International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 3, ISSN No. 2455-2143, Pages 89-93
Published Online December – January. 2016 in IJEAST (http://www.ijeast.com)

DATABASE WATERMARKING TECHNIQUES

Hira Khan GLBTIM
Parul Srivastava GLBTIM
Ritu Dagar GLBTIM
Suhasini GLBTIM

ABSTRACT - Digital watermarking for relational databases emerged as a candidate solution to provide copyright protection, tamper detection, traitor tracing, and maintaining integrity of relational data. Until last decade, most of the work in watermarking was done on the image, video etc., but with a rapid increase in the use of relational databases, the database watermarking has become a great topic of interest for preventing the piracy and asserting ownership on outsourced databases. The paper has been aimed towards studying the various watermarking techniques for relational databases. This is mainly focused on the review of various relational database watermarking techniques and their security analysis.

Keywords - Database Security, Database Watermarking, HMAC, Copyright Protection, Cryptography

I. INTRODUCTION

The recent surge in the growth of the Internet results in offering of a wide range of web-based services, such as database as a service, digital repositories and libraries, e-commerce, online decision support system etc. These applications make the digital assets, such as digital images, video, audio, database content etc. easily accessible by ordinary people around the world for sharing, purchasing, distributing, or many other purposes. As a result of this, such digital products are facing serious challenges like piracy, illegal redistribution, ownership claiming, forgery, theft etc. Digital watermarking technology is an effective solution to meet such challenges. A watermark is considered to be some kind of information that is embedded into underlying data for tamper detection, localization, ownership proof, traitor tracing etc. The piracy of digital assets such as software, images, video, audio and text has long been a concern for the owners of these assets. [1] Data owners use various watermarking techniques to prevent infringement of their copyrighted work. Watermarking is essentially the insertion of a watermark (which is typically the introduction of small errors) in the digital document which doesn't affect the quality and usefulness of the document in a significant manner.

Watermarking techniques are used on various digital media but here we're concentrating on relational databases. There are two phases in watermarking a database:

1. Watermark Insertion: During the insertion phase, a watermark is embedded into the original database using the private key (K). The watermarked database is then outsourced publicly.

2. Watermark Verification: When we need to verify the ownership of a suspicious database, we try to extract the watermark using the same private key (K) used for insertion and then compares the extracted watermark with the original watermark. Based on the similarity between extracted and originally inserted watermark we determine the chances of the ownership claim to be real.

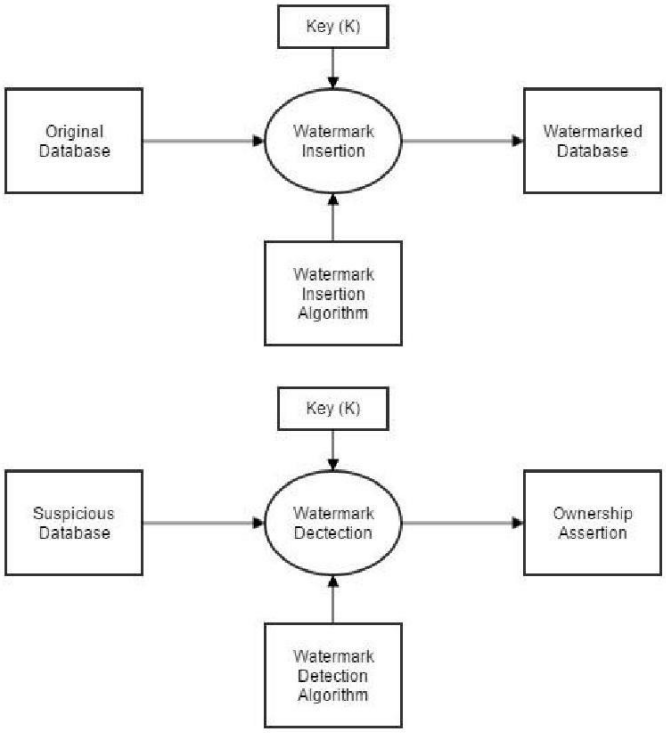

Figure 1: Database watermarking in a nutshell

We can't directly use the techniques which were

originally developed for multimedia data due to differences between relational databases and multimedia. Some of the major differences in database watermarking as compared to multimedia are [5][6][7]:

1. Few Redundant Data: A multimedia object consists of a large no. of bits hence; there is a large amount of space available to hide the watermark. While database consist of the tuples, each tuple represents a separate object so, the watermark is spread over these separate objects.

2. Out of Order Relational Data: The change in a relative spatial/temporal positioning of multimedia object remains unchanged, while in case of database, updates in database may changes the tuples.

3. Frequent Updating: Drop or Replace operation is not possible in multimedia object without causing perceptual changes in the object. While, tuples may simply be dropped by delete operation in database.

4. Human Phenomenon: There are many psycho-physical phenomena based on human visual system and human auditory system which can be exploited for mark embedding. However, one cannot exploit such phenomena in case of relational databases.

Due to these differences, no audio/video/digital media watermarking scheme is applicable for relational databases. There are many more technical challenges in database watermarking due to these differences.

II. LITERATURE SURVEY

The security of relational databases has been a great concern since the expanded use of these data over the Internet.

2.1 Applications of Database Watermarking

The various applications of inserting a digital watermark in relational databases are –

1. Ownership Assertion: Alice (original author of database) can insert a watermark in her original database (R) using some secret key (K) and then outsource the watermarked database to public. When Alice encounters a suspicious database which she thinks is pirated from her own database then she can extract the watermark from suspicious database and can claim her ownership in the suspicious relation. Watermark must survive intentional or unintentional database update operations which may distort or completely remove the watermark.

2. Fingerprinting: its main aim is to determine the original author of the unauthorized copy of a relation. When some original work is pirated by unauthorized duplication and distribution, the Alice would insert a watermark (fingerprint) in each copy of the database. Retrieval of the fingerprint will help in determining the original source of the database.

3. Fraud and Temper Detection: Data integrity (or data origin authentication) becomes an essential requirement for databases used in critical applications such as medical transactions, commercial applications etc. We need to ensure that the content has been originated from an authentic source and it has not been tampered thereafter. It is ensure by inserting a watermark in the underlying data of the relation. When the watermark is extracted later, integrity of the database can be verified by the integrity of the extracted watermark.

2.2 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. **Benign 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. For example, Mallory may Halder R., Pal S., Cortesi A.: Watermarking Techniques ... 3167 convert the data to a different unit of measurement (e.g., Fahrenheit to Celsius). The unnecessary conversion by Mallory would raise suspicion 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.

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

6.2. Invertibility Attack: Mallory may launch an invertibility 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.

8. **Brute Force Attack:** In this case, Mallory tries to guess about the private parameters (e.g. secret key) by traversing the possible search spaces of the parameters. This attack can be thwarted by assuming that the private parameters are long enough in size.

2.3. Desirable Properties of Watermarking Schemes

Some of the desired properties which are essential in a relational database watermarking scheme for it to serve its purpose fully are –

1. Detectability: The watermark must be detectable by the actual owner of the database when claiming the ownership of the database relation.

2. Robustness: The inserted watermark must be robust (unaffected) against the various intentional or unintentional modifications (update, delete, modify etc.) in the data. The watermark must be detectable even after the original database relation has been modified by the attacker.

3. Updatability: The watermarked database should be efficiently updatable with no damage to the previously inserted watermark.

2.4. Watermarking Issues

The important issues that arise in the study of digital watermarking techniques for relational databases are:

1. Capacity: It determines the optimum amount of data that can be embedded in a cover and the optimum way to embed and extract this information.

2. Usability: The changes in the data of the database during watermarking process should not degrade the usability of the data. The amount of allowable change differs from one database to another, depending on the nature of stored records.

3. Robustness: Watermarks embedded in databases should be robust against malicious or accidental attempts at removal without destroying the usability of the database.

4. Security: The security of the watermarking process relies on some private parameters (e.g. secret key) which should be kept completely secret. Owner of the database should be the only one who has knowledge about them.

5. Blindness: Watermark extraction should require neither the knowledge of the original unwatermarked database nor the watermark information. This property is critical as it allows the watermark to be detected in a copy of the database relation, irrespective of later updates to the original relation.

6. Incremental Watermarking: After a database has been watermarked, the watermarking algorithm should

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 3, ISSN No. 2455-2143, Pages 89-93
Published Online December – January. 2016 in IJEAST (http://www.ijeast.com)

compute the watermark values only for the added or modified tuples, keeping the unaltered watermarked tuples untouched.

7. Non-interference: If multiple marks are inserted into a single relational database, then they should not interfere with each other.

8. Public System: Following Kerckhoffs [Kerckhoffs, 1983], the watermarking system should assume that the method used for inserting a watermark is public. Defense must lie only in the choice of the private parameters (e.g. secret key).

9. False Positiveness and False Negativeness: The false hit is the probability of a valid watermark being detected from unwatermarks data, whereas false miss is the probability of not detecting a valid watermark from watermarked data that has been modified in typical attacks. The false hit and false miss should be negligible.

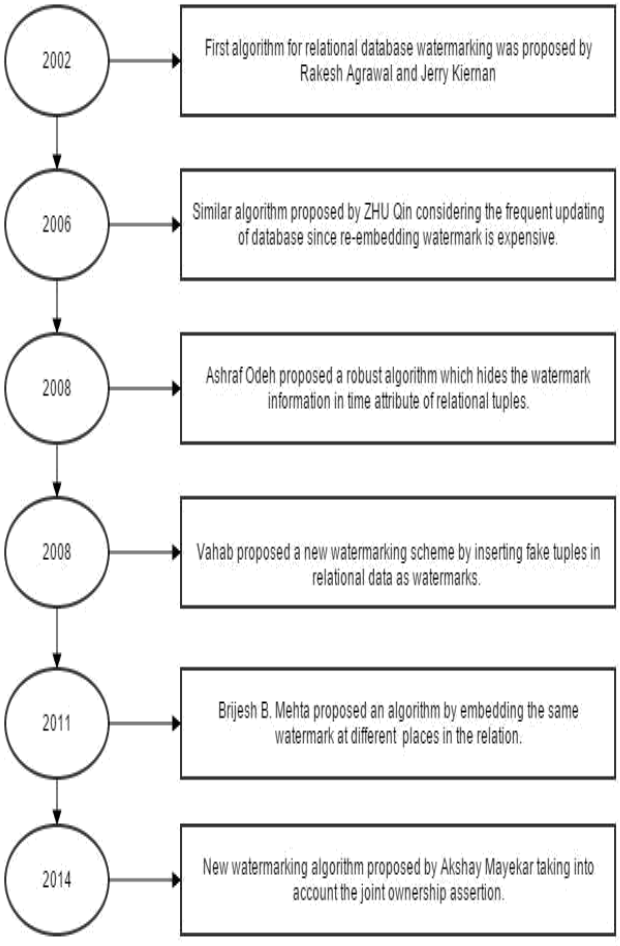

Figure 2: Various Watermarking approaches proposed for relational databases over the period of time.

2.5. Classification of Watermarking Techniques

The watermarking techniques proposed so far can be classified along various dimensions as follows:

1. **Watermark Information:** Different watermarking schemes embed different types of watermark information (e.g. image, text etc.) into the underlying data of the database.

2. **Distortion:** Watermarking schemes may be distortion-based or distortion free depending on whether the marking introduces any distortion to the underlying data.

3. **Cover Type:** Watermarking schemes can be classified based on the type of the cover (e.g. type of attributes) into which marks 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/Detectability:** The detection/ 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 of Marking:** Different watermarking schemes are designed to serve different purposes, namely, integrity and tamper detection, localization, ownership proof, traitor detection etc.

III. CONCLUSION

In this project we reviewed all the work which has been done so far in the field of relational database watermarking. We reviewed 6 papers proposed by different authors on watermarking and fingerprinting of relational databases. All the authors have focused towards the robustness of the technique. All the proposed techniques can be classified on the basis of *(i)* Whether the technique introduces errors in the existing data or adds new fake data, *(ii)* the type of cover where the watermark is embedded /citeraju *(iii)* the type of information which is watermarked. Most of the distortion-based schemes for numerical data have almost similar approach for identifying the candidate bit positions to watermark in the database.

IV. REFERENCES

1. R.Agrawal and J. Kiernan, "Watermarking relational databases," *Proceedings of the 28th international conference on Very Large Data Bases*, 2002.

2. J.LAFAYE, "An analysis of database watermarking

security," *Third Internationl Symposium on Information Assurance and Security*.

3. V. Pournaghshband, "A new watermarking approach for relational data," *Proceedings of the 46th Annual Southeast Regional Conference on XX. ACM*, 2008.

4. Z. Qin, Y. Ying, L. Jia-jin, and L. Yisho, "Watermark based copyright protection of outsourced database," *IEEE*, 2006.

5. R. Halder, S. Pal, and A. Cortesi, "Watermarking techniques for relational databases: Survey, classification and comparison," *Journal of Universal Computer Science, vol. 16*, pp. 3164–3190, 2010.

6. B. B. Mehta and U. P. Rao, "A novel approach as multi-place watermarking for security in database,"

Department of Computer Science and Engineering, S.V. National Institute of Technology, Surat, Gujarat.

7. Abdel-Hamid, A. T., Tahar, S., and Aboulhamid, E. M. (2004). A survey on ip watermarking techniques. Design Automation for Embedded Systems, 9(3):211–

8. Bhattacharya, S. and Cortesi, A. (2009b). A generic distortion free watermarking technique for relational databases. In Proceedings of the 5th International Conference on Information Systems Security (ICISS '09), pages 252–264, Kolkata, India. Springer LNCS, Volume 5905.

9. Qin, Z., Ying, Y., Jia-jin, L., and Yi-shu, L. (2006). Watermark based copyright protection of outsourced database. In Proceedings of the 10th International Database

Engineering and Applications Symposium (IDEAS'06), pages
301–308, Delhi, India. IEEE Computer Society.

10. Zhang et al., 2005 Zhang, Y., Niu, X., and Zhao, D. (2005). A method of protecting relational databases copyright with cloud watermark. International Journal of Information and Communication Engineering, 1:337–341.

11. Agrawal, R., Haas, P. J., and Kiernan, J. (2003a). A system for watermarking relational databases. In Proceedings of the 2003 ACM SIGMOD international conference on

Management of data (SIGMOD '03), pages 674–674, San Diego, California. ACM Press.

12. Agrawal, R., Haas, P. J., and Kiernan, J. (2003b). Watermarking relational data: framework, algorithms and analysis. The VLDB Journal, 12:157– 169.

13. Al-Haj, A. and Odeh, A. (2008). Robust and blind watermarking of relational database systems. Journal of Computer Science, 4:1024–1029. [Bertino et al., 2005] Bertino, E., Ooi, B. C., Yang, Y., and Deng, R. H. (2005). Privacy and ownership preserving of outsourced medical data. In Proceedings of the 21st International Conference on Data

Engineering (ICDE '05), pages 521–532, Tokyo, Japan. IEEE Computer Society.

14. Huang, K., Yue, M., Chen, P., He, Y., and Chen, X. (2009). A cluster-based watermarking technique for relational database. In Proceedings of the 1st International Workshop on

Database Technology and Applications (DBTA '09), pages

107–110, Wuhan, China. IEEE Press.

15. Khanna, S. and Zane, F. (2000). Watermarking maps: hiding information in structured data. In Proceedings of the 11th annual ACM-SIAM symposium on Discrete algorithms (SODA '00), pages 596–605, San Francisco, California, United States. Society for Industrial and Applied Mathematics.

16. Lafaye, J. (2007). An analysis of database watermarking security. In Proceedings of the 3rd International Symposium on Information Assurance and Security (IAS '07), pages 462–

467, Manchester, United Kingdom. IEEE Computer Society.