ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 6, June 2017

Security for Numeric Relational Database Using Traitor Identification: A Survey

Dhanashri Varute[1], Arti Mohanpurkar[2]

Department of Computer Engineering, D.Y.Patil's SOET, Lohegaon, Pune, India

ABSTRACT: The increasing popularity of an internet and its usage has raised the use of digital data at large extent. In this situation, the fundamental need is to provide protection for ownership of data and find out the blameworthy user. This appeals the facilities like content right protection and traitor identification on relational database. To fulfill these features, the fingerprinting schemes should focus on challenges like usability constraint non-violation, minimum distortion, robustness against collusion attack etc. This paper gives a brief overview of fingerprinting schemes which were proposed earlier to address these issues.

KEYWORDS: distortion free; fingerprinting; tardos code; watermarking; collusions

I. INTRODUCTION

Increased amount of internet usage has resulted in sharing huge set of digital data like audios, videos, documents, databases, files, images etc. between massive groups of users. Hence protection of data has become significant need. Fingerprinting is one of the techniques used to protect the valuable digital data from illegal distributions. Now a day's some data mining tools are being used to extract the patterns from a relational databases that include sharing of data. In such scenarios, the techniques like watermarking (to provide ownership protection) and fingerprinting also called traitor tracing (to find source of data and illegal redistribution of data) are used. In fingerprinting, it is required to mark each numerical copy with a unique identifier before distributing the data to user. If in case, the distributor finds any illegal copy of a data on the network at that instant the responsible user who created an unauthorized copy is traced. In our daily life, there are many such applications whose domain of data represents crucial assets (For example medical data, scientific data, stock market data, industrial data, weather data etc.); therefore traitor identification and ownership protection must be carefully advocated. To embed fingerprint in relational database, the user constraints which can admit a bitty amount of error are considered [16]. Furthermore, the distortions in original data are conserved up to valid limits by presenting usability constraints, to maintain the knowledge contained in database. Generating unique copies of the digital documents by embedding the fingerprints in it is a very natural process. Here, the problem will never arise if the buyers will not try to cheat but, there may be some pirates who try to destroy the fingerprint and redistribute the data illegally. That's why to prevent the data and detect the person who is trying to make the fraud; the unique marks are inserted in a digital document of relational database at the locations which are not known to the user [15].

II. LITERATURE SURVEY

A traitor identification system [1] securely embeds the fingerprint for providing protection to numeric relational databases. Digital data of numeric nature calls for preservation of usability and it needs to be done so by achieving the minimum distortion. The insertion technique with reduced time complexity ensures that fingerprint inserted in the form of an optimized error leads to the minimum distortion.

Fingerprint embedding in relational data is gained by considering user constraints which allow a small amount of error to be inserted [2]. Kamran proposed the scheme [2] which performs watermarking and meets different challenges as knowledge preservation, robustness against different attacks, and specifies usability constraints once for all. Every bit

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 6, June 2017

of multi bit watermark created from date-time is inserted in every (numeric) attribute of selected row so as to achieve robustness even after the attack on some selected part of the dataset. As a result of this 100% decoding accuracy is gained by the scheme even though only one watermarked row remains in the database.

Shah et al. [5] suggested a scheme for embedding watermark in the alphabetic data attributes. This method is applicable in all the languages with upper as well as lower character cases. The advantage of this technique is, it doesn't affect the semantic meaning of data if the case is changed from small to capital or vice versa. So the query result doesn't changes.

Ali et al. [6], proposed the technique which is based on using binary image to watermark the relational database. In this the bits of an image are segmented in short strings which are encoded in non-numeric, multi-word attributes of selected tuples of the database. A major advantage of using the space-based watermarking is the large bit-capacity available for hiding the watermark.

Rajneesh et al. [7], presented a secure method that uses both semantic and syntactic techniques to watermark the tuple in a relation. The Watermarking technique is dependent on secret key and on the relation. The algorithm is based on the concept of predefined signals of ASCII characters. A secret key is generated by using these signals only. To embed a watermark the concept of abbreviations for words and also one of syntactic approach is used.

Bedi et al. [8], proposed watermarking technique used for data authentication and integrity of relational database. For integrity verification of tables in the database, the watermark is depending on a secret key and the original copy of relation. The method uses the concept of eigen values for generating the watermark for a record in tuple. Watermark embedding is completed by using eigen values in a non-numeric attribute of a tuple. Detection of the watermark proves authentication and integrity of data.

Pramod et al. [9], presented new scheme for watermarking non-numeric relational databases. This technique uses the voice of copyright holder to generate watermark, and then corresponding insertion and detection algorithms are applied.

Maximum research in fingerprinting is done upon multimedia data. But, since relational data operations and its properties are different, the techniques used for multimedia data cannot be used as it is to relational data [14].

Watermarking scheme which meets the challenges like robustness against several attacks, preserves knowledge in database, keep the balance between conflicting buyer and owner usability constraints is and developed by M. Kamran et al in [2]. The proposed algorithm embeds each bit of a multi-bit watermark in every selected row; as a result it achieves 100 percent decoding accuracy even though only one watermarked row is left in database.

Benjamin Mathon [13] presented an approach for colluding strategies relying on estimation error rate Єa and derived a new worst Єa-Worst Case attack strategy on tardos fingerprinting code.

Julien Lafaye [14] proposed collusion secure pairing algorithm and Integer Linear algorithm by using Tardos code for fingerprinting which considers local and global constraints also on the data but works on modifying least significant bit of numerical values of relational data.

Technique for fingerprinting relational data is presented by Yingjiu Li [15] by extending watermarking scheme proposed in [17]. Here, several measures for robustness of fingerprinting technique are defined along with solution for collusion attacks. This scheme can be used for watermarking as well as fingerprinting, but at the time inserting fingerprint into data the usability of data is not considered.

In [16], Tardos described a fully randomized binary fingerprinting code further tightening the lower bound. This fingerprinting technique is basically known for its short code length. While using the tardos code the owner must be aware about the number of users in advance.

Agarwal et. al [17] presented the watermarking scheme which is based on numeric attributes and bit-level marking. This scheme ensures that some bit positions for some attributes of some tuples in the relation contains specific error values. These bit patterns constitutes a watermark. Selecting parameter for watermarking is based upon computing message authenticated code (MAC) which is calculated by the use of secret key and tuple's primary key. Though the least significant bit (LSB) based data hiding techniques are efficient, shifting LSB by only one position might lead to the watermark loss without much damage to the data. During watermark embedding process ,this scheme assumes unconstrained LSB manipulation.

A binary randomized code [18] that uses concatenation of the partly randomized inner code and outer code along with Marking Assumption where colluder can't alter the marks at which their copies agree is proposed by Boneh and Shaw.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 6, June 2017

III. COMPARATIVE ANALYSIS

In systems presented in [7], [8] and [9] the attacker can randomly select and use a subset of the original data set that might still provide value for the intended purpose. The attacker hopes that, the selected subset will not contain the watermark. However, the algorithm developed in [1] embeds the watermark in the whole database randomly and depending on hash function, so it is very difficult to guess where the watermark is embedded in the database. Also, the detection algorithm checks for a watermarked tuple by the same way as embedding, so it scans a primary key of subset selection attributes. This kind of attack does not changes the primary key or the case of attribute that's why even if attacker attacks only 10% of original database, the system will detect watermark with 100%.
In such type of attacks, an attacker might insert a set of tuple in the original database. The system can be executed with various percentages of insertion. These experiments run on fraction =10, N=13270, α =0.01 and v=3. The system runs twice at each percentage then the results each time are recorded, after that the maximum percentage of detection is taken. In such kind of attack, the attacker may delete a subset of the tuples of the watermarked database by hoping that the watermark will be removed but the detection algorithm scans the marked tuple by primary key, and therefore even ifs an attacker deletes tuple, then also the detection algorithm doesn't effect because detection algorithm increases (totalcount) and (matchcount), iff detect marked tuple which still has marked. This means that Watermark can't be completely removed from all the tuples because if only 5% tuples are there then a watermark can be extracted from the database. As a result the system developed in [1] is considered as very robust against any kind of attacks that did not change the case of attributes or primary key.

IV. CONCLUSION

The fingerprinting technique inserts the fingerprint bits in the numeric relational database subject to usability constraints. This results in minimum distortion in the original data set as well as it facilitates finding the guilty user (if any) who is responsible for illegal redistribution of data set. The system does blind decoding and is considered to be robust against attacks like tuple insertion, tuple deletion, attribute deletion and mix and match. A solution for additive attack is also mentioned. The decoding accuracy is found to be 100% in spite of the numerous attacks. In addition, collusion avoidance is achieved because of the novel fingerprint embedding method. A considerable improvement in the execution time (approximately more than 80%) of the fingerprint insertion algorithm is observed over that of the insertion algorithm of the existing system.

REFERENCES

[1] Arti Arun Mohanpurkar, Madhuri Satish Joshi (2016) "A Traitor Identification Technique for Numeric Relational Databases with Distortion Minimization and Collusion Avoidance" International Journal of Ambient Computing and Intelligence Volume 7, Issue 2, pp 114-137 DOI: 10.4018/IJACI.2016070106, July-December 2016
[2] Kamran, M., Suhail, S., & Farooq, M. (2013). A Robust, Distortion Minimizing Technique for Watermarking relational Databases Using Once-for-all Usability Constraints. IEEE Transactions on Knowledge and Data Engineering, 25(12), 2694–2707. doi:10.1109/TKDE.2012.227
[3] Rakesh Agrawal Jerry Kiernan,"Watermarking Relational Databases",IBM Almaden Research Center,2002.
[4] Yossra H. Ali, Bashar Saadoon Mahdi, "Watermarking for Relational Database by using ThresholdGenerator",Computer Sciences Department, University of Technology Baghdad,2011.
[5] S.A. Shah, Sun Xingming and Hamadou Ali , "Query Preserving Relational Database Watermarking ",Network and Information Security Lab Hunan University, Changsha, Hunan, China ,2010.
[6] Ali Al-Haj, Ashraf Odeh, and Shadi Masadeh, "Copyright Protection of Relational Database Systems", Amman, Jordan,2010 .
[7] Rajneesh Kaur ,Bedi,Purva, Gujarathi,Poonam Gundecha,Ashish Kulkarni,"A Unique Approach for Watermarking Non-numeric Relational Database", International Journal of Computer Applications (0975 – 8887) Volume 36– No.7, December 2011.
[8] Bedi R., Thengade A., Wadhai V., "A New Watermarking Approach for Non Numeric Relational Database", 2011.
[9] pramod a.kharade,vikramsinh m. pokale, vijay d. chougule and vishal v. mangave,"speech based watermarking for non-numeric relational database" , international journal of innovative research in engineering & science,april 2013.
[10] Nahla El_Haggar, Mahmoud M. Elkhouly, Samah S. Abu El Alla," Blind Watermarking Technique for Relational Database",COMPUSOFT, An international journal of advanced computer technology, May-2013.
[11] Abha Tamrakar, Chhattisgarh Swami, "Compression of Watermarked Relational Database for Security and Optimization of Storage Consumption ", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-1, Issue-2, December 2011.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 6, June 2017

[12] Brijesh B. Mehta," A Novel approach as multiplace watermarking for security in database", Dept. of Computer Engineering, S. V. National Institute of Technology, Surat, Gujarat, INDIA-, 2010-2011.
[13] Benjamin Mathon, Patrick Bas, Francois Cayre and Benoit Macq, ―Impact of Watermarking Security on Tardos-Based Fingerprinting‖, IEEE Transactions on information forensics and security, vol. 8, No. 6, june 2013.
[14] Julien Lafaye, David Gross-Amblard, camellia Constantin, and Meryem Guerrouani, ―WATERMILL : An Optimized Fingerprinting System for Databases under Constraints‖, IEEE Transactions on Knowledge and Data Engineering,vol. 20,no. 4, April2008.
[15] Yingjiu Li, Vipin Swarup, Sushil Jajodia, ―Fingerprinting Relational Databases: Schemes And Specialties‖, IEEE Transactions,vol. 2, no. 1, 2005.
[16] G. Tardos, ―Optimal probabilistic fingerprint codes,‖ in Proc. 35th Ann, ACM Symp. Theory of Computing, May 2003, vol . 55, no. 2, pp. 116-125[online]. Available: http://portal.acm.org/citation.cfm?doid=1346331346335, ACM.
[17] R. Agrawal, P.J. Haas, and J. Kiernan, ―Watermarking Relational Data: Framework, Algorithms and Analysis,‖ The VLDB J., vol. 12,no. 2, pp. 157-169, 2003.
[18] D. Boneh and J. Shaw, ―Collusion Secure Fingerprinting for Digital Data,‖ IEEE Trans. Information Theory, vol. 44, no. 5, pp. 1897-1905, 1998.