ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Issue 3, March 2016

A Study of Watermarking Relational Databases

Gishma.K.M[1], Jucy Vareed[2]

M-Tech student, Department of Computer Science & Engineering, Vidya Academy of Science &Technology, Thrissur, Kerala, India [1]

Asst. Prof, Department of Computer Science & Engineering, Vidya Academy of Science &Technology, Thrissur, Kerala, India [2]

Abstract: Digital multimedia watermarking technology was suggested in the last decade to embed copyright information in digital objects such as images, audio and video. However, the increasing use of relational database systems in many real-life applications created an ever increasing need for watermarking database systems. As a result, watermarking relational database systems is now merging as a research area that deals with the legal issue of copyright protection of database systems. Here in this paper, first we look at basics of digital watermarking then applications of watermarking as security in database. We also look at few techniques developed by different researchers in the field of database watermarking and a generic framework of reversible watermarking in relational database.

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

I. INTRODUCTION

The piracy of digital assets such as software, images, video, audio and text has long been a concern for owners of these assets. Protection of these assets is usually based upon the insertion of digital watermarks into the data .The watermarking software introduces small errors into the object being watermarked. These intentional errors are called marks and all the marks together constitute the watermark[1]. The marks must not have a significant impact on the usefulness of the data and they should be placed in such a way that a malicious user cannot destroy them without making the data less useful. Thus, watermarking does not prevent copying, but it deters illegal copying by providing a means for establishing the original ownership of a redistributed copy [2].

Watermarking can be done in two ways:

- Distortion free watermarking
- Non-Distortion free watermarking

In Distortion free watermarking [3] we introduce small errors into the digital assets and mostly these errors are introduced into the least significant bits of the assets so that the usefulness of the data is maintained. In Non Distortion free watermarking,[4] we somewhat extract a signature from the original asset and then we release the asset in the market, so if the asset is tampered or changed then with the help of the digital signature from the original asset it can be seen whether it has been tampered or changed.

Now-a-days, Database Watermarking has become an issue for concern because most of the information available on the internet is in the form of databases. So it is very important for their owners to claim their ownership. But, databases have very little redundancy as compared with multimedia data and this fact makes it very difficult to find enough bandwidth to embed the watermark.

The motivation for database watermarking [5] is to protect databases, especially those published online (e.g., parametric specifications, surveys, and life sciences data), from tampering and pirated copies. A watermark can be considered to be some kind of information that is embedded into underlying data for tamper detection, localization, ownership proof, and/or traitor tracing purposes. Database watermarking techniques complement the Database Protection Act and are becoming increasingly important as people realize that “the law does not now provide sufficient protection to the comprehensive and commercially and publicly useful databases that are at the heart of the information economy”.

Database watermarking consists of two basic processes: watermark insertion and watermark detection, as illustrated in Figure 1. For watermark insertion, a key is used to embed watermark information into an original database so as to produce the watermarked database for publication or distribution. Given appropriate key and watermark information, a watermark detection process can be applied to any suspicious database so as to determine whether or not a

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Issue 3, March 2016

legitimate watermark can be detected. A suspicious database can be any watermarked database or innocent database, or a mixture of them under various database attacks.

There are two types of database on the basis of the attributes they contain:

- Numerical Databases
- Non Numerical Databases

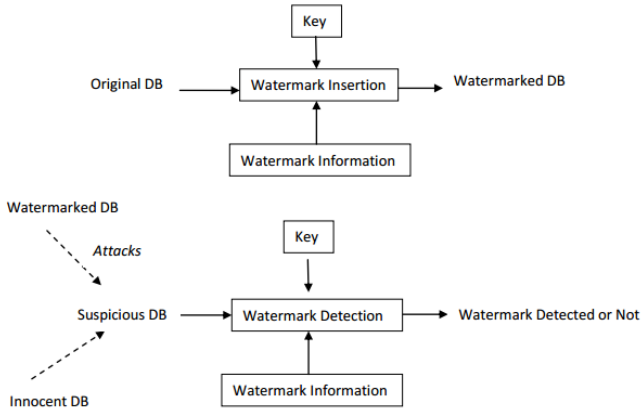

Fig 1. Basic Watermarking Process

Numerical databases contain one or more than one attributes which contain numbers. We assume that these numerical databases provide some kind of leniency in one or more than numerical fields so that we can change the attribute value to embed the watermark. Non numerical [6] databases contain only textual fields which are not liable to be changed. Watermarking in such kinds of databases can be done in two ways:

- Distortion Free Watermarking
- Loose Field Watermarking

In Distortion Free watermarking we extract an watermark from the database and then compare the watermark with the a copy of that database to infer ownership of the database. But in loose field watermarking, we search for the loose fields in the database which is provided by the database owner. In this kind of watermarking scheme, we change one or more than one loose fields to embed the watermark. By loose fields we actually mean the fields which are irrelevant in the database but which are such just present for the sake of more information. Normally we don't require such fields and these fields are assumed are never asked in a query.

The basic processes in database watermarking are quite similar to those in watermarking multimedia data , the approaches developed for multimedia watermarking cannot be directly applied to databases because of the difference in data properties. In general, database relations differ from multimedia data in significant ways and hence require a different class of information-hiding mechanisms. Unlike multimedia data whose components are highly correlated, database relations consist of independent objects or tuples. The tuples can be added, deleted, or modified frequently in either benign updates or malicious attacks. No existing watermarking techniques for multimedia data are designed to accommodate such tuple operations.

II. DIGITAL WATERMARKING

Digital watermarking is the process of embedding information [7] into a digital signal in a way that is difficult to remove. The signal may be audio, pictures or video, for example. If the signal is copied, then the information is also carried in the copy. A signal may carry several different watermarks at the same time. In visible watermarking, the information is visible in the picture or video. Typically, the information is text or a logo which identifies the owner of the media. The image on the right has a visible watermark. When a television broadcaster adds its logo to the corner of transmitted video, this is also a visible watermark.[8]

In invisible watermarking, information is added as digital data to audio, picture or video, but it cannot be perceived as such (although it may be possible to detect that some amount of information is hidden). The watermark may be intended for widespread use and is thus made easy to retrieve or it may be a form of Steganography, where a party communicates a secret message embedded in the digital signal. In either case, as in visible watermarking, the objective is to attach ownership or other descriptive information to the signal in a way that is difficult to remove. It is also possible to use hidden embedded information as a means of covert communication between individuals.

One application of watermarking is in copyright protection systems, which are intended to prevent or deter unauthorized copying of digital media. In this use a copy device retrieves the watermark from the signal before making a copy; the device makes a decision to copy or not depending on the contents of the watermark. Another application is in source tracing. A watermark is embedded into a digital signal at each point of distribution. If a copy of the work is found later, then the watermark can be retrieved from the copy and the source of the distribution is known.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Issue 3, March 2016

This technique has been reportedly used to detect the source of illegally copied movies. Annotation of digital photographs with descriptive information is another application of invisible watermarking.

Digital watermarking techniques can be classified in several ways:

- Robustness:
  - A watermark is called fragile if it fails to be detected after the slightest modification. Fragile watermarks are commonly used for tamper detection (integrity proof). Modifications to an original work that are clearly noticeable are commonly not referred to as watermarks, but as generalized barcodes.
  - A watermark is called semi-fragile if it resists benign transformations but fails detection after malignant transformations. Semi-fragile watermarks are commonly used to detect malignant transformations.
  - A watermark is called robust if it resists a designated class of transformations. Robust watermarks may be used in copy protection applications to carry copy and access control information.
- Perceptibility:
  - A watermark is called imperceptible if the original cover signal and the marked signal are (close to) perceptually indistinguishable.
  - A watermark is called perceptible if its presence in the marked signal is noticeable, but non-intrusive.

III. WATERMARKING OF RELATIONAL DATABASE

Watermarking in relational frameworks is a very young technology that has just begun its maturity cycle towards full deployment in industry-level applications. A watermark can be regarded as some sort of data that is embedded into elemental data for tamper detection, localization, ownership proof, and/or traitor tracing reasons. Database watermarking techniques complement the database protection. The watermarking techniques used for verifying integrity and tamper detection are fragile. Watermark is embedded in these techniques with the intention that it gets collapsed easily under modifications so that any tampering or modifications occur intentionally or maliciously by the attackers could be detected.

Reversible watermarking controls the distortions introduced in the data due to watermark embedding and ensure data recovery along with ownership protection. The benefits and applications of reversible watermarking are:

- **Tamper detection for data integrity**
  - Reversible watermarking technique is used for tamper detection by detecting malicious modifications in the original database relations. Such applications require tamper detection where the intention of the data owner is to preserve the data integrity of the original data.
- **Ownership protection**
  - Reversible watermarking techniques] used to provide data integrity as well as ownership protection without distortions. These techniques can be robust as well as fragile to avoid any alterations in the original data along with preserving digital rights.
- **Original Data Recovery**
  - Reversible watermarking techniques provide high quality datasets (without any alterations) for data mining and other information extraction processes.

Reversible watermarking ensures ownership protection as well as data recovery. Some of the important requirements of reversible watermarking techniques for relational databases are as follows:

- Data Recovery: Watermark embedding and extraction techniques should take care of data recovery.
- Minimizing Data Distortion: Watermark should be embedded in the data in such a way that the modifications are negligible.
- Blind: Watermark should be detected without knowing the original un-marked data and the embedded watermark. (With the passage of time, the original versions of the distributed copies of databases are kept on growing; so, it is difficult to keep the original copy of the data for watermark detection.)
- Imperceptible: The embedded watermark should be imperceptible.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Issue 3, March 2016

- Robust: Watermark should be embedded in data in such an intelligent way that the technique becomes robust against malicious attacks.
- Watermark Capacity: The watermarking technique should utilize the watermark capacity (the available bandwidth for watermark embedding without compromising the data quality).
- Security of Secret Parameters: The security of a watermarking technique mainly depends on secret parameters (e.g. secret key) that should only be private to the data owner.
- Incremental Updatability: The watermark embedding in each tuple should be independent of the other tuples.
- Randomness: The data should be watermarked with the involvement of the information of the whole relational data; so that, the proposed technique could resist malicious attacks.

IV. Related Work

The first irreversible watermarking technique for relational databases was proposed by Agrawal and Kiernan in, "Watermarking relational databases," [5].In this the method the tuples and the attributes possess a private key known only to the owner of the data .The bit pattern is used as the watermark. Reversible watermarking scheme for relational databases was proposed in "Reversible watermarking for relational database authentication," In this technique, histogram expansion is used for reversible watermarking of relational database. Difference expansion watermarking techniques (DEW),[9] by G. Gupta and J. Pieprzyk , in their work "Reversible and blind database watermarking using difference expansion," exploit methods of arithmetic operations on numeric features and perform transformations. The watermark information is normally embedded in the LSB of features of relational databases to minimize distortions.

Another reversible watermarking technique proposed in [10] is based on difference expansion and support vector regression (SVR) prediction to protect the database from being tampered. The intention behind the design of these techniques is to provide ownership proof. Such techniques are vulnerable to modification attacks as any change in the expanded value will fail to detect watermark information and the original data.

Prediction-error expansion watermarking techniques (PEEW) like [20] incorporate a predictor as opposed to a difference operator to select candidate pixels or features for embedding of watermark information. The PEEW proposed technique by Farfoura and Horng is fragile against malicious attacks as the watermark information is embedded in the fractional part of numeric features only.

"Genetic algorithm based on difference expansion watermarking (GADEW) technique "proposed by K. Jawed and A. Khan is used in a proposed robust and reversible solution for relational databases [27]. Selection of tuple and attribute is performed by hashing function, known as message authentication code (MAC).GADEW improves upon by minimizing distortions in the data, and increasing watermark capacity.

M.Kamaran and Fahim Arif proposed [13] "watermarking technique for relational databases with emphasis on re-watermarking attack". It uses bi-level security principle, thus provides robustness against re-watermarking attack. Bi-folded security scheme to detect and resolve conflicting ownership issues in case of re-watermarking attack (also called additive or secondary attacks).

Ali and Bashar Saadoon Mahdi [15] provides the "effective watermarking technique to protect valuable numeric relational data" from illegal duplications and redistributions as well as to claim ownership.The proposed system uses a new hybrid techniques, first technique MAC (Message Authentication Code) that used one way hash function SHA1, second technique is threshold generator base on simple combination of odd number of register and by using secret key in proposed system. Also Sion et al. introduce a" watermark technique for numerical data". This technique is dependent on a secret key, instead of primary key uses the most significant bits of the normalized data set, divides the data set into partitions using markers, and varies the partition statistics to hide watermark bits.

Hazem M. El-Bakry [16] proposed technique is available for any relational database. No delay and no additional time required till the normal calculation end. R.Balasubramaniam [17] present watermarking embeds ownership information in digital content. Watermarking of relational databases as a constrained optimization problem and discus efficient techniques to solve the optimization problem and to handle the constraints.

V. A Framework For Reversible Watermarking Of Relational Databases

Reversible watermarking techniques for relational databases can be generalized into four phases:

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Issue 3, March 2016

preprocessing, encoding, decoding and data recovery. The benefit of providing this framework are: it covers the entire procedure for reversible watermarking of relational databases into four phases; and it can also be used as a reference for developing reversible watermarking techniques for numeric and non-numeric relational databases.[19]

- **Data Preprocessing**
  During data preprocessing step, two sub-modules perform different tasks:selection of a suitable feature for watermark embedding; and formation of a watermark. Statistical measures are used to rank the database features (features in relational databases) according to their importance in the information extraction process. Ranking of features are measured on the basis of their (1) Mutual Information (MI); (2) Information Gain (IG); and (3) Entropy measures.
  The possible ways of watermark generation for relational data reversible watermarking are Evolutionary techniques, such as Particle Swarm Optimization (PSO), Particle Search (PS) and Genetic Algorithm (GA), are employed. Among them GA is a population-based computational model basically inspired from genetic evolution. Watermark Generation through Pseudo-Random Sequence Generator. Watermark Generation through Hashing Techniques, Hashing algorithms such as MD5 and SHA-1 are used to create a watermark string. In some techniques, modulus of hash value of primary key features is calculated and used as watermark information.

- **Watermark Encoding**
  During watermark encoding phase, watermark information is embedded in the selected feature(s). A
  Number of parameters are also computed in encoding phase for use in watermark encoding and
  Decoding phase.

- **Watermark Decoding**
  In watermark decoding phase, the embedded watermark is decoded from the suspicious data. The preprocessing step is performed and decoding strategies are used for recovering watermark. In verifying the watermark information, the original watermark and the detected watermark are compared. Watermark detected should be same as the watermark inserted into the data to prove ownership rights.

- **Data Recovery**
  Original data is recovered in data recovery phase through post processing steps for error correction and recovery. After recovery, the recovered data might also be compared with the original data to ensure that the data quality is not compromised

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

  - **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.

  - **Deletion attack :**
    The Attacker deletes marked tuples from the relational database which leads to synchronization errors.

  - **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.

  - **Insertion attack :**
    Attacker inserts tuples to the data set hoping to disturb the embedded watermark which results in synchronization errors.

VI. CONCLUSION

In this paper we discussed about a genric frame work for reversible watermarking technique on relational databases. Reversible watermarking techniques are used because they are able to recover the original data from watermarked data and ensure data quality. This paper

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Issue 3, March 2016

includes basic concept about digital watermarking then its application in database or say database watermarking .A literature review on different watermarking techniques for database watermarking has been discussed.

REFERENCES

[1]. I.Cox, M. Miller, J. Bloom, and M. Miller, Digital watermarking.Morgan Kaufmann, 2001.

[2]. J. T. Brassil, S. Low, and N. F. Maxemchuk, "Copyright protection for the electronic distribution of text documents," Proceedings of the IEEE, vol. 87, no. 7, pp. 1181–1196, 1999.

[3]. 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.

[4]. A. Hamadou, X. Sun, L. Gao, and S. A. Shah, "A fragile zero-watermarking technique for authentication of relational databases,"International Journal of Digital Content Technology and its Applications, vol. 5, no. 5, pp. 189–200, 2011.

[5]. R. Agrawal and J. Kiernan, "Watermarking relational databases," in Proceedings of the 28th international conference on Very Large Data Bases. VLDB Endowment, 2002, pp. 155–166.

[6]. R. Bedi, A. Thengade, and M. Vijay, "A new watermarking approach for non-numeric relational database," International Journal of Computer Applications, vol. 13, no. 7, pp. 37–40, 2011.

[7]. R. Agrawal, P. J. Haas, and J. Kiernan, "Watermarking relational data: framework, algorithms and analysis," The VLDB journal,vol. 12, no. 2, pp. 157–169, 2003.

[8]. M. U. Celik, G. Sharma, and A. M. Tekalp, "Lossless watermarking for image authentication: a new framework and an implementation," Image Processing, IEEE Transactions on, vol. 15,no. 4, pp. 1042–1049, 2006.

[9]. A. M. Alattar, "Reversible watermark using difference expansion of triplets," in Proc. IEEE Int. Conf. Image Process., 2003, pp. I–501,vol. 1.

[10]. J.-N. Chang and H.-C. Wu, "Reversible fragile database watermarking technology using difference expansion based on SVR prediction," in Proc. IEEE Int. Symp. Comput., Consum. Control,2012, pp. 690–693.

[11]. R. Sion, M. Atallah, and S.Prabhkar, "Right Protection for Relational Data." IEEE Trans. Knowledge and Data Engineering, Vol. 16 no.6, June 2004

[12]. W. Ng and H. Lau, "Effective Approaches for Watermarking XML Data." Department of Computer Science, the Hong Kong University of Science and Technology, Hong Kong, 2005.

[13]. Sabah Suhail1, M. Kamran and Fahim Arif, "Watermarking of Relational Databases with Emphasis on Re-watermarking Attack", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 1, May 2012.

[14]. Vahab Pournaghshband , "A New Watermarking Approach for Relational Data" , 9th WSEAS International Conference on Applied Informatics And Communications (AIC '09) Vol.2, No.3, August 2010

[15]. Dr. Yossra H. Ali & Bashar Saadoon Mahdi , "Watermarking for Relational Database by using Threshold Generator", Computer Sciences Department, University of Technology Eng. & Tech. Journal, Vol. 29, No. 1, 2011.

[16]. Hazem M. El-Bakry, "A New Watermark Approach for Protection of Databases", Mansoura University, IEEE Transactions on Knowledge and Data Engineering, vol. 20, no. 1, pp. 116–129, 2007.

[17]. R.Balasubramaniam, "Data Security in Relational Database Management System", International Journal of Computer Science and Security (IJCSS), Volume (6): Issue (4): 2012.

[18]. S.W. Weng, Y. Zhao and J.-S. Pan "Reversible watermarking resistant to cropping attack", The Institution of Engineering and Technology 2007 (IET Inf. Secure), 2007, 1, (2), pp. 91–95.

[19]. S. Iftikhar, M. Kamran, and Z. Anwar, "Rrw-a robust and reversible watermarking technique for relational data," 2014.