*International Journal of Application or Innovation in Engineering & Management (IJAIEM)*
**Web Site: www.ijaiem.org Email: editor@ijaiem.org**
**Volume 3, Issue 10, October 2014** **ISSN 2319 - 4847**

# A Study of Watermarking Relational Databases

**Miss. Snehal S.Kshatriya[1], Prof.Dr.S.S.Sane[2]**

[1] K. K. Wagh Institute of Engineering Education & Research, Department of Computer Engineering,
Savitribai Phule Pune University
Nashik,India

[2] K. K. Wagh Institute of Engineering Education & Research, Department of Computer Engineering,
Savitribai Phule Pune University
Nashik,India

## ABSTRACT

***Databases most often contain critical information. In today's internet-based application environment, ownership rights protection on relational database is decisive issue because unauthorized changes to data may have serious consequences and result in significant losses for the organization. Ownership protection on relational databases that is shared with purposive receiver desires to develop a watermarking technique that must be robust against different types of attacks and it should retain the knowledge in the databases in order to make them effective for knowledge-aware decision support systems. It is desirable that a database owner need not define usability constraints for every application and for every recipient distinctly. Most of the study available in this field is focused on images, audio, video etc. However, with the requirement of relational database security solution, this paper presents the various watermarking techniques on relational data with certain constraints and analyze their strengths and weaknesses.***

**Keywords:-** data usability constraints, Knowledge preservation, ownership protection, watermarking.

## 1. INTRODUCTION

**S**ecurity of relational database is great concern in today's world because of sharing of data over internet. Data providers create services and make them available to users for searching and accessing purposes. Given that these services may attract more attacks. So it is desire to protect the data and hence providers need some technology that identify the threats and pirated copies unauthorized access to their databases . The increasing use of relational database create a need for watermarking database. . In today's internet-based application environment, ownership rights protection on relational database is decisive issue because unauthorized changes to data may have serious consequences and result in significant losses for the organization. Hence right protection through watermarking becomes an important research topic. The purpose of Digital Watermarking is to protect a data from unauthorized duplication and distribution by enabling provable ownership over the data. These schemes allow the owner of the data to insert a slight watermark into the data. The ownership of data can be proved by a watermark. Secure watermark embedding requires that the watermark must not be easily forged, tampered with or removed from the watermarked data [16]. Slightly inserting watermarking means that the watermark presence is unnoticeable in the data. The blind watermark detection means that it doesn't requires the knowledge of original data and the watermark itself. Watermarking techniques have been developed for multimedia contents such as images, audio, video and text data. It is also used for software and natural language text. But watermarking used for multimedia is different from that of the used for relational data because watermarking relational database has unique and complex requirement. Therefore, relational database watermarking scheme is challenged for frequent updating, relational data out-of-order and data redundancy fewness. Due to such unique requirements and challenges, Watermarking relational databases literature is very limited and it focuses mainly on inserting binary bits in randomly chosen places in databases.

## 2. WATERMARKING RELATIONAL DATABASE

Figure 1 depicts the basic database watermarking technique[10]. Watermark embedding phase includes a private key K (known only to the owner) which is used to embed the watermark bits into the original database to form watermarked database. The watermarked database is then made publicly available. To verify the right ownership of a doubtful database, the verification process is performed. In this process the mistrustful database is taken as input and by using the private key K which is used during the embedding phase, the embedded watermark (if present) is extracted from watermarked database and it is compared with the original watermark information.

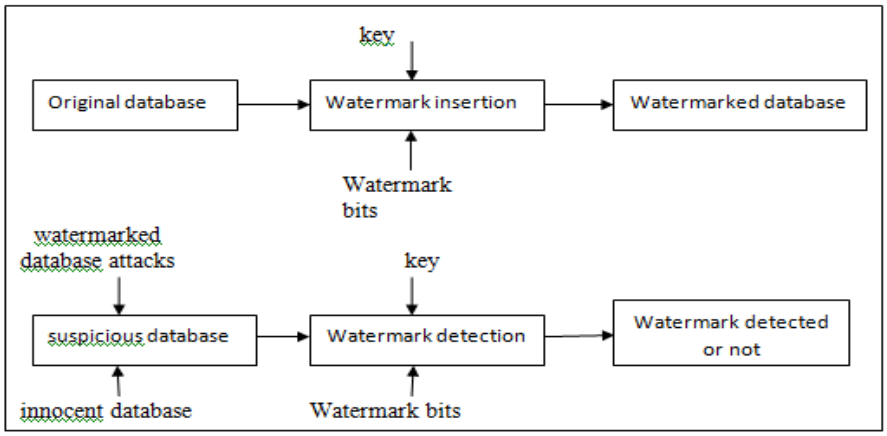


**Figure 1** Basic Watermarking Technique

The watermarked database must preserve the following properties:

**Robustness:** Watermarking process should be robust against different types of malicious attacks. The watermarking algorithm should be developed in such a way that it should be difficult for an attacker to delet or alter the watermark from database without violating the knowledge of the data.

**Usability:** Watermarking technique should not results in distortion of data and knowledge in the databases should be preserved. i.e. Data should be useful after watermark embedding process.

**Blindness:** Watermark extraction should not require the knowledge of the original database and watermark itself.

**Security:** Watermarked tuples, attributes, bit positions that are selected for embedding watermark bits should be kept secret and it should be only known by having the knowledge of a secret-key. (i.e.Owner of the database)

**2.1 Application of Digital Watermarking for Relational Databases**

Digital Watermarks for relational databases are useful in many applications:

1) **Ownership Assurance:** For ownership protection watermarking can be used. To assure ownership of a relational database, Owner of the database can embed a watermark into his data by using some private parameters which is known only to him. Then watermarked database can be made publicly available. Later, suppose Owner suspects that the data published by someone else has pirated from his data. To avoid ownership confusion, Owner can proved the presence of his watermark in attacker's data. Hence watermark detection have to be used to survive against various malicious intentions [4].
2) **Fingerprinting:** Fingerprinting is used to identify a betrayer. The applications where content is publicly available over a network, the owner of data would like to discourage unauthorized distribution and dublication of data by embedding a distinct watermark in each copy of the content. If an unauthorized copies of the data are found, then the original data can be determined by extracting the fingerprint [4].
3) **Fraud and Tamper Detection:** Critical applications such as commercial transactions or medical applications use data, it will originated from a specific source and it will not been modified, manipulated or destroyed. This can be achieved by embedding a watermark in the underlying data of the database. The watermark is extracted by using private parameter associated with the source. Fraud in the data is verified by checking the integrity of original data to that of extracted watermark [7]

**2.2 Different Attacks**

In fragile watermarking, integrity verification is done while in robust watermarking, the embedded watermark should be robust against various types of attacks. This attacks includes removing or distorting the watermark. The watermarked database may suffer from various types of attacks which are created intentionally and unintentionally and it may damage or erase the watermark.[13]

1) **Benign Update:** In this type of attack, the marked tuples may be inserted, removed or updated. It may make embedded watermark detectable or undetectable. This may done unintentionally.
2) **Value Modification Attack:** In this type of attack, watermarks are destroyed by altering one or more bits in the watermarked data. 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.
3) **Subset Attack:** Attacker may consider a subset of the tuples or attributes of a watermarked relation. Attacker may delete or update tuples or attribute and hope for watermark has been lost.
4) **Collusion Attack:** This attack requires the attacker to have access to multiple watermarked copies of the same content.

5) **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, attacker may claim for ownership by adding his own watermark in owner's data.
7) **Subset Reverse Order Attack:** In this type of attack, attacker exchangs the order or positions of the tuples or attributes in data which may remove or disturb the watermark.

**2.3 Classification of Watermarking Techniques**

In this paper, we try to cover the details of various watermarking techniques. To limit the survey area we classify techniques based on:

1)**Watermark Information:** Different watermarking embed different types of watermark information into the database. (e.g. image, text, sound etc.)
2) **Distortion:** Watermarking may be distortion-based or distortion- free depending on whether the marking introduces any distortion to the data. Distortion-based watermarking techniques includes slight changes in the original data during embedding phase but the degree of change should be tolerable and should not make the data useless. In distortion-free watermarking scheme, the watermark insertion phase does not depend on any specific type of attribute and does not introduce any distortion in the original data
3) **Cover Type:** Watermarking can be classified based on the type of the cover i.e. type of attributes into which watermark bits are embedded.
4) **Granularity Level:** The watermarking can be performed by modifying or inserting information at bit level or higher level (e.g. character level or attribute level or tuple level).
5) **Verifiability:** The verification process may be deterministic or probabilistic in nature, it can be performed blindly or non-blindly, it can be performed publicly (by anyone) or privately (by the owner only).
6) **Intent:** Different watermarking schemes are designed for various purposes, namely, integrity and tamper detection, localization, ownership proof, traitor detection etc.

## 3. RELATED WORK

In this section, an overview of existing watermarking techniques for relational database are provided. The objective of this survey is clearly understand the limitations of existing schemes. Agrawal and Kiernan [3] uses bit-level watermarking algorithm. In this algorithms, attributes of a chosen subset of tuples are selected to embed watermark bit information. The value of a hash function determines which attribute should be select for watermark embedding process. Some bit locations will be marked amongst least significant bits of the attribute[1], [3], [6]. Agrawal and Kiernan [3] proposed a bit-resetting algorithm which set the LSB of the selected attribute of the selected subset of tuples. The attribute selection for watermark embedding is based on computation of message authenticated code, where MAC is calculated by using the tuple's primary key ans the secret key. This technique assumes unconstrained LSB manipulation during watermark embedding process. Such out-of-bound modification of data generates undesirable results. LSB-based data hiding techniques are efficient, but an attacker can easily destroy watermark by simple manipulation of data. (for example shifting LSB) Gupta, G. and Pieprzyk, J [5] propose a reversible watermarking scheme that is the modified version of Agrawal and Kiernan's. In this technique, it first extracts a bit OldBit from the integer portion of the attribute value and then replace it by the watermark bit and embeds it in the fraction portion of the attribute value. Thus, the watermark bit can be recovered and the attribute can be restored to its unmarked value by replacing the watermark bit with the original bit OldBit extracted from the fraction part. Thus the original database can be recovered along with the ownership proof. R. Sion, M. Atallah, and S. Prabhakar [8] uses Statistical-property watermarking algorithm. In this algorithm, watermark bits are embedded in actual data distribution properties of subset of tuples. The whole database is divided into a maximum number of unique, non-overlapping subsets of tuples. A watermark bit is embedded in each selected subset of tuples by making slight difference in some of the data values. Then average value of subset and variance values are reach some value which is depending on whether the watermark bit is 0 or 1. The data partitioning concept is vulnerable to watermark synchronization errors, particularly in the case of tuple insertion and deletion attacks, because the position of marker tuples is disturbed by these attacks. Such errors may be reduced if marker tuples are stored during watermark embedding phase and that same stored marker tuple will used for the data partitions again during watermark decoding phase. But this violates the concept of "blind decoding" of watermark. Furthermore, the threshold selection for bit decoding includes arbitrarily chosen thresholds without following any optimality criteria. This will results in error in the decoding process. The concept of usability bounds on data is used in this technique to control distortions introduced in the data during watermark embedding. However, an attacker can corrupt the watermark by launching large scale attacks on large number of rows. The decoding accuracy is depend on the usability constraints defined by owner of the data. Hence decoding accuracy is degrade if an attacker violates these bounds. An important shortcoming of this approach is that the data owner needs to specify usability

constraints separately for application that will use data every time. Zhang, Z, X. Jin, M. Wu, E. Tang [17], [18] uses Image-based watermarking algorithm. In this algorithms, Watermarking is done by using an image. The image contain bits which are used to represent the watermark bits. These bits are embedded in selected locations in database and if these bits are recovered correctly then it can be used to reconstruct an embedded image. This watermarking technique can be considered as a sub-class of the bit-level watermarking algorithm. A watermarking scheme proposed by Al-Haj, A. and Odeh, A.[9] is based on hiding binary image in non-numeric multi-word attributes spaces of subsets of tuples. The watermark is partitioned in m string each containing n bits. The database is also partitioned into non-overlapping subset each containing m tuples. The watermark image is embedded into each m-tuple subset. The embedding process is done as follows: suppose the integer representation of the ith, i € [1 ...m], short string is di .A double space is created after di words of the pre-selected nonnumeric, multi-word attribute of ith tuple in the subset. The extraction process counts the single spaces present before double space which shows inserted short binary string. This algorithm embeds the same watermark for all non-intersecting subsets of the database and so it is robust against subset addition, subset deletion attacks. Another advantage for space- based watermarking is that large bit-capacity available for embedding the watermark. However, it may suffer from watermark removal attack if attacker replaces all double spaces between two words (if exist) by single space for all tuples in the relation. M. Kamran, Sabah Suhail, and Muddassar Farooq [2] uses Random bit pattern watermarking algorithm. They proposed a method for numeric data. It is a robust against various attacks and efficient watermarking scheme for relational databases. In this method, a robust watermark algorithm is used to embed watermark bits into the original data set. The watermark embedding algorithm takes a secret key and the watermark bits as input and converts a original data set into watermarked data set. Watermark bits are generated from UTC (Coordinated Universal Time) datetime which is the primary time standard used to synchronize the time all over the world [19]. These bits are given as input to the watermark encoding function. Then data set is partitioned into non-overlapping partitions by using the secret key in conjunction with a cryptographic secure hash function. To minimize distortions, only few tuples are selected for watermarking. Then watermark bits are embedded in the selected tuples using a robust watermarking function. This technique embeds each bit of the watermark in every selected tuple of each partition; as a result, it is robust against malicious attacks even only one watermarked row is left in the data after an attack. The watermarked data set is delivered to recipient where an attacker aims at destroying the watermark by launching different types of attacks. The decoding algorithm is blind and its decoding accuracy does not depend on the usability constraints. As a result, 100 percent decoding accuracy is achieved irrespective of the amount of data alterations made by an attacker in the watermarked data. But The decoding accuracy may decrease in case of combination of different attacks. And if an attacker alters the original data to some signed or zero-valued data then the decoding accuracy might be decreased. This technique is best suited for data sets that contain unsigned numeric attributes. This scheme depend critically on presence of primary key attribute. If there is no primary key or if attacker alter/destroy key then scheme will not work.

## 4.CONCLUSION AND FUTURE WORK

In this paper, we reviewed different techniques on watermarking relational databases that embeds the watermark bits in the database set by partitioning it. Every author worked for the robustness of the technique. Many watermarking techniques are based on different watermark information, most of these techniques are designed for numerical database and are distortion based. There are almost similar steps to identify attribute then tuple and then marking position for the watermark. Finally, we observe that usability of the watermarked database and deterministic detectability leaves so many queries in mind for future research. Most of these techniques used a single attribute of a tuple to embed a watermark. So, this work will be extended towards embedding the same watermark at different attributes at different places. Therefore, it will be difficult for attacker to remove watermarks from different places from the database. Most of these techniques are also depend on presence of primary key. So we will also extend the work to find solution if there is no primary key.

## REFERENCES

[1] Agrawal, R., P. Hass and J. Kiernan, 2003.Watermarking relational data: Framework, algorithms and analysis. Int. J.Very Large Data Bases, 12: 157-169.
[2] 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, DECEMBER 2013
[3] Agrawal, R. and J. Kiernan, 2002. Watermarking relational databases. Proceeding of the 28th International Conference on Very Large Databases, (ICVLD’02), Hong Kong, China, pp: 1-12.
[4] http://en.wikipedia.org/wiki/Digital_watermarking

[5] Gupta, G. and Pieprzyk, J. "Database relation watermarking resilient against secondary watermarking attacks" In Proceedings of the 5th International Conference on Information Systems Security (ICISS 09), 2009 pages 222–236, Kolkata, India. Springer LNCS, Volume 5905.
[6] Gross-Amblard, D., 2003. Query preserving watermarking of relational databases and XML documents. Proceedings of the 22nd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems**,** June 9-11, New York, USA., pp: 191-201. DOI: 10.1145/773153.773172
[7] G.H. Gamal, M.Z. Rashad and M.A. Mohamed "A Simple Watermark Technique for Relational Database" Mansoura Journal for Computer Science and Information Systems Vol. 4, No.4, Jan2008.
[8] R. Sion, M. Atallah, and S. Prabhakar, "Rights Protection for Relational Data," IEEE Trans. Knowledge and Data Eng., Vol. 16, no. 6, pp. 1509-1525, Dec. 2004.
[9] Al-Haj, A. and Odeh, A. (2008). "Robust and blind watermarking of relational database systems" Journal of Computer Science, 2008 4:1024–1029.
[10][10] Bhattacharya and Cortesi] - "Database authentication by distortion-free watermarking." In Proceedings of the 5th International Conference on Software and Data Technologies (ICSOFT '10), 2010 pages 219–226.
[11] A. Deshpande and J. Gadge, "New Watermarking Technique for Relational Databases," Proc. Second Int'l Conf. Emerging Trends in Eng. and Technology (ICETET), pp. 664-669, 2009.
[12] M. Farfoura and S. Horng, "A Novel Blind Reversible Method for Watermarking Relational Databases," Proc. IEEE Int'l Symp. Parallel and Distributed Processing with Applications, pp. 563-569, 2010.
[13] 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, pp.3165-3190
[14] D. Allan, N. Ashby, C. Hodge, and H.-P. Company, The Science of Timekeeping. Hewlett-Packard, 1997.
[15] B. Schneier, Applied Cryptography. John Wiley, 1996
[16] R. Wolfgang, C. Podilchuk, and E. Delp. Perceptual Watermarks for Digital Images and Video. Proceedings of the IEEE 87:1108.1126, July 1999
[17] Wu, M., E. Tang and B. Liu, 2000. Data hiding in digital binary image. Proceeding of the IEEE International Conference on Multimedia and Expo, July 30-Aug. 02, New York, USA., pp: 393-396. DOI: 10.1109/ICME.2000.869623
[18] Zhang, Z., X. Jin, J. Wang and D. Li, 2004. Watermarking relational database using image. Proceeding of the International Conference on Machine Learning and Cybernetics, Aug. 26-29, IEEE Xplore Press, USA., pp: 1739-1744. DOI: 10.1109/ICMLC.2004.1382056
[19] S. Horng et al., "A Blind Reversible Method for Watermarking Relational Databases Based on a Time-Stamping Protocol," Expert Systems with Applications, vol. 39, no. 3, pp. 3185-3196, 2011
[20] Y. Li and R. Deng, "Publicly Verifiable Ownership Protection for Relational Databases," Proc. ACM Symp. Information, Computer and Comm. Security, pp. 78-89, 2006.
[21] S. Bhattacharya and A. Cortesi, "A Distortion Free Watermark Framework for Relational Databases," Proc. Fourth Int'l Conf. Software and Data Technologies (ICSOFT '09), pp. 229-234, 2009.
[22] R. Halder, S. Pal, and A. Cortesi, "Watermarking Techniques for Relational Databases: Survey, Classification and Comparison," J. Universal Computer Science, vol. 16, no. 21, pp. 3164-3190, 2010
[23] Lafaye, J. "An analysis of database watermarking security" In Proceedings of the 3rd International Symposium on Information Assurance and Security (IAS '07),2007 pages 462–467, Manchester, United Kingdom. IEEE Computer Society

## AUTHOR

**Snehal Kshatriya** received the B.E. degrees in Information Technology from K.B.Thakare College of Engineering, Nashik, Savitribai Phule Pune University in 2012. Now pursuing M.E. in Computer Engineering from K. K. Wagh Institute of Engineering Education & Research, Nashik, India.

**Prof. Dr.S.S.Sane** received M. Tech (CSE) from IITB, Ph D from COEP, University of Pune. Currently working as Vice Principal, Professor & Head of Dept. of Computer Engineering, Prof. In-charge Central Library in K K Wagh Institute of Engineering Education & Research, Nashik, India