Scientific Journal Impact Factor (SJIF): 3.518

International Journal of Modern Trends in Engineering and Research

IJMTER

ISSN (ONLINE) : 2349 - 9745

ISSN (PRINT) : 2393 - 8161

REVERSIBLE AND ROBUST WATERMARKING FOR RELATIONAL DATABASES

Sampada Shukla[1], Shivani Vanarse[2], Sonali Paradhi[3] and Pooja Panpatil[4]

[1,2,3,4] Computer Engineering , K.K.Wagh College Of Engineering, Education and Research

Abstract- In the today's digital world, data is being generated at an excessive pace due to the increasing usage of Internet and cloud computing. Data is stored in different digital formats such as images, audio, video, and natural -language texts and also as relational data. Relational data in particular is shared extensively in research communities and in virtual data storage locations in the cloud. Their main purpose is to work in a collaborative environment and make data openly available so that it becomes useful for knowledge extraction and decision making. These databases are used effectively in collaborative environments for information extraction; consequently they are vulnerable to security threats such as ownership rights and data tampering. Watermarking is a way to enforce ownership rights over relational data and for providing means for tracking data tampering. When ownership rights are enforced using watermarking, the data undergoes certain modifications due to which, the data quality gets compromised. Reversible watermarking is a technique employed to ensure that the data in question retains its original quality along with data recovery. However, these techniques are usually not safe against attacks and do not provide any mechanism to watermark only a particular attribute by also considering its role in further knowledge discovery. Therefore, a robust reversible watermarking technique is required that ensures: (I) Watermark encoding and decoding by accounting for role of the features in knowledge discovery. (II) Recovery of original data in the case of active malicious attacks. To overcome the problem of data quality degradation and various forms of malicious attacks, a robust and reversible watermarking technique for string and numerical relational data will be implemented.

Keywords- Reversible watermarking, data recovery, robustness

I. INTRODUCTION

Watermarking techniques have been historically used to Ensure security in terms of protection of ownership and tamper proofing for a wide variety of data formats. This includes images, audio, video natural language processing software relational databases and more. Reversible watermarking techniques can ensure data recovery along with ownership protection. RRW mainly comprises of a (I) Preprocessing of data phase,(II) Watermark Encoding Phase, (III) Attacker Channel,(IV) Watermark Decoding Phase and (V) Data Recovery Phase. In data preprocessing phase, hidden parameters are defined and several strategies are used to rank and analyze the features that are to be watermarked. In RRW, all the tuples of the selected feature can be marked by selection of a low distortion watermark; therefore, the attacker will need to attack all the tuples to corrupt the watermark to overcome the effect of the majority voting scheme. Attacking all the tuples is not feasible for the attacker since he doesn't have knowledge of the original data or the constraints of usability and that compromises its usefulness to a great extent. Furthermore, since RRW can afford to embed watermark bits in a large fraction of the tuples of selected feature; it achieves a high robustness against heavy attacks. However, marking all tuples is not necessary. RRW can be configured in such a way that the data owner is able to choose a fraction for watermarking if required. RRW beats existing state of the art reversible watermarking techniques like GADEW, DEW and PEEW in terms of performance. These techniques work by embedding the watermark in portions of the data to ensure minimum distortion; therefore, they recover original data with degraded data quality and lack robustness. RRW has successfully overcome drawbacks of these techniques and is also immune to some extent against these heavy attacks.

1.1 Do We Need New Watermarking Techniques for Relational Data?

There is a rich body of literature on watermarking multimedia data . Most of these techniques were initially developed for still images and later extended to video and audio sources. While there is much to learn from this literature, there are also new technical challenges due to the differences in the characteristics of relational and multimedia data. These differences include: _A multimedia object consists of a large number of bits, with considerable redundancy. Thus, the watermark has a large cover in which to hide. A database relation consists of tuples, each of which represents a separate object. The watermark needs to be spread over these separate objects. _The relative spatial/temporal positioning of various pieces of a multimedia object typically does not change. Tuples of a relation on the other hand constitute a set and there is no implied ordering between them. _Portions of a multimedia object cannot be dropped or replaced arbitrarily without causing perceptual changes in the object. However, the pirate of a relation can simply drop some tuples or substitute them.

II. DIFFERENT TYPES OF ATTACKS

Generally, the digital watermarking for integrity verification is called fragile watermarking as compared to robust watermarking for copyright protection. In a Robust watermarking scheme, the embedded watermark should be robust against various attacks which aim at removing or distorting the watermark. While in a fragile watermarking scheme, the embedded watermark should be fragile to modifications so as to detect and localize any modification in presence of different attacks.

The watermarked database may suffer from various types of intentional and unintentional attacks which may damage or erase the watermark, as described below:

1. Update: In this case, the tuples or data of any watermarked relation are processed as usual. As a result, the marked tuples may be added, deleted or updated which may remove the embedded watermark or may cause the embedded watermark undetectable (for instance, during update operation some marked bits of marked data can be erroneously flipped). This type of processing are performed unintentionally.

2. Value Modification Attack:

- **Bit Attack:**

This attack attempts to destroy the watermark by altering one or more bits in the watermarked data. More information about the marked bit position makes attack more successful. However, in this case usefulness of data is crucial: more alternation may result the data completely useless.

Bit attack may be performed randomly which is known as Randomization Attack by assigning random values to certain bit positions; or by Zero Out Attack where the values in the bit positions are set to zero; or may be performed by inverting the values of the bit positions, known as Bit Flipping Attack.

- **Rounding Attack:**

Mallory may try to lose the marks contained in a numeric attribute by rounding all the values of the attribute. Success of this attack depends on the estimation of how many bit positions are involved in the watermarking. Underestimation of it may cause the attack unsuccessful, whereas overestimation may cause the data useless.

- **Transformation:**

An attack related to the rounding attack is one in which the numeric values are linearly transformed. Among users.

3. Subset Attack: Mallory may consider a subset of the tuples or attributes of a watermarked relation and by attacking (deleting or updating) on them he may hope that the watermark has been lost.

4. Superset Attack: Some new tuples or attributes are added to a watermarked database which can affect the correct detection of the watermark.

5. Collusion Attack: This attack requires the attacker to have access to multiple fingerprinted copies of the same relation.

- **Mix-and-Match Attack:**

Mallory may create his relation by taking disjoint tuples from multiple relations containing similar information.

- **Majority Attack:**

This attack creates a new relation with the same schema as the copies but with each bit value computed as the majority function of the corresponding bit values in all copies so that the owner cannot detect the watermark.

6. False Claim of Ownership: This type of attack seeks to provide a traitor or pirate with evidence that raises doubts about merchant's claim.

- **Additive Attack:**

Mallory may simply add his watermark to Alice's watermarked relation and try to claim his ownership.

- **Inevitability Attack:**

Mallory may launch an inevitability attack to claim his ownership if he can successfully discover a fictitious watermark which is in fact a random occurrence from a watermarked database.

7. Subset Reverse Order Attack: Attacker enjoys this attack by exchanging the order or positions of the tuples or attributes in relation which may erase or disturb the watermark.

III. MATH

The mutual information of every feature to be watermarked with all other features is calculated by using the following equation:

$$MI(A,B)= \sum_{a}\sum_{b} P_{AB(a,b)} \log P_{AB}(a,b)/ P_{A(a)} P_{B(b)}$$

Where MI(A,B) is the degree of correlation of features which measuring the marginal probability distributions as PA(a),P_B(B) and the joint probability distribution PAB(a,b).

Then MI of one feature with all other features can be calculated by using the following relation:

MI_{FI}= (MI_{FIJ})

Where i, j ¼ 1, 2 . . . ; ft with i 6¼ j, and ft are the total number of features. Then the value of MI of each feature is used to rank the features.

IV. APPLICATIONS OF DIGITAL WATERMARK FOR RELATIONAL DATABASES

Digital Watermarks for relational databases are potentially useful in many applications, including:

1. Ownership Assertion: Watermarks can be used for ownership assertion. To assert ownership of a relational database, Alice can embed a watermark into her database *R* using some private parameters (*e.g.* secret key) which is known only to her. Then she can make the watermarked database publicly available. Later, suppose Alice suspects that the relation *S* published by Mallory1 has been pirated from her relation *R*. The set of tuples and attributes in *S* can be a subset of *R*. To defeat Mallory's ownership claiming, Alice can demonstrate the presence of her watermark in Mallory's relation. For such a scheme to work, the watermark has to survive intentional or unintentional data processing operations which may remove or distort the watermark.

2. Fingerprinting: Fingerprinting aims to identify a traitor. In the applications where database content is publicly available over a network, the content owner would like to discourage unauthorized duplication and distribution by embedding a distinct watermark (or fingerprint) in each copy of the database content. If, at a later point in time, unauthorized copies of the database are found, then the origin of the copy can be determined by retrieving the fingerprint.

3. Fraud and Tamper Detection: When database content is used for very critical applications such as commercial transactions or medical applications, it is important to ensure that the content was originated from a specific source and that it had not been changed, manipulated or falsified. This can be achieved by embedding a watermark in the underlying data of the database. Subsequently, when the database is checked, the watermark is extracted

V. Helpful Hints

A. Figures and Tables

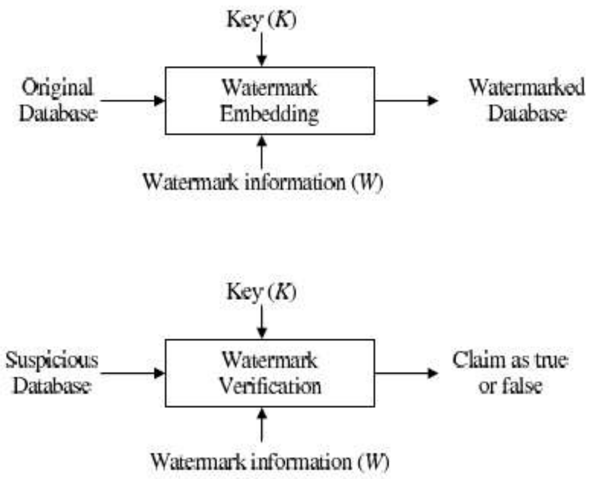

Figure 1: Basic Watermarking Technique

Reversible watermarking can be configured so that the data owner can choose a fraction for watermarking when required. RRW outperforms current state of the art reversible watermarking techniques including PEEW, DEW, and GADEW. These techniques embed the watermark in partitions of the data so as to ensure minimum distortion; therefore, the data they recover lacks robustness and quality. RRW has very successfully overcome drawbacks of these techniques and is also immune against heavy attacks to a large extent.

VI. Conclusion

Irreversible watermarking techniques make changes in the data to an extent that data quality of data gets compromised. Reversible watermarking techniques come to rescue in such scenarios because they are able to recover original data from watermarked data and ensure quality of data to some extent. However, these techniques are not immune against attacks — specially those techniques that target some selected tuples for watermarking.

REFERENCES

[1] J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, “Secure spread
[2] spectrum watermarking for multimedia,” IEEE Trans. Image Process.,vol. 6, no. 12, pp. 1673–1687, Dec. 1997..
[3] Y.-R. Wang, W.-H. Lin, and L. Yang, “An intelligent watermarking method based on particle swarm optimization,” Expert Syst. Appl., vol. 38, no. 7, pp. 8024–8029, 2011..