Secured Relational Database Watermarking Using Genetic and Firefly Optimization Algorithms

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

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

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

Abstract— *The information technology plays a vital role during the rapid growth of using information systems comprising relational data. Also, the relational information is important in many areas such as medical, military, forensics and stock market, where data should be distributed to users from a centralized database. The relational databases which is used in shared environment for extracting the information; inevitably they are susceptible to security threats concerning data tampering and ownership rights. The outsourcing data results in a number of threats such as alteration of data, deletion of data, accessing by unauthorized people, producing unauthorized copies. Recently, many methods are proposed to watermark databases to protect digital rights of owners. When the watermark is enforced on data for protecting ownership right, the data quality get compromised. For protecting the data quality a robust and reversible watermarking technique for numeric and non-numeric relational data is required. Particularly, watermarking techniques based on optimization draw attention, which results in improving watermark capacity and lower distortion. In this paper, A Robust Reversible Watermarking (RRW) with Genetic Algorithm (GA) and Firefly Algorithm (FFA) is proposed to embed watermark into relational databases. GA and FFA are optimization techniques which are biologically inspired. Best attribute values are selected efficiently by the FFA and later GA is used for creating the optimum watermark string which ensures reversibility without data quality loss. Experimental results indicate that FFA and GA have reduced complexity and results in improved watermark capacity and less distortion.*

Keywords— Reversible Digital Watermarking, Relational Database, Robustness, Genetic Algorithm, Firefly Optimization Algorithm.

I. Introduction

The advancement of computers, digital communications and the internet in the banking and social media welfares are expanding rapidly with different advent features where exchanging of digital data becomes very simpler [1]. Also with this the exponential increase in internet users is also growing rapidly together with the accessibility of technological knowledge which also may lead to new proportions of crime, as the digital data like image, text, audio, video can be easily duplicated and modified [2]. A major weakness of digital technology is how easily illegal and unauthorized reproduction and distribution of digital objects is achieved. Watermarking adds a level of protection to the copyright of digital assets [3]. It is the process of making deliberate alterations in a digital object, providing that they can be detected further. In general, watermark is small, hidden perturbations in the database used as an evidence of its origin, which means to determine the paternity of the object [3]. Inserting mark into original data is used to find the ownership [4]. Watermark should not significantly affect the quality of original data and should not be able to destroy easily. The marks are applied by an encoder. The detection of the watermark is achieved through the use of a key which is known only to data owner. Watermarking methods has been proposed for multimedia, digital documents, software and more recently, databases. Consequently, it has become relevant to explore suitable watermarking techniques for ownership rights protection of relational databases that should be robust, imperceptible, with blind decoding [4]. Also, once the owner of data embeds the watermark, the distortions in the original data are kept within certain limits, which are defined by the usability constraints; to preserve the knowledge contained in the data. An intended recipient wants the data owner to define tight usability constraints so that he gets accurate data. Remember, the robustness of a watermark is measured by the watermark decoding accuracy [5].

This paper presents semi-blind and robust reversible watermarking (RRW) technique with FFA and GA for non-numerical and numerical relational data. FFA and GA an optimization algorithms are employed in the proposed RRW technique, where FFA algorithm is used for efficiently selecting the best attribute values to increase watermark capacity and producing lower distortions in relational data and GA is used to embed watermark in the selected attribute to achieve an optimal solution that is feasible for the existing problem at hand and does not violate the defined constraints. The watermark optimal value created through GA is embedded into the selected feature of the relational database in such a manner, through which the data quality is preserved. Consequently, RRW provides a better and robust solution for data recovery that is resilient and reversible against heavy attacks.

II. Background

The robust reversible watermarking technique, tries to overcome the problem of data quality degradation by allowing recovery of original data along with the embedded watermark information and maintaining the robust nature of data. Digital watermarking can also be modeled as an optimization problem as demonstrated by some recent research works and that use PSO for watermarking different data formats and the results are quite encouraging [1].

The novel watermarking scheme for numeric database attributes which is efficient in defeating a range of attacks that may be used to destroy or remove the mark is proposed in [2]. The watermark might be any digital object related to the underlying data, for example a logo, a text message, an image a sound, a speech signals, etc. The encoding algorithm can be applied to each tuple independently, i.e., the watermarked database can support normal user modifications by simply using the encoding algorithm to those involved tuples, without affecting any other ones [2].

A relational database robust watermarking algorithm based on embedding a binary image watermark into numerical attributes, and textual attributes with arbitrary number of words is proposed in [3]. This method is robust against common attacks such as subset addition, alteration and deletion attacks. The main aspect of this method is to embed different watermark signals into numerical data and textual data in relational databases, respectively. The embedded watermark can be obtained only by the owner of the database who owns the secret key [3].

A novel watermark decoding algorithm ensures that its decoding accuracy is independent of the usability constraints proposed in [4].This algorithm embeds every bit of a multi-bit watermark in each selected row as in a numeric attribute with the objective of having maximum robustness even if an attacker is somehow able to successfully corrupt the watermark in some selected part of the data set. The method proposed in [4] provides solutions to resolve conflicting ownership issues in case of the additive secondary attack in watermarked database.

A novel right protection scheme that will establish the ownership of EMR data and consequently, will make its illegal sale very difficult is proposed in [5]. This proposed scheme introduced the concept of information-preserving watermarking. Once the watermark is computed, it is inserted by utilizing the knowledge of correlation of those features that have not much impact on the diagnosis. The proposed scheme in [5] used the Particle Swarm Optimization (PSO) to create an optimized watermark that once inserted into an EMR does not change the diagnosis rules.

The subsequent structure of this paper is as follows: the brief introduction of RRW technique with two optimization algorithms FFA and GA is given in Section I. Section II discusses background. Section III discusses previous work. Section IV discusses existing methodologies. Section V describes analysis and discussion. Section VI discusses proposed methodology. Section VII discusses the possible outcomes and result. Finally section VIII concludes this paper.

III. Previous Work done

Saman et al. (2015) [1] proposed RRW technique which is robust and reversible where data quality is preserved by taking into account the importance of the features in knowledge discovery. RRW outperforms existing state of the art reversible watermarking techniques including DEW, GADEW and PEEW. These techniques embed the watermark in partitions of the data to ensure minimum distortion; therefore, recover original data with degraded data quality and lack robustness. RRW has overcome drawbacks of these techniques and is also resilient against heavy attacks.

Theodoros Tzouramanis et al. (2011) [2] proposed a novel watermarking scheme for relational data which is efficient against a range of attacks that may be issued to remove or destroy the watermark. The proposed scheme provides experimental results for a variety of parameter settings, revealing the robustness of this proposed scheme in numerous possible attacks. The encoding algorithm can be applied to each tuple independently.

Lizhong Zhang et al. (2011) [3] proposed method works for numerical attributes and textual attributes, with watermark blindness. This method makes stego-channel of watermark expand to non-numerical data, helpful for improving the robustness of the watermark. Here, linefeed character and carriage return character, representing 0 and 1 of watermarking bits respectively, are inserted into textual data, which does not change original appearance and meaning of textual data in relational databases.

M. Kamran et al. (2013) [4] proposed technique is able to meet the conflicting “robustness requirement” of the data owner and “minimum distortions requirement” of the intended recipient. This scheme tries to focus a balance between the conflicting requirements of database owners, who require soft usability constraints, and database recipients who want tight usability constraints that ensure minimum distortions in the data.

M. Kamran et al. (2012) [5] proposed technique, rank attributes on the basis of their importance in a decision making process. The proposed scheme objective is to identify weak attributes by developing a knowledge model that correlates the effect of an attribute on the decision making process. The proposed technique uses an optimization technique to first create a watermark which ensures that data usability constraints are not violated. Once the watermark is created, the proposed scheme embed it in real-time into an EMR system.

IV. Existing Methodology

A. Robust Reversible Watermarking Technique

A robust and semi-blind reversible watermarking (RRW) technique for numerical relational data has been introduced, in which an optimal watermark value is created through the GA and inserted into the selected numeric feature of the relational database in such a way that the data quality remains intact. RRW mainly comprises a (1) data pre-processing phase, (2) watermark encoding phase, (3) attacker channel, (4) watermark decoding phase and (5) data recovery phase. The major contributions of this work are: (a) the design of an intelligent reversible watermarking technique for relational data that ensures data recovery without compromising data quality, and (b) a robust data recovery scheme that is resilient against malicious attacks [1]. Using the scheme in [1], the value of a numeric feature is recovered using following equations:

$$D_r = D'_{W_r} + \beta,$$
$$D_r = D'_{W_r} - \beta.$$

Where, D_r denotes a tuple in the database table, D'_{W_r} denotes the tuple in the watermarked database table.

B. The Novel Watermarking Scheme

The Novel Watermarking scheme embeds the watermark on the group basis. The tuples are uniformly divided into |W| groups, using a mixed sequence of $\log_2$|W| msbs and lsbs of the attribute that will be watermarked and, afterwards, one bit of watermark information is stored in each group. Therefore, the only information which needs to be saved in a safe storage regarding this process is $\log_2$|W| + 1 short integer values. It is obvious that a great advantage of the proposed method against other group based watermarking techniques is that there is no need to store large quantities of information related to the constructed groups of tuples, like the number of groups, the number of tuples in each group, the tuples that define the borders of each group, the parameters of the function which distributed the tuples in the groups, and any other related information regarding the groups' content. Therefore the proposed method offers an almost blind decoding process [2].

C. The Novel Relational Database Robust Watermarking Scheme Suitable For Some Numerical and All Textual Data

The Novel method presents a relational database robust watermarking algorithm based on embedding a binary image watermark into numerical attributes, and textual attributes with arbitrary number of words. This is done by embedding special mark and watermark bits into textual attributes and numerical attributes respectively. Carriage return character and linefeed character, representing 1 and 0 of watermarking bits respectively, are inserted into textual data, which does not change original appearance and meaning of textual data in relational databases. For numerical data, Watermarks are embedded into one of least significant bits (LSB) of the optional attributes [3]. Using the scheme in [3], the location of the embedded watermark bit for each textual data is given by:

$$Lt_{i,j} = H_i \bmod length(r_i.A_j), 1 \leqslant i \leqslant \eta, 1 \leqslant j \leqslant v.$$

Where $Lt_{i,j}(1 \leqslant i \leqslant \eta, 1 \leqslant j \leqslant v)$ denote the remainder of tuple hash divided by textual data length.

D. Robust, Novel and Efficient Watermarking Scheme for Relational Databases

A robust and efficient watermarking scheme used for ownership protection of relational databases shared with collaborators or intended recipient's demands developing a watermarking scheme that must be able to meet various challenges like: the proposed scheme is robust against different types of attacks that an intruder could launch to corrupt the embedded watermark; the proposed scheme is able to preserve the knowledge in the databases to make them an effective component of knowledge-aware decision support systems; the proposed scheme tries to strike a balance between the conflicting requirements of database owners, who require soft usability constraints, and database recipients who want tight usability constraints that ensure minimum distortions in the data. Here the date-time attribute is used to generate the watermark bits [4]. Using the scheme in [4], in a data set D, a function Φ is used to calculate data selection threshold for constructing D'T from D:

$$\Phi : D \rightarrow D'_T.$$

The data selection threshold for an attribute is calculated by using the following equation:

$$T = c * \mu + \sigma$$

Where, μ is the mean, σ is the standard deviation of the values of an attribute A in D, and c is the confidence factor.

E. An information preserving watermarking scheme

An information-preserving right protection of EMR systems framework is introduced. This framework operates in two modes: 1) Watermark encoding, and 2) Watermark decoding. In the watermark encoding phase, the main goal is to determine a watermark that, once inserted into an EMR does not alter the important features to an extent that the patient is misdiagnosed. Similarly, the goal of the decoding phase is to accurately detect a watermark in an efficient manner [5]. Using the scheme in [5], the decoding algorithm selects one row at a time, for every watermarked feature, and calculated $\eta_{d_{f_c}}$ as:

$$\eta_{d_{f_c}} = \beta_{f_c} * (f_{cw})$$

Where, $\eta_{d_{f_c}}$ denotes the detected amount of change in the value of a feature after an attack of the watermark bit b.

V. Analysis and Discussion

In RRW all the tuples of the selected feature were watermarked to achieve robustness. This proposed technique results have shown 100 per cent accuracy in both watermark detection and data recovery. The major contributions of this work includes: (1) the design of an intelligent reversible watermarking technique for relational data that ensures data recovery without degrading data quality, and (2) a robust data restoration scheme that is resilient against subset alteration, subset deletion and subset insertion attacks. This technique is not tested with non-numeric data and shared distributed environment which has to be considered in further work [1].

A novel watermarking scheme is used for relational data which is efficient against a range of attacks that may be issued to remove or destroy the watermark. In this scheme the proposed method for watermarking numeric relational data is provided. This proposed scheme embeds the watermark on the group basis. The proposed method offers an almost blind decoding process. To ensure the robustness of the watermark, only, the watermark to modified tuples in order to keep the distribution of tuples per watermark bit as uniform as possible is required [2].

A novel method suitable for watermarking some numerical data and all textual data of relational databases is proposed here. Experiment shows that even the database suffers from approximately 70% of various attacks, the watermark detected match reaches up to 95%. Only the data owner having secret key is viable to obtain the embedded watermark [3].

A novel watermark decoding algorithm that ensures that its decoding accuracy is independent of the usability constraints is proposed here. The date-time is used to generate the watermark bits. The proposed algorithm embeds every bit of a multibit watermark in each selected row in a numeric attribute. The robustness of this proposed watermarking scheme is proved by analyzing its decoding accuracy under different types of malicious attacks using a real world data set. This scheme provides solutions for secondary attacks [4].

A novel information-preserving technique for watermarking EMR is presented. The benefits of this technique are preserving high ranking features, empirically selecting the length of watermark to ensure real-time computability constraints, ensuring the usability constraints, and decoding the watermark using majority voting based on a novel watermark decoder. The proposed scheme has performed extensive experiments to test the effectiveness and accuracy of the proposed watermarking technique [5].

TABLE I. COMPARISON BETWEEN EXISTING METHODOLOGY

Watermarking Techniques	**Advantages**	**Disadvantages**
Robust Reversible Watermarking technique	1) It provides data quality and data recovery. 2) RRW is also evaluated through attack analysis.	1) This technique is used only for numerical data. 2) This technique is not used for shared databases in distributed environment.
Novel Watermarking Scheme	The encoding algorithm can be applied to each tuple independently.	The proposed scheme is unable to make correct watermark recovery decisions in view of brute force and mix-and-match type of attacks.
The Novel Relational Database Robust Watermarking Scheme	The proposed scheme can be applied on numeric as well as non-numeric data.	The proposed method cannot provide 100% extraction of the affected data.
Robust, Novel and Efficient Watermarking Scheme for Relational Databases .	The proposed scheme, implemented watermark decoding algorithm ensures that its decoding accuracy is independent to the usability constraints.	The proposed technique is restricted to numeric unsigned data only.
An information preserving watermarking scheme	The proposed scheme, ensuring the usability constraints.	The proposed technique is not tested on nonnumeric strings data and medical images.

VI. Proposed Methodology

A new robust reversible database watermarking scheme using Firefly and Genetic bio-inspired optimization algorithms, were proposed here. The proposed method uses RRW technique to reverse the original database after watermark extraction and it also uses optimization algorithms called firefly algorithm to determine the best candidate pairs to embed the watermark and genetic algorithm for creating the optimum watermark string. The prime advantage of the Firefly Algorithm is its easy implementation and its run time efficiency. The proposed method mainly consists of two algorithms: watermark insertion and watermark extraction. Watermark insertion algorithm embeds the watermark information into specially selected tuples with RRW. Firefly Algorithm determines which attributes are more appropriate for watermark embedding on the selected tuples. Watermark extraction algorithm extracts the specially embedded watermark information from the database and compares it with the original data. This algorithm also reverses the watermarked database in the original form after watermark extraction and verification.

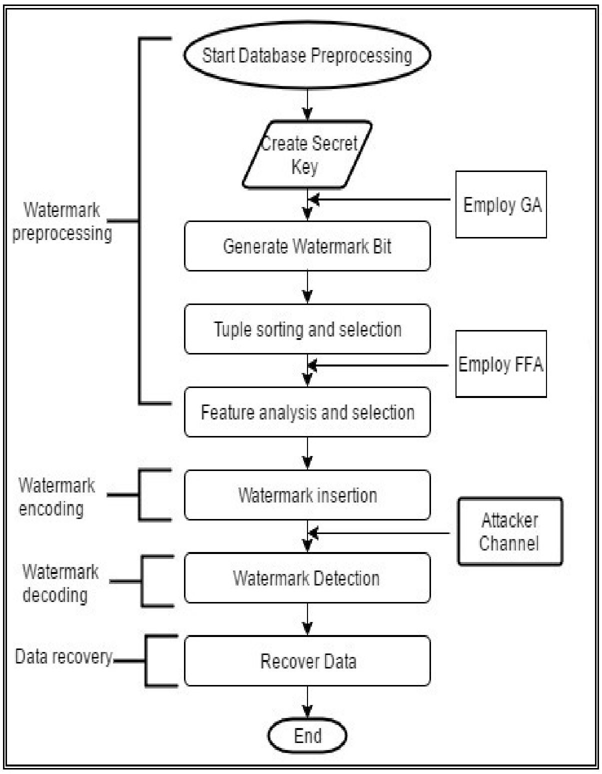

Fig 1: Flow Diagram of Proposed Framework

Fig 1. Shows the flow diagram summarizing the main phases of the proposed watermarking technique as follows:

1) The first phase is of Watermark pre-processing –

This phase prepares the database for watermarking process as the first step. An optimum watermark string is created in this phase by employing genetic algorithm, that ensures reversibility without data quality loss and generate the watermark bits. A robust watermark algorithm is used to embed watermark bits into the data set. The watermark embedding algorithm takes a secret key and the watermark bits as input and converts a data set into watermarked data set. The columns of the database are sorted according to the names of attributes and the tuples of the database are sorted according to the primary key. Sorting operation ensures algorithm’s robustness. Firefly algorithm outputs the best firefly for dataset.

2) The second phase is of Watermark encoding –

This phase embeds watermark data into dataset according to the best firefly in the selected feature(s). Finally, the watermarked data for destined recipients is generated. The attacker channel comprises subset deletion, subset alteration and subset insertion attacks generated by the intruder. These malicious attacks alter and modify the original data and try to pervert its quality.

3) The third phase is of Watermark decoding –

In this phase the embedded watermark is decoded from the suspicious data. In order to achieve this pre-processing step is repeated again, and decoding strategies are used to recover the watermark. Semi-blind nature of RRW is used mainly for data reversibility in case of heavy attacks.

4) The fourth phase is of Data recovery-

Original data is recovered in data recovery phase, through post processing steps for error correction and recovery.

VII. Possible Outcomes and Result

Experimental results of the proposed method indicates improved watermark capacity with less distortion faster than similar works reported in the existing techniques. The propose method is efficient and time complexity is better than existing systems. RRW with FFA and GA algorithms are evaluated for: (1) interrogating effect on the data quality of the elementary data; (2) robust quality against malicious attacks; and (3) recovery of the original data. Robustness of RRW is demonstrated through an immense attack analysis. The attack analysis can be performed on uniformly distributed synthetic real life datasets or some standard datasets available for testing.

VIII. Conclusion

Relational database watermarking has become an active research as the demand for sharing information increases rapidly. In this paper, a robust and reversible technique for watermarking numerical and non-numerical data of relational databases is proposed. The main contribution of the proposed technique is to allow recovery of a large portion of the original data even after being subjected to malicious attacks. The attack analysis is also used for evaluating the proposed RRW technique where the watermark is detected with maximum decoding accuracy. Firefly and Genetic, the new bio-inspired optimization algorithms, are adopted and used with RRW in this work to both minimize distortion and reduce complexity during database watermarking. The number of experiments has been conducted with different number of tuples attacked.

IX. Future Scope

From Observation, the scope to be studied in future work, the propose method can be added more efficient methods that will develop robust watermarking scheme that can be used to watermark shared databases in distributed environments where different members share their data in various proportions.The future concern is at making correct watermark recovery decisions in view of other types of attacks, for example, brute force and mix-and-match attacks.

Acknowledgment

I, Priya G. Tale thankful of my supervisors Dr. R. V. Dharaskar and Dr. V. M. Thakare for helping to complete this research paper.

References

[1] Saman Iftikhar, M. Kamran, and Zahid Anwar, “RRW—A Robust and Reversible Watermarking Technique for Relational Data”, IEEE transactions on knowledge and data engineering, Vol. 27, No. 4, Pg. No. 1132-1145, April 2015.

[2] Theodoros Tzouramanis, “A Robust Watermarking Scheme for Relational Databases”, 6th International Conference on Internet Technology and Secured Transactions, Pg. No. 783-790, December 2011.

[3] Lizhong Zhang, Wei Gao, Nan Jiang, Liqiu Zhang, Yan Zhang, “Relational Databases Watermarking for textual and numerical data”, International Conference on Mechatronic Science, Electric Engineering and Computer, Page No. 1633-1636, August 2011.

[4] M. Kamran, Sabah Suhail, and Muddassar Farooq, “A Robust, Distortion Minimizing Technique for Watermarking Relational Databases Using Once-for-All Usability Constraints”, IEEE transactions on knowledge and data engineering, Vol. 25, No. 12, Pg. No. 2694-2707, December 2013.

[5] M. Kamran and Muddassar Farooq, “An Information-Preserving Watermarking Scheme for Right Protection of EMR Systems”, IEEE transactions on knowledge and data engineering, Vol. 24, No. 11, Pg. No. 1950-1962, November 2012.