Architecture for Recoverable Watermarking for Distributed Databases

Sana A. Rashid Hatture[1], Prof. Dr. Suhas D. Raut[2]

[1]M.E. Student, Dept. of Computer Science & Engineering, N.K. Orchid College of engineering and technology, Solapur, Maharashtra, India
[2] Ph.D. Professor, Dept. of Computer Science & Engineering, N.K. Orchid College of engineering and technology, Solapur, Maharashtra, India

---***---

Abstract - *Distributed database security has become an important issue. This paper provides method for providing security to shared databases in distributed environment. Such approach is applied for protecting numerical data in relational database. Advancement in information technology is playing an increasing role in the use of information systems comprising relational databases. These databases are used effectively in collaborative environments for information extraction; consequently, they are vulnerable to security threats concerning ownership rights and data tampering. Watermarking is advocated to enforce ownership rights over shared relational data and for providing a means for tackling data tampering. The proposed system provides security to the shared databases in distributed environment by recoverable watermarking where different users share their data in various proportions.*

Key Words: **Distributed Database, Relational Database, Recoverable Watermarking and Numerical Data.**

1. INTRODUCTION

Security 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 creates 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 [1].

The copyright protection inserts evidence into the digital objects without lossless of its quality. Whenever, the copyright of a digital object is in question, this information is extracted to identify the right full owner. Digital watermarking is the solution of embedding information in multimedia data. There are many techniques used to protect copyrights. Digital contents in the form of text document, still images motion picture, and music etc. are widely used in normal life nowadays. With the rapid grown of internet users, it boots up transaction rates (file sharing, distribution or change). Trend goes up dramatically and continues growing everyday due to convenient and easy to access. It is, hence, copyright protection becomes more concerned to all content owners. One of the best solutions to this problem is watermarking data. Watermarking of relational databases is very important point for the researches; because the free databases available on the internet websites are published without copyrights protection and the future will exploding problems.

2. LITERATURE REVIEW

Following are the techniques to ensure security in terms of ownership protection.

- Watermarking techniques have historically been used to ensure security in terms of ownership protection and tamper proofing for a wide variety of data formats. This includes images, audio, video, natural language processing software, relational databases [2] and more.
- Reversible watermarking techniques can ensure data recovery along with ownership protection.
- Fingerprinting (transactional watermarks), data hashing, serial codes are some other techniques used for ownership protection [3]. Fingerprints are used to monitor and identify digital ownership by watermarking all the copies of contents with different watermarks for different recipients. Primarily this type of digital watermarking tries to identify the source of data leakage by tracing a guilty agent. In hashing, digital contents can be saved by performing a one-way hash function whereby the data contents do not change. If the hash of the original and tampered data is the same, data authenticity can be verified but ownership cannot be proved easily.
- Serial or classification codes are used for filtering of inappropriate contents over the Internet and are mainly applicable to images, audio and video.

Reversible watermarking tries to overcome the problem of data quality degradation by allowing recovery of original data along with the embedded watermark information. Following are some existing systems:

- The first irreversible watermarking technique for relational databases was proposed by Agrawal et al. [2]. Similarly, the first reversible watermarking scheme for relational databases was proposed by Zhang et al. [4]. In this technique, histogram expansion is used for reversible watermarking of relational database. Zhang et al. proposed a method of distribution of error between two evenly distributed variables and selected some initial nonzero digits of errors to form histograms. Histogram expansion technique is used to reversibly watermark the selected nonzero initial digits of errors. This technique keeps track of overhead information to authenticate data quality. However, this technique is not robust against heavy attacks (attacks that may target large number of tuples).

- Difference Expansion Watermarking techniques (DEW), exploit methods of arithmetic operations on numeric features and perform transformations [5]. The watermark information is normally embedded in the LSB of features of relational databases to minimize distortions.

- Another reversible watermarking technique proposed 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 [6].

- Genetic Algorithm based on Difference Expansion watermarking (GADEW) technique is used in a proposed robust and reversible solution for relational databases [7]. GADEW improves upon the drawbacks mentioned above by minimizing distortions in the data, increasing watermark capacity and lowering false positive rate. To this end, a genetic algorithm is employed to increase watermark capacity and minimize introduced distortion. This is because the watermark capacity increases with the increase in number of features and the genetic algorithm runs on more features to search the optimum one for watermarking. However, watermark capacity decreases with the increase in watermarked tuples. GADEW used the distortion measures to control distortions in the resultant data.
- Prediction-error expansion watermarking techniques (PEEW) incorporate a predictor as apposed to a difference operator to select candidate pixels or features for embedding of watermark information [8]. The PEEW proposed technique by Farfoura et al. is fragile against malicious attacks as the watermark information is embedded in the fractional part of numeric features only. In this particular scenario, the scheme works because the intention of the attacker is to preserve the usefulness of the data; otherwise, he can easily compromise the fractional part.

- Robust and Reversible Watermarking (RRW): Genetic algorithm (optimization algorithm) is employed in the robust and reversible watermarking technique (RRW) to achieve an optimal solution that is feasible for the problem at hand and does not violate the defined constraints. An optimal watermark value is created through the GA and inserted into the selected feature of the relational database in such a way that the data quality remains intact. Mutual Information, a well-known information theory (concept), statistically measures the amount of information that one feature contains about the other features in a database. In RRW, mutual information is used to select a suitable (candidate) feature from the database for watermarking. In RRW, the knowledge of mutual information for every candidate feature is also employed to compute the watermark information. Thus, it is ensured that the data quality will not be affected [9]*.

3. PROPOSED SYSTEM ARCHITECTURE

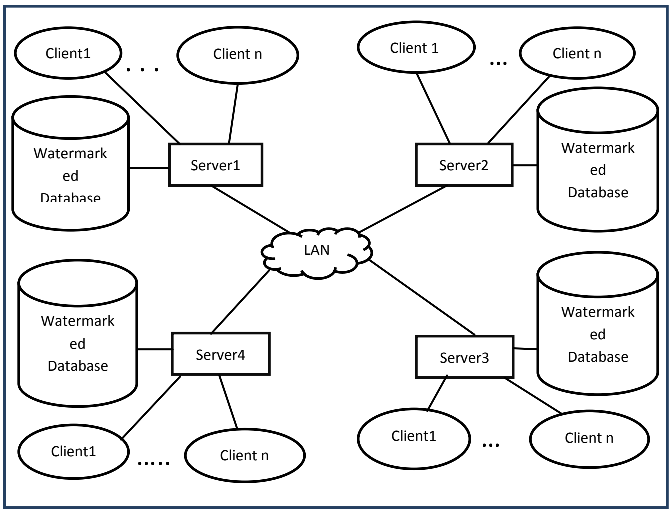

Fig -1: Proposed System Architecture

Above figure1, depicts proposed system architecture. The proposed system uses relational database. Here users share their data in various proportions. It provides security to shared databases in distributed environment. This is done by recoverable watermarking shared databases.

The proposed system consists of four servers. These servers are connected to each other via LAN and connected to respective client directly. Here database is relational database. Database design starts from global schema and processed by designing fragmentation and allocating fragments to different servers. Data is distributed by fragmenting the data and storing at different servers. Here horizontal fragmentation is done. Each Replica is located at least at 2 servers. Each server add watermark to numerical data in relational database. Only authorized client can access watermarked data. Here read authorization is provided to client and log is maintained for read only access. When authorized client send request to server, server checks validity of client. For this server maintains client table and checks user identification number and password match. If match found then client is valid otherwise refused.

The protocol communication between client and server is inter process communication call that is remote procedure call. Procedure call can be local or remote. Server gives reply with watermarked data when procedure call is local. For remote procedure call server forward client request to all servers and once it receives reply send result of requested procedure to client.

4. CONCLUSIONS

The proposed system provides security to the shared databases in distributed environment by recoverable watermarking numerical data where different users share their data in various proportions.

REFERENCES

[1] Snehal S. Kshatriya and Dr. S. S. Sane, "A Study of Watermarking Relational Databases," in International Journal of Application or Innovation in Ebgineering & Management , vol. 3,issue 10, October 2014 .

[2] 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.

[3] S. Subramanya and B. K. Yi, "Digital rights management," Potentials, IEEE, vol. 25, no. 2, pp. 31–34, 2006.

[4] Y. Zhang, B. Yang, and X.-M. Niu, "Reversible watermarking for relational database authentication," Journal of Computers, vol. 17, no. 2, pp. 59–66, 2006.

[5] G. Gupta and J. Pieprzyk, "Database relation watermarking resilient against secondary watermarking attacks," in Information Systems Security. Springer, 2009, pp. 222–236.

[6] J.-N. Chang and H.-C. Wu, "Reversible fragile database watermarking technology using difference expansion based on svr prediction," in Computer, Consumer and Control (IS3C), 2012 International Symposium on. IEEE, 2012, pp. 690–693.

[7] K. Jawad and A. Khan, "Genetic algorithm and difference expansion based reversible watermarking for relational databases," Journal of Systems and Software, 2013.

[8] D. M. Thodi and J. J. Rodriguez, "Prediction-error based reversible watermarking," in Image Processing, 2004. ICIP'04. 2004 International Conference on, vol. 3. IEEE, 2004, pp. 1549–1552.

[9] Saman Iftikhar, M. Kamran and Zahid Anwar," A Robust and Reversible Watermarking Technique for Relational Data," in Knowledge and Data Engineering, val X, No : XX,, IEEE, 2015