COMPARATIVE ANALYSIS OF FRAGILE WATERMARKING TECHNIQUES FOR WEB DATABASES

[1.] Dr. Vidhi Khanduja, [2.] Dr. Roshini Rawal
[1,2]Associate Professor
[1.] Computer Engineering Department, [2.] Applied Science And Humanities Department
[1,2] SAL Institute of Technology and Engineering Research, Ahmedabad, Gujarat, India.

Abstract: With the Internet becoming part and parcel of our lives, there is an increase in vulnerability to copyright and piracy threats as well. Digital Watermarking has emerged as an effective TPM to protect databases that are either shared or outsourced from illegal infringement. The state-of-art in the domain of Integrity losses has been analysed and the results of different fragile watermarking techniques for relational and decision systems is discussed. Different phases of fragile watermarking technique is analyzed. Additionally, theoretical analysis of the tamper detection is done analysing various cases w.r.t. watermark extracted and re-generated. False Hit Rate is also considered as a measure for security analysis.

IndexTerms: Fragile Watermarking ,Tamper Detection, Technological Protection Measures, Web Databases.

I. INTRODUCTION

With the Internet becoming part and parcel of our lives, there is an increase in vulnerability to copyright and piracy threats as well. Technological techniques are devised to protect digital content such as images, video and databases. Such methods are called “Technological Protection Measures” (TPM) [1]. They include the methods that are used to control access to copyright digital material or to prevent users from copying the protected content. Digital Watermarking has emerged as an effective TPM to protect databases that are either shared or outsourced from illegal infringement.

Databases are collection of knowledge assembled by the combined efforts of mankind across regions and through ages. Majority of the material on internet is dynamically created from digital databases. Significant efforts in terms of human effort, money and creative inputs to build these databases is required. They are thus considered as an Intellectual Property. However, with the increase in ease of availability of recent technologies to alter them, it is crucial to protect digital databases against potential misuse.

Watermarking provides the method to protect them. They are classified into two broad categories:

1. **Robust Watermarking Technique:** Watermarking techniques are designed to protect the ownership issues related to databases on web [2].

2. **Fragile Watermarking Technique:** Watermarking techniques are designed to protect the integrity of databases on web.
   In this work, we discuss and analyse the fragile watermarking techniques that tackle integrity issues related to web databases.
   Fragile watermarking techniques are designed to have four phases i.e.
   i. *Watermark Preparation*: Watermark that is to be inserted into databases is prepared. This may use secret parameters to enhance level of security.
   ii. *Watermark Insertion:* Prepared watermark bits are embedded within the database using secretly selected parameters.
   iii. *Watermark Extraction*: To prove the integrity losses, watermark is extracted. This step uses the algorithm which is just the reverse of embedding algorithm.
   iv. *Tamper Detection:* This phase decides whether the database is tampered with or not. Thus, integrity losses are detected in this step. The watermark inserted is compared with watermark extracted to get the results.

Section 2 discusses various watermarking techniques using cryptography. In section 3 we provide the applications of Information Security in watermarking techniques followed by security analysis of these techniques in section 4. We finally conclude the work in section 5.

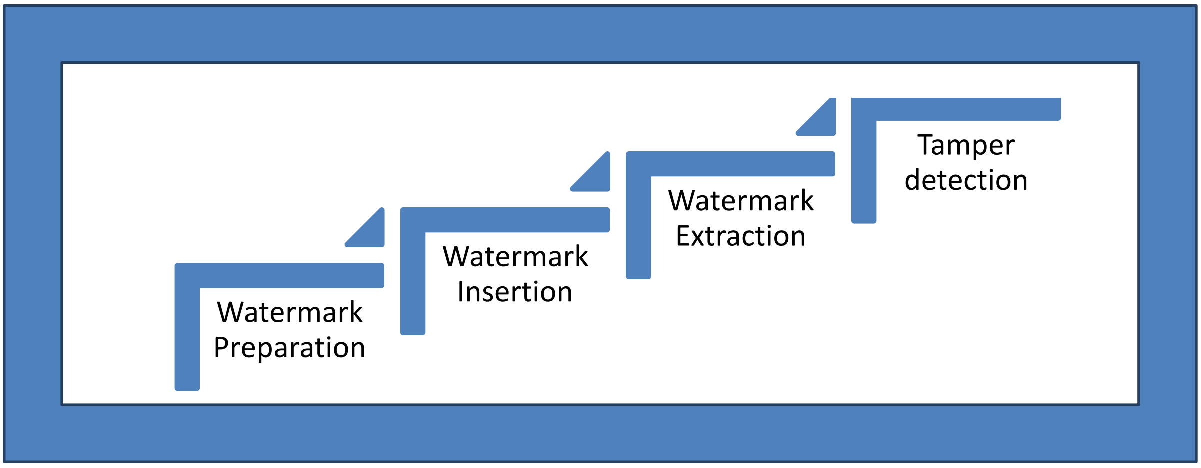

Fig.1. Phases of Fragile Watermarking Technique

II. Literature Survey

Relatively less works exist on fragile watermarking to protect databases from being tampered with. The watermark usually acts as a signature of the database. They are based on the principle that the even a minute change made to the database will immediately alter its signature and hence its watermark. The tampering effect can therefore be detected. We enumerate the fragile database watermarking techniques in table 1.

Li *et.al.* forthput a distortion-free method for categorical data. In their distortion-free technique, entire database relation is first divided among predefined number of groups, and then watermark bits are inserted. These bits are then independently tested in every group based on group hash values [3]. Digital encoding of the complete database is used to prepare watermark that acts as its signature. This is then inserted into the relation by reordering tuples. However, rearranging of tuples is considered to be a malicious alteration. This may result in alteration of watermark even though there is a change in original values in the database.

Later, Guo *et. al.* provided the solution to this problem by embedding the watermark bits into autonomous groups of tuples that are securely pre-decided[4]. Two different groups of watermarks are embedded into Least Significant Bits (LSBs) of the attributes of a tuple within a group using secure key. This results in localization of modifications made to the database. However, the technique produces considerable distortions as two LSBs per attribute for all the attributes of all the tuples are modified. Later, Khan *et.al.* proposed another fragile technique [5]. The core idea of their technique is to create watermark based on local characteristics of database relation like frequency distribution of digit, length and range. In [6], Khataeimaragheh *et.al.* proposed a similar technique that can identify and correct modifications by inserting watermarks generated from each attribute value. Thus, actual data can be recovered. However, the probability of accurately detecting, localizing and rectifying errors reduces drastically when the number of errors exceeds two.

Recently, Camara *et.al.* proposed a scheme which divides the database into vivid set of square matrices [7]. Watermark is created by determinant and minor of the square matrix. This watermark is then registered with trusted third party (TTP) for integrity verification instead of inserting into the database. Several other fragile watermarking techniques which have contributed to literature exist in [8-10].

Next, we found that the special databases called Decision Systems (DS) are handled recently. There is a deep ocean of applications of DS using rough sets. It is widely used in medicine for the treatment and diagnosis of various diseases, analyses of images such as X-ray etc. and analysis of medical data of patients [11-12]. Other fields include banking, economics, business, environmental case's studies, Information sciences, seasonal weather forecasting, decision sciences and many more [12]. Decision Systems (DS) are information systems having a decision attribute that categorize them into distinct classes. Other attributes in decision systems are called decision attributes. Authors have applied rough set theory for classificatory analysis of decision systems [13]. The classification and comparison of different database watermarking schemes is depicted in Table 1.

Table 1. Classification of various fragile watermarking techniques

Proposed Schemes	**Watermark Type**	**Target attribute**	**Contribution**
Li et.al.[3]	Created from Database	Categorical	Distortion-free watermarking
Guo et.al.[4]	Created from Database	Numeric	Embed the watermark in two LSBs of numeric attribute
Khan et.al[5]	Created from Database	--	Distortion-free, use of TTP
Khataeimaragheh *et.al* [6]	Created from Database	Numeric	Embed watermarks in LSB of attribute
Camara et.al. [7]	Created from Database	---	Distortion-free, use of TTP.
Khanduja et.al. [13]	Created from selected features of database	---	Integrity Protection of decision systems, use of rough set theory for integrity protection.

III. SECURITY ANALYSIS

Theoretical analysis of the fragile watermarking technique is done. This includes various cases where detection of modification is done. We now discuss them in detail:

3.1. Tamper Detection: At receiver end, databases are verified to detect temperedness. Watermark W_x is first extracted from the suspected database. This is then compared bit-wise with the actual watermark. This can be obtained by regenerating it from suspected database *W_R*. Let the original embedded watermark is *W_o*. We now consider the following cases:

i. Database is not tampered: If the shared/outsourced database is not modified then the data and embedded watermark will not change. Therefore, W_R=W_o and W_X=W_o. Thus, both watermarks will match (W_R==W_X).

ii. Database is tampered with, in such a way that watermark is not modified: If the data is altered then, watermark generated from database will be different from the original watermark. Since, the embedded watermark is not altered, watermark extracted will be same as original watermark. Thus, $W_X \neq W_R$, and integrity loss will be detected.

iii. Database is tampered with, in such a way that data is not modified: If the data is not altered then, watermark generated from database will be same as the original watermark. Since, the embedded watermark is altered, watermark extracted will be different from the original watermark. Thus, $W_X \neq W_R$, and integrity loss will be detected.

iv. Both the data and the inserted watermark were altered. This case has a very little chance of the two watermarks generated after the alterations turns out to be exactly the same. Hence $W_X \neq W_R$ and the tampering are detected.

3.2 False Hit Rate: It is defined as a probability of extracting original watermark from a non-watermarked database. Depending upon the technique, final watermark is generated by concatenating all the watermarks embedded within the database.

Let a watermark bit b_x be inserted N_x times and τ be the threshold that decides acceptable majority level. Each bit $b_x{}^*$ i.e. extracted from a non-watermarked relation has same probability of *0.5* to match or not match the original embedded bit in the watermark. We now take P_b as the probability that at least τ (*majority threshold*) portion out of N_x can be noticed from non-watermarked relation by sheer chance as [14]

$$P_b = \sum_{j=\tau*Nx}^{N_x} b\left(j; N_x, \frac{1}{2}\right) = B\big((\tau*N_x+1); N_x, 0.5\big) \qquad (1)$$

Where, $b(j;n,p)$ be the probability of getting j success in n Bernoulli trials with probability p for success and $1-p$ for failure [2].

$$b(k;n,p) = {}^{n}_{k}C * p^{k} * (1-p)^{n-k} \qquad (2)$$

For a watermark of length *Lw*, the false hit rate P_{FHR} is given by:

$$P_{FHR} = B\,(\tau_w; L_w, P_b) \qquad (3)$$

where τ_w is watermark length threshold on entire database, L_w is length of watermark, $B(j;\ n,p)$ is known as cumulative binomial probability exhibiting the probability of getting at least j success from n Bernoulli's trials. Hence, one can increase the chances of preventing a false hit by choosing larger values of τ as well as τ_w, as the false hit probability declines substantially.

IV. CONCLUSION

In this work, authors have analyzed the fragile watermarking techniques for web databases. The state-of-art for protecting Relational databases as well as Decision systems are discussed and analyzed how watermark is prepared and where it is inserted. The protection of Decision System uses Rough Sets theory to generate the watermark from the attributes to be protected. The watermark is then embedded inside the Decision System for protection against tampering. Theoretical analysis of the tamper detection is done analyzing various cases w.r.t. watermark extracted and re-generated watermark. False Hit Rate is also considered as a measure for security analysis.

Future work holds protection of other databases such as No SQL databases and Object-Oriented databases against ownership and tamper detection. Techniques should focus on strengthening the robustness for tackling ownership issues and utilizing watermark to carry useful information within the targeted database.

REFERENCES

[1] Khanduja, V. (2017). Database watermarking, a technological protective measure: Perspective, security analysis and future directions. *Journal of information security and applications*, *37*, 38-49.

[2] Khanduja, V., Chakraverty, S., & Verma, O. P. (2015). Watermarking categorical data: algorithm and robustness analysis. *Defence Science Journal*, *65*(3), 226-232

[3] Li Y., Guo H. and Jajodia S., “Tamper detection and localization for categorical data using fragile watermarks”, ACM workshop on Digital Rights Management, pp: 73-82, 2004.

[4] Guo H., Li Y., Lui and Jajodia S., “Fragile watermarking scheme for detecting malicious modifications of database”, Elsevier, Information Sciences, Vol. 176(10), pp: 1350–1378, 2006.

[5] Khan A., Husain S.A., “A fragile zero watermarking scheme to detect and characterize malicious modifications in database relations,” The Scientific World Journal, Vol. 2013, Article ID 796726, pp: 1-16, 2013.

[6] Khataeimaragheh H. and Rashidi H., “A Novel Watermarking Scheme for Detecting and Recovering Distortions in Database Tables”, International Journal of Database Management Systems, Vol 2(3), pp: 1-11, 2010.

[7] Camara L., Li J., Li R. and Xie W., “Distortion-Free Watermarking Approach for Relational Database Integrity Checking”, Hindawi Publishing Corporation, Mathematical Problems in Engineering, Vol. 2014, pp. 1-10, 2010, http://dx.doi.org/10.1155/2014/697165.

[8] Genin, E., & Franco-Contreras, J. (2019). Dynamic Watermarking-Based Integrity Protection of Homomorphically Encrypted Databases–Application to Outsourced Genetic Data. In *Digital Forensics and Watermarking: 17th International Workshop, IWDW 2018, Jeju Island, Korea, October 22–24, 2018, Proceedings* (p. 151). Springer.

[9] Abbasi, F., &Memon, N. A. (2018, April). Reversible Watermarking for the Security of Medical Image Databases. In *2018 21st Saudi Computer Society National Computer Conference (NCC)* (pp. 1-6). IEEE.

[10] Niyitegeka, D., Coatrieux, G., Bellafqira, R., Genin, E., & Franco-Contreras, J. (2018, October). Dynamic Watermarking-Based Integrity Protection of Homomorphically Encrypted Databases–Application to Outsourced Genetic Data. In *International Workshop on Digital Watermarking* (pp. 151-166). Springer, Cham.

[11] Abdel-Basset, M., & Mohamed, M. (2018). The role of single valued neutrosophic sets and rough sets in smart city: Imperfect and incomplete information systems. *Measurement*, *124*, 47-55.

[12] Pawlak, Z. (1997). Rough sets. In *Rough sets and data mining*(pp. 3-7). Springer, Boston, MA.

[13] Khanduja, V., & Chakraverty, S. (2018). Fragile Watermarking of Decision System Using Rough Set Theory. *Arabian Journal for Science and Engineering*, *43*(12), 7621-7633.