without any loss of data. Difference-expansion suggests a high-capacity, reversible technique for embedding the data. But the problem with these techniques is that they suffer from unsatisfactory distortion at low capacities for watermark embedding and lack of capacity control because of the need for embedding a location map. All these problems have been overcome by this proposed method. This paper also proposes a reversible watermark embedding methodology called prediction-error expansion. In this technique they make use of the interaction built-in the neighborhood of a pixel than difference-expansion. In short, they have combined together both the Prediction-error expansion and histogram shifting technique combined together to form an effective approach towards data embedding.

Efficient Reversible Watermarking Based On Adaptive Prediction-error Expansion and Pixel Selection

In this paper [7] they had modified the concept of Prediction Error Expansion (PEE) and more efficient reversible watermarking approach is proposed. Modification to the PEE is carried out by incorporating two new strategies, namely, adaptive embedding and the selection of the pixels. The previous techniques used PEE in such a way that the data is embedded uniformly. But in the proposed technique, they adaptively embed 1 or 2 bits into expandable pixel according to the local complexity. This expansion of pixels with large prediction errors is avoided thereby reduces embedding impact by lowering the maximum alteration to pixels. Also, within a single embedding process PEE allows very large payload and hence increases the capacity limit of traditional PEE. The selection of the pixels is done by detecting smooth area and leave rough pixels unchanged. Therefore, a more sharply distributed prediction-error histogram and better visual quality of watermarked image is observed when compared to the traditional PEE.

A Robust Watermarking Approach For Large Databases

In this paper [8] they introduce a robust, blind, resilient and reversible; image based watermarking methodology for large databases. Watermarking is especially useful in publicly available web-databases. In order to assure data quality, detect the alterations made in the database and also to safeguard the ownership rights, this paper proposed an algorithm for watermarking based on parameterized tuple partioning and white spaces, using a public watermark. For watermarking they use the bit strings of the image. One bit from the bit string image is inserted in all the tuples of a single partition, and the same process is repeated for the rest of the partitions. The schema proposed here shows an extraordinary decrease in the detection of watermark due to the numerous attacks and as a result the tuples in the database gets distorted highly.

PROPOSED WORK

In this paper, we propose a robust and semi-blind reversible watermarking approach for securing numerical relational dataset. Here they combined the efficiency features of the previous papers such as robustness, semi-blind nature and the reversible watermarking techniques to make this proposed

technique highly efficient than the existing ones. There are mainly four phases in this proposed system. They are pre-processing phase, watermark encoding phase, attacker channel, watermark decoding phase and the data recovery phase.

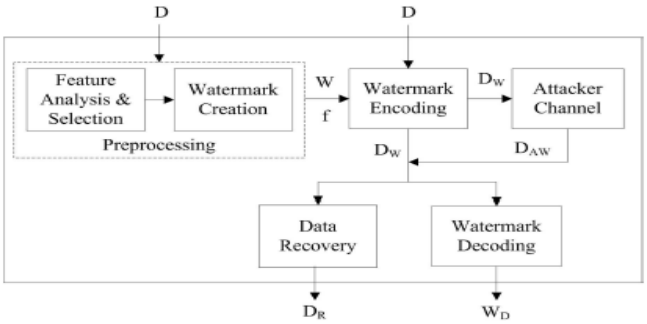

Figure 2: Overall Architecture

The first phase is the pre-processing. During the first phase, two tasks are carried out. They are feature selection and the generation of the watermark. In order to select the feature first they are ranked according to their importance in the knowledge extraction process. Ranking is done on the basis of mutual information (MI). MI calculates the range to which the features are correlated. This is measured by using the joint probability distribution and the marginal probability distribution.

Depending upon the MI value of all the features, the owner of the database defines a secret threshold. The threshold is calculated for all the features by using the confidence factor, mean and standard deviation. For watermarking, the features whose MI value less than the threshold is selected. The features that have larger MI will not be attacked by the attacker because in that case the knowledge cannot be preserved for the future decision making process. Hence, due to the secret threshold the attacker is forced to attack the features with smaller MI without any proper knowledge of which features have been watermarked. The next task is to generate watermark. An optimum watermark bit string is carried out by an evolutionary technique known as Genetic Algorithm (GA). The optimum string is found by searching in the possible solution space and remains intact even after watermark embedding.

GA is inspired from genetic evolution. An iterative mechanism is proposed for the optimal string. It starts with random generation of individuals initially. In order to select individuals for the next generation, they make use of evaluation function. Evaluation function is calculated by measuring the fitness of individuals. Operations of GA include selection, crossover, mutation and replacement. The quality of each individual is evaluated by the fitness function. The individuals with high fitness value are selected for the next generation. The selection technique used here is the tournament selection. Two randomly generated individual strings are compared and the better one is choose among two. After the selection of the parent individual string, the crossover operation is applied to the offspring. The crossover operator uses a random point for the operation, through which the flipping of the bits takes place. Then all the offspring are subjected to mutation. Every bit is flipped by the mutation operator with pre-specified probability function. Until the individuals in the population are completed this process is continued. As we embed the watermark information, the original data undergoes certain changes, but these changes should be in a way that the quality of the data should not get compromised. For this purpose, the GA is populated with a constrained fitness function to get the best value for the amount of change a feature can withhold so that quality remains intact. Data quality is assured by making mean, variance and the mutual information of the original as well as watermarked data same.

The next phase is the watermark encoding. Here the optimum watermark string length and the amount of change that is acceptable calculated from pre-processing phase is used. Then the optimum value change is added to each tuple of the selected feature

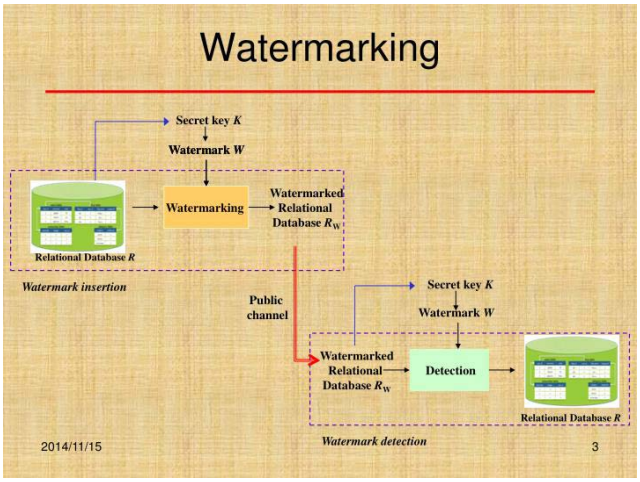

Figure 3: Watermark Encoding

The main task of this phase is to insert watermark into the database and the mutual information remains unchanged even after the insertion of watermark information. Based on MI and the secret threshold value, the data owner can choose the number of features for watermark embedding. The main goal of this technique is in such a way that data quality should not get changed even after watermark embedding. After the encoding process, the data are then released to the communication channel where attackers can modify the data by corrupting the watermark information and making the data suspicious.

The third phase is the watermark decoding. Embedded watermark is decoded from the distrusted data Here mainly there are two tasks.

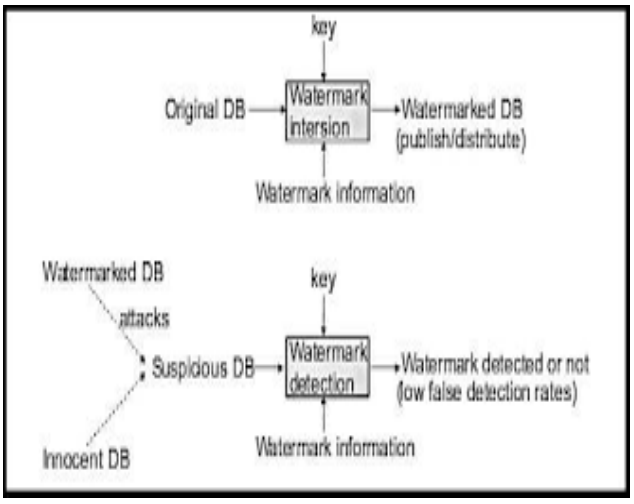

Figure 4: Watermark Decoding

The first task is to locate the features that have watermarks. Detection of the watermark bits is carried out from LSB towards MSB. Since it is easy to detect the effect of the last encoded bit of the watermark detection of the watermark bits is carried out in the reverse order. The next task is to decode the bits according to percentage change values of the watermarked data. The watermark information is verified by comparing the original watermark and the detected watermark.

The final phase is the data recovery. Here original data is recovered. Also conversion of the decoded watermark bits to get the watermark information.

EXPECTED RESULTS

The proposed watermarking approach is expected to provide effective performance than the existing ones. In this proposed watermarking following Keep Performance Indicator (KPI) are chosen to compare the performance of the watermarking:

Data Distortion Ratio

The watermarking approach used here will use the optimum bandwidth for embedding the watermark. Therefore, this technique is expected to keep data quality throughout the database. The distortion does not have any impact on the original data.

Resilient Against Attack Ratio

The watermarking approach uses the robust nature. This technique is expected to perform in such a way that it is highly resilient against many types of attacks such as alteration attacks, insertion attacks and the deletion attacks.

Watermark Detection

The proposed technique will make the watermark to be detected with or without the presence of attacks.

Data Recovery

The technique proposed uses reversible nature to recover the data.

CONCLUSION AND FUTURE WORK

In this paper we proposed both robust reversible watermarking techniques for numerical relational dataset. Irreversible watermarking techniques will make modification in such a way that the data quality will get negotiated. But for the proposed reversible watermarking technique, the data quality remains intact even after embedding watermark information. Here GA is proposed to find the optimum for the selection of features based on MI. Here we mainly focus on watermarking numerical data of the relational databases. Even after being subjected to heavy attacks large portion of the data is able to recover from the original data. This is the main contribution of this paper. This approach improves the performance by combining all the efficient features of the previous techniques. The future work of this paper is to embed the watermark in the distributed environment where different databases are shared by different people in various ratios.

REFERENCES

1. Y. Zhang, B. Yang, and X.-M. Niu, "Reversible watermarking for relational database authentication," Journal of Computers, vol. 17,no. 2, pp. 59–66, 2006.
2. 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.
3. J.-N. Chang and H.-C. Wu, "Reversible fragile database watermarking technology using difference expansion on svr prediction," in Computer, Consumer and Control (IS3C), 2012 International Symposium on. IEEE, 2012, pp. 690–693.E. H.
4. K. Jawad and A. Khan, "Genetic algorithm and difference expansion based reversible watermarking for relational databases,"Journal of Systems and Software, 2013.
5. 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.
6. D. M. Thodi and J. J. Rodriguez, "Expansion embedding techniques for reversible watermarking," Image Processing, IEEE Transactions on, vol. 16, no. 3, pp. 721–730, 2007.
7. X. Li, B. Yang, and T. Zeng, "Efficient reversible watermarking based on adaptive prediction-error expansion and pixel selection," Image Processing, IEEE Transactions on, vol. 20, no. 12, pp.3524–3533, 2011.
8. E. Sonnleitner, "A robust watermarking approach for large databases," in Satellite Telecommunications (ESTEL), 2012 IEEE First AESS European Conference on. IEEE, 2012, pp. 1–6.