Watermarking of relational databases: Survey

Rucha D. Kulkarni[1], **Dipak V. Patil**[2]

[1]*Student, Department of Computer Engineering*

Gokhale Education Society's R.H. Sapat College of Engineering, Management Studies and Research, Nasik, Maharashtra, India

[2]*Associate professor, Department of Computer Engineering*

Gokhale Education Society's R.H. Sapat College of Engineering, Management Studies and Research, Nasik, Maharashtra, India

---***---

Abstract - ***Watermarking techniques provides candidate solution to ensure security in terms of ownership protection and tamper proofing for a various types of data. Digital watermarking of multimedia content is more commonly known. There is rich body of literature for watermarking of multimedia data. But watermarking of relational databases is emerging area as compared to multimedia data. Though it is emerging various techniques are proposed to ensure security against variety of attacks, ownership right protection and data tampering. This paper focuses on different techniques that have been proposed to provide solutions for security problems related with relational data. This survey provides different techniques classified according to their intent along with type of watermarking.***

Key Words: *Relational database, watermarking*

1. INTRODUCTION

Now a day's digital data can be accessed and exchanged through computer via internet has growing extensively which is very simple task. As digital data available publicly it can be easily modified by unauthenticated user and can steal rights of it. So, data security is an essential area that provides variety of solutions for protection of different data formats.

Watermarking is one of the popular and extensively used techniques that ensure security in terms of ownership protection and tamper proofing for a various data formats. Fingerprinting, data hashing, serial codes are some other techniques used for ownership protection [2]-[4].Using these other techniques one can identify source of data leakage but can't protect data from being leaked. Digital watermarking provides a strong method of protecting digital data from modification, copyright protection by embedding a secret code directly into the data. The embedded secret code, called watermark, can be used in various applications. Watermarking has the property that it can provide ownership protection to digital content inserting watermark unique to the owner. The embedded watermark can subsequently be used for proving and claiming ownership. It is very important to protect the ownership of databases, many times making copy of databases may get ignored. We only care about is relational database is authentic and unmodified, and if modified discovered and recovered.

Initially watermarking is restricted only up to multimedia content such as images, audio, video [7]-[10] etc. Particularly image watermarking is used while transmission of messages from one party to another. Processing of relational database watermarking differs that of watermarking techniques that are applied to multimedia data, cause is difference in properties of data. As relational data is independent and discrete compared to multimedia data is continuous.

Thus watermarking particularly for relational databases was proposed very firstly by [11]. The technique was irreversible in nature i.e. it can't regenerate original data from watermarked data using secret key. Further after few years reversible watermarking techniques get proposed by [12] that can regenerate data without comprising original quality.

Watermarking techniques mainly used to protect publicly available data from being tampered, protect ownership [13] of that data, ensure integrity [14] and such other purposes.

Some of the important approaches of watermarking are introduced in this paper as follows: Section 2 gives general processing for watermarking that are explained using fig.1. Section 3 includes various Surveys' on variety of watermarking techniques and finally in Section 4 conclusion was made.

2. GENERAL STRUCTURE OF WATERMARKING

There are main four phases they are as follows-
Watermark creation-
To insert watermark into selected data of original

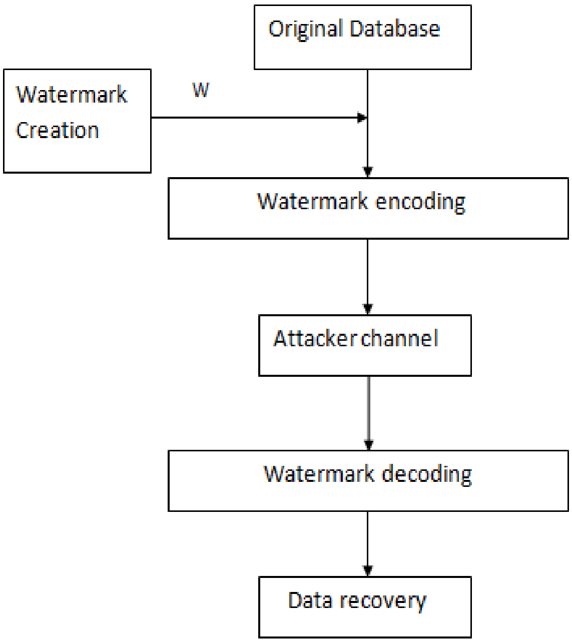

Fig -1: General structure of watermarking

database first it is essential to create optimal watermark. There are various methods that can be use to create watermark such as Genetic algorithm [1], Particle swarm optimization [15]-[17], simulated annealing [18] etc.

Watermark encoding-
In this phase watermark created in first phase is embedded in selected part of data from relational database. There are varieties of methods in a literature that can be applied to encode a watermark.

Watermark decoding-
Decoding is third phase in watermarking process which is used to extract embedded watermark from data which have been undergoes through watermarking. Way and method of encoding as well as decoding changes as technique changes.

Data recovery-
This phase is involved in overall watermarking process according to nature of watermarking i.e. whether it is reversible or irreversible. If watermarking is able to generate original data then data recovery phase is involved in watermarking process otherwise it is not.

3. ANALYSIS OF WATERMARKING TECHNIQUES

Watermarking database relation is one of the several areas which demands research on focus owing to commercial implications of database theft. Digital watermarking for relational databases emerged as candidate solution that provides copyright protection, tamper detection, traitor tracing, and maintaining integrity to the relational data. Watermarking of relational databases can be classified into two parts based on their ability to regenerate original data. They can be viewed as irreversible watermarking and reversible watermarking technique. Image watermarking has rich literature comparing to relational database, though in reality there are many techniques that are proposed to provide solution against problem of data security. Such techniques and there details are given as below,

3.1 Irreversible watermarking techniques

Agrawal and kierman[11] proposed first well-known conventional irreversible database watermarking scheme for watermarking numerical values in relational databases. The fundamental assumption is the watermark database that can tolerate a small amount of errors. In this scheme, the private key K, used for copyright verifiability, concatenated with primary keys of the tuple & is the seed for the pseudo random number generator algorithm which decides the tuples, attributes with inside a tuple,and bit positions within an attribute to be marked. Only when the chances of attacker have access to the private key, it also detects the watermark with high possibility. The technique survives several attacks and which preserves mean and variance of all numerical attributes. This scheme cannot be directly applied to watermarking categorical data since no matter how small it is, to a categorical value may render the value meaningless.
Agrawal, R., Haas, P. J., and Kiernan, J. [19] proposed technique is robust amongst various forms of malicious attacks as it starts updating to the data. The technique ensures there are some positions of some of the attributes stating some of the tuples contain specific values. The specific location and value changes algorithmically & are determined under control of a secret key known only to the owner of the data.bit pattern constitutes the watermark. Only one has access to the secret key can watermark & is detected with high possibility.
Sion, R [13] introduced a novel method of rights protection for categorical data through watermarking to solve this problem. This scheme involves watermarking of categorical attribute by changing some of its value to other value to the attribute (e.g. red is changed to green) if such change is tolerable in certain applications. New watermark embedding channels were used for relational data with categorical types. Novel watermark encoding algorithms

are designed and analyzed important theoretical bounds including mark vulnerability in this scheme.

3.2 Reversible watermarking techniques

Recently, there has been little progress in database watermarking, most of the watermarking schemes modeled irreversible database watermarking scheme firstly proposed by Agrawal and Kiernan[11]. Reversibility is the potentiality to generate the original relation from the watermarked relation using a secret key. There are several techniques to implement reversibility, some of them have been mentioned below.

3.2.1 Techniques for handling ownership protection/copyright protection

Udai Pratap Rao, et.al [2] proposed a scheme mainly for protection of ownership rights. In these technique bits of binary image is inserted in relational database. It minimizes variation in watermarked relational database. Proposed technique found robust against malicious attacks.
G.Shyamala, et.al [3] provides mechanism for proofing its ownership basing the secure embedding of a robust imperceptible watermark in relational data. Technique is resilient for the watermark synchronization error because it uses a partitioning approach that does not require marker tuples.
Jun Ziang Pinn and A. Fr. Zung, [4] proposed an efficient database watermarking algorithm based on inserting binary image watermarks in numeric mutilator attributes of selective database tuples. Proposed technique must be suitable for different area like e banking, multimedia industry,film industry etc.
Theodoros Tzouramanis,[5] Proposed a novel watermarking technique that accurately identifies true owner of database. It results in resilience to range of attacks. Algorithm presented in paper works on bits of tuples in relational data. After analysis of results produced it is found that proposed technique protect embedded data from errors.
G.Shyamla, et.al [6] proposed a security mechanism that helps to resolve ownership conflicts over watermarked dataset in case of additive attacks. And this method provides less distortion and maximum accuracy for decoding. It proves the robustness of watermarking scheme by analyzing its decoding accuracy under different types of malicious attacks using a real world dataset. It also provides solutions to resolve conflicting ownership issues in case of the additive attack.

3.2.2 Histogram expansion

Zhang et al. [12] proposed first reversible watermarking of relational database to achieve less as well exact authenticate of relational databases via expansion on data error histogram.This method has distributive error within two evenly distributed variable as some initial nonzero digits of errors to form histograms. Histogram expansion technique is reversible watermarking scheme possesses the ability of perfect restorative to the original attributive data from untampered watermarking relational databases, thus guaranteeing a clear and exact tampered or not authenticate without any worry about causing any permanent distortion to the database. This technique keeps track to overhead information to authentically data quality. This technique is not robust against heavy attacks.

3.2.3 Difference expansion watermarking technique (DEW)

Difference expansion refer to series of arithmetic operator on two integer value and a bit that result into a pair of modified integer from the original pairing integer the bit would be regenerated [20][21]. Difference expansion has previously been applied in image watermarking, but application in database watermarking introduction with an additional constraint of limiting distortion. Initially DEW was applied to secure image and after when it is needed to provide security to relational database it was applied to it. Proposing a high capacity algorithm based on the different expansion of triplets which is developed for embedding reversion watermark with reasonable level of image distortion. The algorithm uses a spatial and spectral tripling of pixel to hide a pair of bit which allows the algorithm to hide a large amount of data.
G.Gupta et.al [23] proposed a reversible watermarking method that has ability to recover watermarked database and blind in nature that means to recover original data, secret key along with watermarked data. Previously proposed methods aren't either reversible or completely blind in nature. To overcome this problem reversible blind watermarking method was proposed which recover original data with high quality provides ownership identification and resistant to secondary watermarking attacks.
Gaurav Gupta and Josef Pieprzyk [24] proposed an improvement over the reversible and blind watermarking scheme presented , identifying and eliminating a critical problem with the previous techniques related with reversible watermarking. The proposed scheme provides high security against secondary watermarking Attacks such as subtractive, distortive, shuffling, and additive attacks. The scheme introduces less distortion in the data items, with an adjustable upper bound, in order to maintain usability. Experiments showing that the average watermark detection rate is estimated about 91% of even with attacker distortion of its half to the attributes. In addition with this Even if the attacker re-watermarks an already marked database relation, the rightful owner can be identified by the virtue of reversibility.

3.2.4 SVR (Support Vector Regression)

Jung-Nan Chang and Hsien-Chu Wu [25] proposed scheme that detects database tamperer by embedded importance characterize of the originality of database. as ore additive with support vector regression (SVR) is applied to get the predicted each protective attributed value. The association rule of frequent pattern tree (FP-tree) data mining is used to detect the relationship existing with the protected attribute and others as well in the database. Support vector regression (SVR) is to be applied to concertedly predict each protective attribute value. If the protective database is distort then SVR function will still predict the protected values. Then, an examination of the difference between original protective and predicted value allow the extraction of the watermark. Data which has been tampered can be found by comparing original watermark with the extracted one. FP-tree mining method is used to reduce SVR train time. Moreover, if the database cannot be attacked then the proposed method can recover the original attribute values.

3.2.5 GADEW (Genetic Algorithm based on Difference Expansion Watermarking)

Khurram Jawad, Asifullah Khan [26] introduced new robust technique for reversible watermarking approach for the protection of relational databases. While the approach is on the idea of difference expansion as on utilizing genetic algorithm (GA) to improve watermark capacity and to reduce distort error. The proposed approach is reversible and therefore, distortion introduced after watermark insertion can be fully restored. GA introduces some randomness in DEW technique, thus making it is difficult to the attacker to predict attribution. Security of the watermarking system is also enriched by reduction on the distort and minimize abrupt changes caused by DEW. They have achieved this by two measures added in the fitness function of GA, first by using the knowledge of their neighbor value of the relational database as well in second by minimizing the distortion introduced while selecting attributes resulting with minimum distortion.

3.2.6 PEEW (Prediction Error Expansion Watermarking Technique)

M. E. Farfoura and S.-J. Horng, [27] presented a novel blind reversible watermarking method that ensures us the ownership of protection in the area of Relational Database of water marking. Whereas previous technique has been mainly concerned with introducing permanent errors in the actual data, as our approach assure 100% recovery of the original database. In the proposed method, as using a reversing data embedding technique so called prediction error expansion on the integers as well to achieve its reversible action. The watermark detection can be successfully completed even on 70% of watermarked relation tuples are deleted.

Mahmoud E. Farfoura, el.al [28] utilized a reversible data embedding technique called prediction-error expansion on integers to achieve reversibility, by introducing a novel blind reversing watermarking method assuring ownership protection in the area of RDB watermarking. This assures full recovery of the original database relation after the watermark has been detected and authenticated.

D. M. Thodi and J. J. Rodriguez [29] proposed a alternative solution to distortion by using histogram shifting technique. The method improves distortion performance at low embedding capacity. To improve this new technique was presented called prediction error expansion. Along with this it was found that combination of prediction error and histogram shifting methods was effective to lower down distortion.

Mahmoud E. Farfoura ,Shi-Jinn Horng a, et.al [30] designed and utilize an authentication protocol based on an efficient time-stamp protocol, and we propose a blind reversing watermarking method that assure ownership protection in the area of relative database watermarking. as well previous techniques has mainly concerned with introducing permanent errors with original data, approaching ensure one hundred percent recovery of original database relative to the owner-specific watermark has detected and authenticated. In the proposed watermarking method, utilizing a reversible data embedding technique so called prediction error expansion on integer achieving reverse action. The detection of the watermark would be completed successful even in 95% of a watermarked relation tuples deleted even. There are extensive analysis showing the proposed scheme is robust opponent to various forms of database attacks including adding as well as deleting and shuffling or modifying tuples or attributes.

X. Li, B. Yang, and T. Zeng, [31] presented a technique to improve the embedding capacity i.e. reversible watermarking using a adaptive predicative error expansion & pixel selection. This method is a improvement to conventional PEE by adding two new techniques adaptation with embedding & pixel selection as well with it. Instead of uniform embedding we deceptively embed one or two bits are expandable pixels as prescribed by the regional complexity.

3.2.7 Watermarking techniques using soft computing optimization methods

K. E. Parsopoulos and M. N. Vrahatis, [15] proposed a optimization technique that is used to solve constraint optimization (CO) problem. It is found that PSO (Particle Swarm Optimization) is acting like a good alternative to other optimization techniques. Non-stationary, multistage, penalty function is implemented in this paper and results are compared with evolutionary algorithms like genetic algorithm.

R. Hassan, et.al [16] did comparison between two popular and emerging optimization techniques PSO and GA. Authors tested better solution by both the methods based on three problems. And Solution to these problems gave a indication about which optimization method is better than other in which case of problem. A last according to tested problems it is found that PSO slightly outperformed than GA.

M. Kamran and M. Farooq, [17] presented a Information preserving watermarking scheme for Electronic Medical Records(EMR). The proposed scheme was implemented with the intent of building EMR system to so medical system strong in future. As medical domain is extremely sensitive because records provided in it suggest relevant diagnosis, thus records of patients are vulnerable to the attacks and the scheme provide strong solution to the same.

3.2.8 Other watermarking techniques

Erik Sonnleitner[32] proposed a watermarking algorithm based on parameter that tuple partitioning o watermarking and white spaces while using it in public watermark. The watermarking scheme is non intrusive resilient as well blind to reversibly and its suitable for databases of any sizes with reasonable performances on embedding and extraction. Moreover authors emphasize locatability of malicious changes with the scope that of predefined tuple sets and support incremental water marked to conceive with dynamic natural data base system.In addition with this which are shown with whitespace substitution for the purpose of information hiding within database relations offers significant potential for watermarking scenarios.

Javier Franco-Contreras, Gouenou Coatrieux, et.al[33] proposed the robust reversible watermarking modilized by originally proposed under Vleeschouwer et al for the images protection of given relational databases. The propose scheme states relative angular position of the circular histogramic centeric mass of one numerical attribute for message displaying and embedding. It can be used for verifying databases authenticate and for traceability when identifying database origin after it has been modified. Evaluation of this scheme is done in terms of capacity, distortion, and robustness with two common database modifications against it. addition and removal of tuples. To that end that even model the impact of the embedding process and database modified for the probability even distributive to the center mass position.

M. Kamran Sabah Suhail, and Muddassar Farooq[34] proposed a robust and efficient water marking scheme for relation to the database that enables to meet four challenges.The technique is robust against different types of attacks that of an intruder could launch to corrupt the embedded watermark. Second is it is able to preserve the databases to made them an effective component of knowledge-aware decision third to strike the balance between requirements for database owners who requires softly usable constraints as well databases recipient who need tight usability constraint which ensure minimum distortions in data and last but not least, it is not require that a database owner defines usability constraint even for each typed applications and recipients separately.

Mohamed Shehab [35] presented a mechanism for a proof ownership basing on the secure embedding of a imperceptible watermark to be robust in relational data. Authors formulate the water marked relational databases as optimized problem discussing efficient techniques to solving the optimized problem as well to handle the constraints. Watermarking technique presented here is resilient to watermark the synchronization errors because it uses a partition of approach that does not require marker tuples. This approach overcomes a major weakness in previously proposed watermarking techniques. Watermark decoding is based on a threshold technique characterized with optimal threshold techniques that minimize the probability of decoding errors. After implementation of presented technique it is found. This technique is proper enough to tuple deletion alteration as well for insertion attacks.

4. CONCLUSIONS

In this paper we survey the different watermarking techniques for relational databases. Initially watermarking techniques for relational data couldn't produce original data after extracting watermark from selected attributes. As research goes in forward direction reversible watermarking techniques proposed by researchers as have seen in survey. These techniques provide candidate solution to problems of data security, ownership right protection and other problems. Current work in this research area try to provide reversible as well as robust solution so that data that undergoes watermarking won't comprise with quality, integrity and it would be resilient to attacks.

REFERENCES

[1] Saman Iftikhar, M. Kamran, and Zahid Anwar, "RRW—A Robust and Reversible Watermarking Technique for Relational Data" , IEEE Transactions On Knowledge And Data Engineering, Vol. 27, No. 4, APRIL 2015.

[2] Udai Pratap Rao, Dhiren R. Patel, Punitkumar M. Vikani, "Relational Database Watermarking for Ownership Protection", 2nd International Conference on Communication, Computing & Security [ICCCS-2012]

[3] G.Shyamala, I.Jasmine Selvakumari Jeya, M.Revathi, "Secure and Reliable Watermarking in Relational Databases", *International Journal of Computer Trends and Technology (IJCTT) – volume 11 number 1 – May 2014*

[4] Jun Ziang Pinn and A. Fr. Zung, "A new watermarking technique for secure database", International Journal of Computer Engineering & Applications, Vol. I, No. I

[5] Theodoros Tzouramanis,"A Robust Watermarking Scheme for Relational Databases", 6th International Conference on Internet Technology and Secured Transactions, 11-14 December 2011, Abu Dhabi, United Arab Emirates 2011 IEEE

[6] G. Shymala, C. Kanimozhi, S. P. KAVYA, "An Efficient Distortion Minimizing Technique for Watermarking Relational Databases", International journal of scientific research and Technology research, Vol.04,Issue.11,May-2015

[7] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," IEEE Trans. Image Process., vol. 6, no. 12, pp. 1673–1687, Dec. 1997.

[8] I. Cox, M. Miller, J. Bloom, and M. Miller, "Digital Watermarking".Burlington, MA, USA: Morgan Kaufmann, 2001.

[9] P. W. Wong, "A public key watermark for image verification and authentication," in Proc. IEEE Int. Conf. Image Process., 1998, vol. 1,pp. 455–459.

[10] P. W. Wong and N. Memon, "Secret and public key image watermarking schemes for image authentication and ownership verification," IEEE Trans. Image Process., vol. 10, no. 10, pp. 1593–1601, Oct. 2001.

[11] Rakesh Agrawal Jerry Kiernan, "Watermarking Relational Databases", Proceedings of the 28th VLDB Conference,Hong Kong, China, 2002

[12] Yong Zhang, Bian Yang, and Xia-Mu Niu,"Reversible Watermarking for Relational Database Authentication" ,J. Comput., vol. 17, no. 2,pp. 59–66, 2006.

[13] Sion, R. (2004). "Proving ownership over categorical data". In Proceedings of the 20th IEEE international conference on data Engineering ICDE, April 2004 (pp. 584–596).

[14] Yingjiu Li Huiping Guo Sushil Jajodia, "Tamper Detection and Localization for Categorical Data Using Fragile Watermarks", *DRM'04,* October 25, 2004, Washington, 2004 ACM

[15] K. E. Parsopoulos and M. N. Vrahatis, "Particle swarm optimization method for constrained optimization problems," Intel.Technol.–Theory Appl. New Trends Intell. Technol., vol. 76, pp. 214–220, 2002.

[16] R. Hassan, B. Cohanim, O. De Weck, and G. Venter, "A comparison of particle swarm optimization and the genetic algorithm," in Proc. 46th AIAA/ASME/ASCE/AHS/ASC Struct., Struct. Dyn. Mater. Conf., 2005, pp. 1–13.

[17] M. Kamran and M. Farooq, "An information-preserving watermarking scheme for right protection of EMR systems," IEEE Trans. Knowl. Data Eng., vol. 24, no. 11, pp. 1950–1962, Nov. 2012.

[18] Richard William Piper Meana, "Approximate Sub-Graph Isomorphism ForWatermarking Finite State Machine Hardware", January 2013

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

[20] Jun Tian, "Reversible Data Embedding Using a Difference Expansion", IEEE Transactions on circuits and systems for video technology, Vol. 13, No. 8, August 2003

[21] Jen-Bang Feng, Iuon-Chang Lin, Chwei-Shyong Tsai, and Yen-Ping Chu, "Reversible Watermarking: Current Status andKey Issues", International Journal of Network Security, Vol.2, No.3, PP.161–171, May 2006

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

[23] G. Gupta and J. Pieprzyk, "Reversible and blind database watermarking using difference expansion," in Proc. 1st Int. Conf. Forensic Appl. Tech. Telecommun., Inf., Multimedia Workshop, 2008, p. 24

[24] G. Gupta and J. Pieprzyk, "Database relation watermarking resilient against secondary watermarking attacks," in Information Systems and Security. New York, NY, USA: Springer, 2009, pp. 222–236.

[25] 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.

[26] K. Jawad and A. Khan, "Genetic algorithm and difference expansion based reversible watermarking for relational databases,"

[27] J. Syst. Softw., vol. 86, no. 11, pp. 2742–2753, 2013

M. E. Farfoura and S.-J. Horng, "A novel blind reversible method for watermarking relational databases," in Proc. IEEE Int. Symp.Parallel Distrib. Process. Appl., 2010, pp. 563–569.

[28] D. M. Thodi and J. J. Rodriguez, "Prediction-error based reversible watermarking," in Proc. IEEE Int. Conf. Image Process. 2004, vol. 3,pp. 1549–1552.

[29] D. M. Thodi and J. J. Rodriguez, "Expansion embedding techniques for reversible watermarking," IEEE Trans. Image Process.,vol. 16, no. 3, pp. 721–730, Feb. 2007.

[30] M. E. Farfoura, S.-J. Horng, J.-L. Lai, R.-S. Run, R.-J. Chen, and M. K. Khan, "A blind reversible method for watermarking relational databases based on a time-stamping protocol," Expert Syst. Appl., vol. 39, no. 3, pp. 3185–3196, 2012.

[31] X. Li, B. Yang, and T. Zeng, "Efficient reversible watermarking based on adaptive prediction-error expansion and pixel selection," IEEE Trans. Image Process., vol. 20, no. 12, pp. 3524– 3533, Dec. 2011.

[32] E. Sonnleitner, "A robust watermarking approach for large databases," in Proc. IEEE First AESS Eur. Conf. Satellite Telecommun.,2012, pp. 1–6.

[33] Javier Franco-Contreras,Gouenou Coatrieux,Fréderic Cuppens Nora Cuppens-Boulahia, Christian Roux, Robust "Lossless Watermarking of Relational Databases Based on Circular Histogram Modulation", IEEE Transactions on information forensics and security, Vol. 9, No. 3, March 2014

[34] 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

[35] Mohamed Shehab,Elisa Bertino,Arif Ghafoor, "Watermarking Relational Databases Using Optimization-Based Techniques", IEEE Transactions On Knowledge And Data Engineering,, Vol. 20, NO. 1, Jan 2008