International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 24 (2017) pp. 15754-15758

An Invisible Watermarking Technique for Integrity and Right Protection of Relational Databases

Murugan R
Research Scholar, Research & Development Centre, Bharathiar University, Coimbatore, Tamilnadu, India and Associate Professor, MES College Marampally, Aluva, Kochi, India.

Orcid Id: 0000-0003-3137-8236

Jaseena K U
Assistant Professor, MES College Marampally, Aluva, Kochi, India.

John T Abraham
Assistant Professor, Bharata Mata College, Thrikkakara, Kochi, India.

Abstract

In recent years, everything is trending toward digitalization. With the rapid development of the Internet technologies and wide applications of databases, databases need to be transmitted conveniently over the network. Extensive use of the Internet coupled with the tremendous growth in database applications has created a huge demand for database security. This urges us to devise new data protection techniques to protect and secure the data with vital significance. In this respect, digital watermarking works as a tool for integrity, authentication and copyright protection of relational databases over the Internet.

In this paper we propose an invisible watermarking algorithm for the protection of relational databases. The watermark is logically embedded in the database and a watermark key is generated. Later the watermark is extracted to prove the identity. The proposed algorithm is suitable for any type of attribute that used to house numerical, non-numerical or categorical data. The experiments show that the new method is efficient as well as effective for authentication and right protection of relational databases.

Keywords: Digital Watermarking, Relational Databases, Security, Integrity, Authentication, Right Protection.

INTRODUCTION

The rapid development of Internet technology offers a wide range of web-based services, such as database as service, digital repositories and libraries, e-commerce, online support system, etc. The applications on the web make the digital content such as images, videos, audios, text, database, easily accessible to people around the world for sharing, purchasing, distribution, computation, etc. This results in challenges like piracy, illegal redistribution, ownership claiming, and forgery. Digital watermarking technology is an effective solution to meet such challenges [1]. Fraud and Tamper Detection is generally used for very critical applications such as commercial transactions or medical applications. It is important to ensure that the content was originated from a specific source and that it had not been changed, manipulated or falsified. This can be achieved by embedding a watermark in the underlying data of the database. Subsequently when the database is checked, the watermark is extracted using a unique key associated with the source and the integrity of the data can be verified with by using the extracted watermark.

The digital watermarking is a significant development and technique for inserting information into an image, text, video, audio or database. Later this can be extracted for a variety of purposes like integrity enforcement, authentication and right protection [2]. The wide utilization of database applications over Internet has created a crucial anxiety as far as ownership and copyright protection for relational databases are concerned. Recently, copyright protection of database is important in the field of information technology. This is partly because that the digital data is very easy to be tampered, peculated and illegally copied. Digital watermarking is an approach to cope with this situation. Protection from the piracy of digital assets is usually based upon the embedding of digital watermarks into the data. Watermarking relational database is a solution for ownership proofing, tamper detection, copy right protection, etc. In this paper, we have proposed an imperceptible, zero and robust watermarking algorithm for relational databases using combined image and text.

The rest of the paper is organized into different sections. The concept of digital watermarking and the constraints of using watermarking in relational databases are explained in DIGITAL WATERMARKING section. A comprehensive survey on this topic is presented in RELATED WORKS and the new algorithms proposed for embedding and extraction of watermark are illustrated in PROPOSED TECHNIQUE sections. The EXPERIMENTAL RESULTS describes the experiment, results and analysis of the proposed work. The last section concludes this paper with summaries.

DIGITAL WATERMARKING

A digital watermark is considered to be some kind of information that is embedded into a digital asset for tamper detection, ownership proof, traitor tracing, etc. [1]. A digital watermarking may be perceptible or imperceptible, depends on its visibility in the watermarked content [3][4]. Watermarking may be classified as image watermarking, video watermarking, audio watermarking, text watermarking, database watermarking based on the type of document to be watermarked.

Again, a watermark may be robust or fragile. In robust watermarking, the modification to the watermarked content will not affect the watermark and in fragile type, the watermark gets destroyed when the watermarked content is modified or tampered with. To insert watermark into a digital content, we can follow any one of the two techniques such as spatial-domain technique or transform-domain technique. On the basis of the requirements of data, watermarks can be classified into blind or informed. In process of extraction of watermark, the original content is not required in the case of blind watermarking. But in the latter case, the original content is required. The zero watermarking technique is a type of blind watermarking in which the original content is not modified while embedding the watermark.

Watermarking techniques used for text and multimedia could not be used for watermarking relational databases. The relational data differs from multimedia data in many respects:

Few Redundant Data:

Multimedia objects consists of large number of bits providing large cover to hide watermark, whereas the database object is a collection of independent objects, called tuples. The watermark has to be embedded into these tuples.

Out-of-Order Relational Data:

The relative spatial/temporal positions of different parts or components in multimedia objects do not change, whereas there is no ordering among the tuples in database relations as the collection of tuples is considered as set.

Frequent Updating:

Any portion of multimedia objects is not dropped or replaced normally, whereas tuples may be inserted, deleted, or updated during normal database operations[1]. Similarly, watermarking techniques for text typically exploit special properties of text formatting and semantics. For example, watermarks are often introduced by altering the spacing between words and lines of text [5].

RELATED WORKS

For integrity verification, authentication and copyright protection of relational database many watermarking techniques have been proposed. Agrawal et al. introduced a watermarking technique based on numeric data type attribute and marking is done at bit-level [6]. This technique used markers to locate tuples to hide watermark bits in the least significant bits.

Sion et al. introduced a watermark technique for numerical data [4]. This technique also dependent on a secret key which used the most significant bits of the normalized data set. In this technique, the data set is divided into partitions using markers and the partition statistics is varied to hide watermark bits.

Li et al. proposed a fragile scheme for tamper detection of categorical data [7]. In this scheme, the database relation is first divided into partitions in which a watermark is embedded by physically modifying the order of tuples. The scheme does not allow any legal update.

H. M. EL-Bakry and N. Mastorakis proposed a new approach for protecting the ownership of relational databases [8]. This approach is suitable for both textual and numeral data. The proposed technique used two different methods for generating secret function for both textual and numeric data.

An invisible zero watermarking algorithm for protecting text documents using combined image and text watermark is proposed [2][9]. Here the inherent properties text document is used and a watermark key is generated and this watermark key is utilized later in order to extract watermarks from the text document.

Ali Hamadou et al. proposed a novel method based on zero-watermarking approach for content integrity checking of database relations [11]. The proposed scheme does not change the content of the host data, thus resolving the inherent conflict between security and imperceptibility. Furthermore, there is no constraint on the type of attributes being watermarked.

Anuj Kumar Dwivedi et al. presented a detailed survey of database watermarking techniques with certain constraints and analyzed their strengths and weaknesses [12]. It also discussed various types of database attacks and narrated in detail the challenges of database watermarking. Ibrahim Kamel proposed a nonconventional scheme for R-tree data structures that does not change the values of the attributes [13].

In this paper, we have proposed an invisible zero watermarking algorithm for integrity and copyright protection of relational databases. This approach can be applied for both textual and numeral data. The proposed technique uses a single method for generating secret key for attributes having both textual and/or numeric data.

PROPOSED TECHNIQUE

To protect the relational databases, it is proposed to have a robust and zero watermarking algorithm utilizing combined image and text mark. A watermark that is a quite remarkable and unique logo or signature of a particular individual or an organization who owns the copyright of a digital asset is that what is suggested [10].

The relations in the relational databases can be altered in different ways. The ordering of tuples can be changed in many instances by different sort orders of tuples and even the ordering of attributes also can be changed. Sometimes this may damage the watermark information which we have inserted into the relation. That is, we may lost the watermark information simply without changing the values of the tuples in the relation. But we can take different kinds of measures to avoid the loss of watermark information due to the above said manipulations. In the case of relational databases, there are chances of tampering/alteration of data by way of addition, deletion or modification of tuples or even by changing one or more attribute values of one or more tuples. So we need to consider each attribute values in every tuples in the relation as an object for inserting watermark information, then only we can identify whether the database get altered/tampered. If we kept the watermark information in the relation, then the chances of loss of watermark are more. The proposed algorithm is inserting the watermark information in to the database in a logical way only, without performing a physical insertion.

In the proposed algorithm, the copyright of owner of the database is logically embedded as a mark and generated a watermark key. The watermark key along with the watermark is kept with the Certification Authority (CA), where the original owner is registered. Later the watermark is extracted from a suspected/received database using watermark key with the help of watermark extraction algorithm. After that both the original and extracted watermarks are compared to calculate the watermark accuracy in order to prove authenticity of the original owner and the integrity of the database.

The proposed watermarking process involves two phases,

1. Watermark embedding, and
2. Watermark extraction.

The Process of Watermark Embedding

In the database the watermark is embedded and this algorithm is called embedding algorithm. In fact the inputs for embedding are the original database plus the watermarks that is the original database, image and the text. The embedding process is depicted in Figure-1.

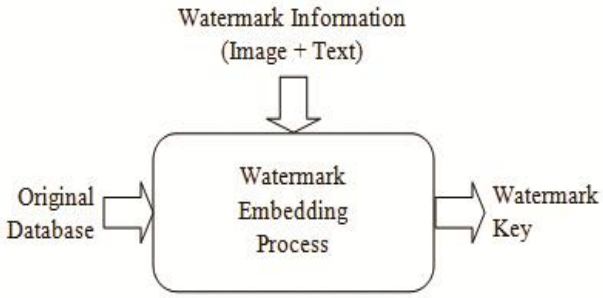

Figure-1: Process of Watermark Embedding

1. Inputs are database and combined image and text watermark.
2. For each field, generate a new attribute based on the data stored in other records in the available database using a secret function. The secret function is given as follows:

   $$\beta = [H(n, \alpha, \rho)] / n$$

   where, α is the numbers characters per attribute, ρ is the ASCII value of the character in the attribute, n is the number of tuples.
3. Thus an extra record has been created and stored in an array RDT (Relational Database table) in the form of alphanumeric characters.
4. The watermarks (watermarks obtained after text preprocessing and image preprocessing) and the RDT are given as input to the key generation algorithm, which is described in Section 4.3.
5. The key generation algorithm generates the watermark key, which is to be registered with a certification authority (CA), a trusted third party, along with the watermark for copyright protection.

The Process of Watermark Extraction

The extracting algorithm is something which extracts the watermark and that is why it is known as extracting algorithm. The watermark key is generated by giving the watermark (image and text) and the malicious database as inputs. Figure-2 shows the watermark extraction process. For tamper detection or right protection, the newly generated key is compared with the original key.

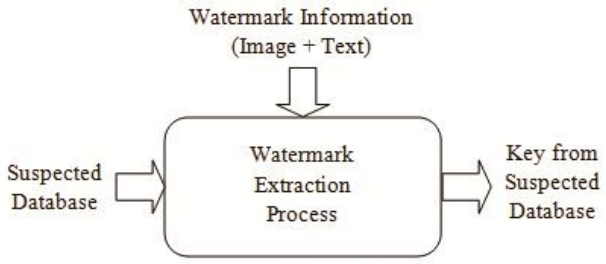

Figure-2: The Process of Watermark Extraction

1. The watermark and the received suspected database are the inputs.
2. For each field, generate a new attribute based on the data stored in other records in the available database using a function. The function β is given as follows:

   $$\beta = [H(n, \alpha, \rho)] / n$$

   where, α is the numbers characters per attribute, ρ is the ASCII value of the character in the attribute, n is the number of tuples.

3. Thus an extra record has been created and stored in an array RDT in the form of alphanumeric characters as in Embedding Process.
4. Using the generated RDT and the watermark, generate the new watermark key using key generation algorithm described in section 4.3.
5. Compare the newly generated watermark key with the original key to check the originality of the database.

Key Generation Algorithm

The algorithm used for generating watermark key is described below:

The watermark (W) is initially separated into image (WImg) and text (WTxt). WImg is then converted to alphabet and we obtain an alphabetical watermark (WT). The algorithm used for embedding watermark is presented below:

1. Input W, and RDT.
2. Split W into W_{Img} and $W_{Txt.}$
3. Preprocess W_{Img} and $W_{Txt.}$
4. Convert W_{Img} to text WT.
5. Merge WT and W_{Txt} to get W.
6. Compute watermark_length from W.
7. Generate watermark key using steps 8 to 11.
8. I=1, J=1.
9. While (I < watermark_length) repeat step 10 and 11.
10. If (W(I) = RDT(I)) Then

    Key (J) = 0

    Else

    Key (J) = 1, Key(J+1) = RDT(I) + 3
11. Increment I by 1, J by 2.
12. Output Key.

    W: watermark, W_{Img}: image watermark, W_{Txt}: text watermark, WT: text watermark generated from W_{Img}.

First the watermark is split into image and text watermarks. The preprocessing of text and image watermarks is done to make the watermark characters (W_{Txt} and WT). Using RDT and textual watermarks, the watermark key is generated using the algorithm described above.

EXPERIMENTAL RESULTS

We conducted and tested the algorithm of our proposed scheme using Matlab and Spread Sheet, on a workstation with a real life data, the Forest Cover Type dataset, available at http://kdd.ics.uci.edu/databases/covertype/covertype.html.

There are 581,102 tuples in the dataset, each with 10 integer attributes, 44 Boolean attributes, and 1 categorical attribute. A partial set of data values are given in Figure-3. For watermark construction, we used all types of attributes involved in the database.

	A	B	C	D	E	F	G	H	I	J	K	L	BB	BC
1	2596	51	3	258	0	510	221	232	148	6279	1	0	0	5
2	2590	56	2	212	-6	390	220	235	151	6225	1	0	0	5
3	2804	139	9	268	65	3180	234	238	135	6121	1	0	0	2
4	2785	155	18	242	118	3090	238	238	122	6211	1	0	0	2
5	2595	45	2	153	-1	391	220	234	150	6172	1	0	0	5
6	2579	132	6	300	-15	67	230	237	140	6031	1	0	0	2
7	2606	45	7	270	5	633	222	225	138	6256	1	0	0	5
8	2605	49	4	234	7	573	222	230	144	6228	1	0	0	5
9	2617	45	9	240	56	666	223	221	133	6244	1	0	0	5
10	2612	59	10	247	11	636	228	219	124	6230	1	0	0	5
11	2612	201	4	180	51	735	218	243	161	6222	1	0	0	5
12	2886	151	11	371	26	5253	234	240	136	4051	1	0	0	2

Figure-3: Partial Forest Cover Type dataset

For watermarking image we have selected an educational institutions logo and the preprocessing of the image before converting into text has been done as shown in Figure-4.

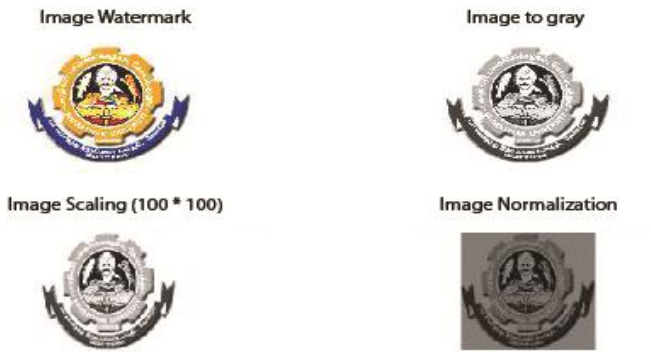

Figure-4: Image preprocessing in the proposed work

The key generated with the help of watermarking embedding process can be kept in the certifying authority for future authentication of the above database. Figure-5 depicts a partial form of key generated after the watermark embedding process. The original key and the key generated using the suspected database can be used for checking the integrity and the originality of the database.

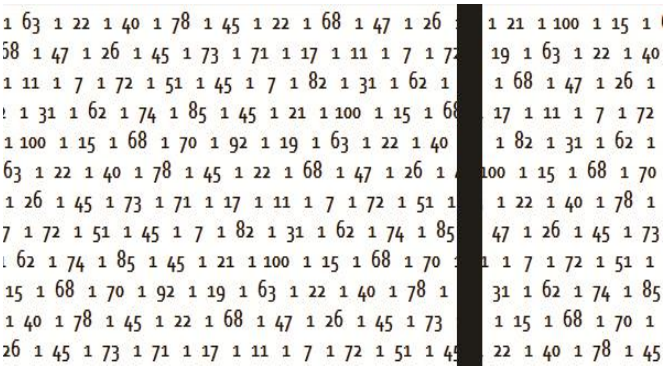

Figure 5: Partial key generated from the Forest Cover Type dataset and watermark information

In all types of database attack such as insertion attack, deletion attack, alteration attack and integrity checking and protecting copy right information, we proved the proposed method overcome the pitfalls found in other types watermarking methods. Most of the researchers embedded the watermark information in the attribute values of various tuples. These marks usually make changes in the data value and the database become useless.

Table I: Comparison of the proposed technique and the technique by Hazem M. El-Bakry and Nikos Mastorakis[8]

Items Compared	Proposed Technique	Compared Technique
Type of algorithm used	A single algorithm for all types of attributes.	Separate algorithm for numeric and character type attributes.
Insertion/Deletion/Alteration	100 % temper detection even small changes in the data values of any attribute. Ordering tuple will also take care.	Tamper detection is not achievable in all the cases. Reordering of tuples will not affect the key.
Synchronization Error	Not vulnerable to synch error, since the technique does not require a marked tuple in the detection process.	It requires a special marked tuple in the marked database.
Usability of data values	Maintaining 100% integrity in the watermarked database.	Interchange of data values of same attributes of two tuples will be change any difference.

CONCLUSION

Digital watermarking can be effectively utilized for integrity, authentication and copyright protection of relational databases and this has of great importance in research. In this paper, a new watermarking technique is specified that uses combined image and text watermark for the protection of relational databases. The proposed technique uses a unique method for digital watermarking to relational databases irrespective of the types values the database hold, that is textual, numeric, logical or categorical. Thus the proposed method is highly recommendable for relational databases with attributes having different data types.

REFERENCES

[1] Raju Halder, Shantanu Pal and Agostino Cortesi.(2010). Watermarking Techniques for Relational Databases: Survey, Classification and Comparison. In *Journal of Universal Computer Science,* vol. 16, no.21 (2010), 3164-3190.

[2] Z. Xiao-hua1, M. Hong-yun,L. Fang, "A New Kind of Efficient Fragile Watermarking Technique", *Acta Electronica Sinica,* 2004.

[3] Z. Jalil, A. M. Mirza, "A Review of Digital Watermarking Techniques for Text Documents", *IEEE,* 2009.

[4] Sion R., Atallah, M., and Prabhakar, S. (2005). Rights protection for categorical data. *IEEE Transactions on Knowledge and Data Engineering*, 17: pp. 912–926.

[5] N. Chotikakamthorn. Electronic Document Data Hiding Technique using Inter-character Space. In *Proceedings of the 1998 IEEE Asia-Pacific Conference on Circuits and Systems(IEEE APCCAS).* 1998, pp. 419-422.

[6] Agrawal, R. and Kiernan, J. (2002). Watermarking relational databases. In *Proceedings of the 28th international conference on Very Large Data Bases (VLDB '02)*, pp. 155–166, Hong Kong, China. VLDB Endowment.

[7] Li, Y., Guo, H., and Jajodia, S. (2004). Tamper detection and localization for categorical data using fragile watermarks. In *Proceedings of the 4th ACM workshop on Digital rights management (DRM '04)*, pp. 73–82, Washington, DC, USA. ACM Press.

[8] Hazem M. El-Bakry and Nikos Mastorakis.(2009). A New Watermark Approach for Protection of Databases. In *Proceedings of the 9th WSEAS International Conference on APPLIED INFORMATICS AND COMMUNICATIONS (AIC '09).*

[9] Jaseena K.U., and Anita John, *"Text Watermarking using Combined Image and Text for Authentication"*, In *International Journal of Computer Applications (0975 – 8887)*, Volume 20 – No.4, April 2011.

[10] Z. Jalil, A. M. Mirza, "An Invisible Text Watermarking Algorithm Using Image Watermark", *Innovations In Computing Science and Software Engineering,* 2010.

[11] Ali Hamadou, Xingming Sun, Lingyun Gao, Saeed A. Shah, "A Fragile Zero-Watermarking Technique for Authentication of Relational Databases", In *International Journal of Digital Content Technology and its Applications* Vol.5 No.5, May 2011.

[12] Anuj Kumar Dwivedi, Dr. B. K. Sharma, Dr. A. K. Vyas, "Watermarking Techniques for Ownership Protection of Relational Databases", In *International Journal of Emerging Technology and Advanced Engineering* Vol 4, Special Issue 1, February 2014.

[13] Ibrahim Kamel, "A schema for protecting the integrity of databases", In *Journal - Computers and Security,* Volume 28 Issue 7, October, 2009 pp. 698-709.