Watermarking Of Relational Databases

Mohit H Bhesaniya[1], Kunal Thanki[2]
[1]*Department of computer engineering, AITS, Rajkot,Gujarat, India.*
[2]*Department of computer engineering, Government polytechnique, porbandar.Gujarat, India.*
[1]mohit.bhesaniya@gmail.com
[2]kunal.thanki11@gmail.com

Abstract: As a tool for storing and managing data, relational database is widely used in many information systems. It is very critical issue to protect the copyright of relational data. As an invention of information hiding the digital watermark techniques have been attracting more interests in both research and industrial fields. Today database watermarking becomes the research topic because of the increasing use of relational database systems. In order to make watermarking Information more intuitive and easier to identify and to give authorization to the database, watermarking is been proposed. Watermark describes information that can be used to prove the ownership of data such as the owner, origin or recipient of the content. Watermarking can be used for ownership verification of a database by inserting an imperceptive watermark in such a way that it provides robustness and security against attempts to remove the watermark.

Keywords: Database Security, Database Watermarking, Copyright Protection, Watermark Embedding, Watermark Detection.

I. INTRODUCTION:

Recently, the copyright protection of database is important in the field of information technology. This is partly because that the digital data is very easy to be tampered, peculated and illegality copied. Digital watermarking is an approach to cope with this situation. Protection from the piracy of digital assets is usually based upon the embedding of digital watermarks into the data.

Watermarking technique can be used for ownership verification of a digital product by inserting an imperceptive watermark in such a way that it provides robustness against attempts to remove the watermark. It also provides a promising method of protecting digital data from illicit copying and manipulation by embedding a secret code directly into the data. In general, the database watermarking techniques consist of two phases: ***Watermark Embedding*** and ***Watermark Verification***. During watermark embedding phase, a private key *K* (known only to the owner) is used to embed the watermark *W* into the original database. The watermarked database is then made publicly available. To verify who the owner of a suspicious database the verification process is performed where the suspicious database is taken as input and by using the private key K the embedded watermark is extracted and compared with the original watermark information.The novel approach for database watermarking is that in which watermark embedding is done by some secret key and for the verification process there must be no need of original database. This means that there should be blind watermark verification available.

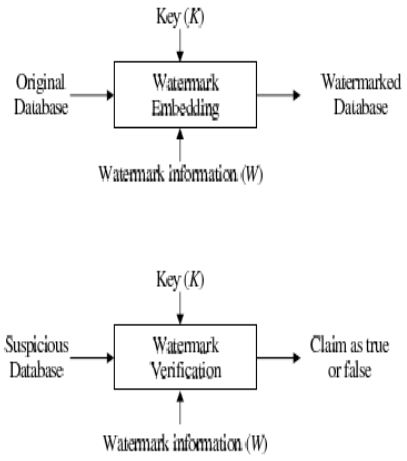

FIG. Fig.1 WATERMARKING PHASES

II. LITERATURE SURVEY:

In the year of 2000 S. Khanna proposed the idea of controlling the security of database with digital watermark which arouses the researcher's interests on watermarking database [3].

First method for relational database watermarking was exposed by Agrawal and Kiernan in 2002. This Method marks only numeric attributes and marking done at a bit level.For embedding images as a watermark in database there are basically following methods. First method was proposed in 2004 by Zhang [2]. This method uses the order of the marked data for embedding image. Due to the fact that the pixels of image have relative position, and they are sure to be placed in order, this method is vulnerable to conflicting order of embedded marks by some subset attacks.In 2008 Sun [7] introduced another robust technique for embedding images into the database. This method transforms some meaningful images as copyright information into a bit flow. In this method unlike other algorithms, location of each mark bit (pixel) is calculated according to the hash value that is related to one of the database tuples. Hence this technique can resist against mark out of ordering. On the other hand as this algorithm uses the remainder of hash value modulo length of watermark, to calculate and find specific mark bit location, the algorithm will have a problem if length of watermark (image) increases, or number of database tuples decreases. So this method is vulnerable in placing all of the mark (image) bits and the probability of missing mark bits is notable, especially in case of embedding big size image into the small size database.

Wang et al. [8] describes an image based watermarking scheme which embeds a scrambled image based on Arnold transform. This technique converts a scrambled image into binary string. Suppose the length of this binary string is L. And whole database is divided into L groups. It computes 1 hash value using private key, database's primary key and order of an image. According to this hash value, a particular group among all L groups are found. The ith bit of binary string is inserted into the specific bit position of the attribute value. This ith bit is chosen algorithmically [8]. The efficiency of this technique is improved because it depends on many factors like private key, scrambling Number and order of an image. This technique uses only one fixed attribute to insert watermarks.Cao et al. [9] introduced a new technique that uses EMC (Encrypted Mark Code) to convert the original image into bit flow and then, the similar steps are used as in [8]. This technique does not consider the order of the image also. At last, the usefulness of the database is checked. The modification is committed if acceptable, otherwise rolled back.Watermarking of relational databases is relatively very new research area that deals with problem of relational database and its copyright protection. Therefore literature in this area is very limited.

2.1 Types of Database Watermarking

The need for watermarking database relations is to deter their piracy, identify the unique characteristics of relational data which pose new challenges for watermarking and provide desirable properties of a watermarking system for relational data. Proving ownership rights on outsourced relational databases is a crucial issue in today's internet based application environments and in many content distribution applications.

2.1.1 Distortion base watermarking

The watermarking techniques in this category introduce small changes in the underlying data of the database during embedding phase. The degree of changes should be such that any changes made in the data are tolerable and should not make the data useless. The watermarking can be performed at bit level, or character level, or higher such as attribute or tuple level, over the attribute values of type's numeric, string, categorical, or any [5].

2.1.2 Distortion free watermarking:

Most of the distortion free watermarking techniques are fragile in the sense that in addition to the ownership claiming, they aim at maintaining the integrity of the information in the database. The watermark insertion phase does not depend on any specific type of attribute and does not introduce any distortion in the underlying data of the database [5].

2.2 Applications of Digital Watermark for Relational Databases:

2.2.1 Ownership Assertion:

Watermarks can be used for ownership assertion. To assert ownership of a relational database, watermark can embed a bit into database R using some private parameters (e.g. secret key) which is known only to user. Then the watermarked database can make publicly available. Later, suppose any suspects that the relation S published by some has been pirated from the relation R. The set of tuples and attributes in S can be a subset of R. To defeat ownership claiming, one can demonstrate the presence of her watermark in its relation. For such a scheme to work, the watermark

has to survive intentional or unintentional data processing operations which may remove or modify the watermark.

2.2.2 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.

2.2.3 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.

2.3 Different types of attacks:

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

2.3.1 Benign Update

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 [6].

2.3.2 Deletion attack

The Attacker deletes marked tuples from the relational database which leads to synchronization errors.

2.3.3 Alteration attack

Attacker alters the data values of the tuples which leads to disturbance in the watermark. Altering the data values violates the usability constraints and makes the data useless.

2.3.4 Insertion attack

Attacker inserts tuples to the data set hoping to disturb the embedded watermark which results in synchronization errors.

2.4 Cryptography Overview:

Cryptography is used for user authentication and it protects data from theft and alteration. There are basically three types of mechanisms used to obtain above goals.

2.4.1 Secret Key Cryptography (SKC)

It uses a single key for both encryption and decryption Examples: Data Encryption Standard (DES), Advanced Encryption Standard (AES), Rivest Ciphers, and Seeded.

2.4.2 Public Key Cryptography (PKC)

It uses one key for encryption and another for decryption. Examples: RSA, Diffie Hellman and ElGamal.

2.4.3 Hash Functions

Uses a mathematical transformation to irreversibly "encrypt" information Examples: Message Digest (MD) algorithms, hash of variable length, whirlpool etc.

Table 1: Comparison of Distortion base watermarking:

Proposed scheme	Watermark Information	Cover Type	Granularity Level	Verifiability	Intent
AHK Algorithm	Meaningless bit pattern	Numeric	Bit level	Blind, Private	Ownership proof
Gupta et. Algorithm	Meaningless bit pattern	Numeric	Multi bit Level	Blind , reversible, private	Ownership proof
Image Based Watermarking	Image	Numeric or non numeric multi word	Bit level ,whole attribute value or char level	Blind private	Ownership proof and temper detection
Speech based	Owners speech	Numeric	Bit level	Blind, Private	Ownership proof
Content based	Database content	Numeric	Multi bit Level	Blind ,private	Ownership proof and/or temper detection and localization
Cloud model Based	Cloud model with three characteristics Expected value ,entropy, hyper entropy	Numeric	Whole attribute value	Non blind Private	Ownership proof
Categorical attribute based	Meaning full binary string	Categorical	Bit level	Blind and Private	Ownership proof
Fake tuple Based	Fake information obtain from the database contain	Database table	Tuple level	Blind private	Ownership proof
Virtual attribute based	Database contain	Database table	Attribute Level	Blind , deterministic, private	Temper detection
Others	Meaningful information of any type	Numeric	Bit level	Blind or non blind or private	Ownership proof
Fingerprinting Techniques	Meaningful fingerprint identifying buyers uniquely	Numeric	Bit level	Blind private	Trailer detection

Table 2: Comparison of Distortion free watermarking:

Proposed scheme	Watermark Information	Cover Type	Granularity Level	Verifiability	Intent
Permutation Based	A part of group level Hash value	Tuples positions	Tuple level	Blind private	Temper detection
Characteristics Based	White image with owner‘s mark at four corner	Nil	Nil	Blind private	Temper detection
Binary form Relation	Relation in binary form	Nil	Nil	Blind public or private	Ownership proof or temper detection
R tree based Scheme	Numeric value identifying owner	Order of entries in r tree nodes	R tree nodes	Blind private	Temper detection

III.CONCLUSIONS:

In this paper we reviewed various techniques for database watermarking. We can insert an image as watermark in our database in various ways. We have also derived drawbacks of various watermarking methods. We should note that all the techniques of watermarking database can only be applied for real time protection of database and copyright protection. These methods cannot be used to prevent piracy or illegal copying of database.

Methods of Watermarking of relational databases are basically divided into two types, Distortion Based and Distortion Free methods. Watermarking of Relational Database can be implemented by Image Based Watermarking which is actually Distortion Based Watermarking. After studying various research papers and approaches we can assure that image based watermarking can be efficiently used for Security and authentication of database. Image based database watermarking is also useful for ownership protection. This technique will be useful for fraud and temper detection also.

ACKNOWLEDGEMENTS:

I am heartily thankful to my guide, Prof. J.N.Rathod, who encourages me and provided me necessary guidance. I also want to thank my family and friends for helping and supporting me.

REFERENCES:

[1]. Andri Furmanyuk, Mykola Karpinskyy, Bohdan Borowik, ‖ Modern Approaches to the Database Protection”. IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems” 2007.

[2]. Zhi-Hao Zhang, Xiao-Ming jin, jain-Min wan, ―Watermarking Relational Database Using Image. IEEE Proceedings of the Third International Conference on Machine Learning and Cybernetics, Shanghai, 26-29 August 2004.

[3]. Sanjeev Khanna, Francis Zane.‖ Watermarking maps: hiding information in Structured Data. Int’l Conf. SODA 2000, San Francisco, California, USA.

[4] R.Barnett ―Digital watermarking: applications, techniques and challenges. Electronics &

Communication Engineering Journal AUGUST 1999.

[5] Raju Halder, Shantanu Pal, Agostino Cortesi, Watermarking Techniques for Relational Databases: Survey‖, Classification and Comparison. Journal of Universal Computer Science, vol. 16, no. 21, 2010.

[6] R. Agrawal and J. Kiernan." Watermarking relational databases". In Proceedings of the 28th International Conference on Very Large Databases VLDB, 2002.

[7] Jianhua Sun, Zaihui Cao, Zhongyan Hu, 2008. Multiple Watermarking Relational Databases Using Image, In IEEE International Conference on Multimedia and Information Technology, p. 373-376.

[8] Chaokun Wang, Jianmin Wang, Ming Zhou, Guisheng Chen, Deyi Li, 2008. Atbam: An Arnold transform based method on watermarking relational data, In Proceedings of the 2008 International Conference on Multimedia and Ubiquitous Engineering, p. 263-270.

[9] Zhongyan Hu, Zaihui Cao, Jianhua Sun, 2009. An Image Based Algorithm for Watermarking Relational Databases, In Proceedings of the 2009 International Conference on Measuring Technology and Mechatronics Automation, p. 425-428.

