International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-7, July 2018 E-ISSN: 2347-2693

Relational Databases Watermarking Technique Based on Specific String Verification

Anuj Kumar Dwivedi[1*], B. K. Sharma[2], A. K. Vyas[3]

[1]CSE Department Ajay Kumar Garg Engineering College, Ghaziabad, India
[2]MCA Department Ajay Kumar Garg Engineering College, Ghaziabad, India
[3]E & T Department Jodhpur National University, Jodhpur, India

Available online at: www.ijcseonline.org

Accepted: 20/July/2018, Published: 31/July/2018

Abstract: With the speedy growth of web base environment and extensive requirement of databases in many important fields, the databases owners are need to be care full about how to protect and verify originality of the database. Digital watermarking is an effective solution to protecting the copyright of databases from illegal copying by using the inherent properties of relational databases. In this paper, we propose a relational database watermarking scheme that partition the whole database in groups by using the tuple hash value of the databases. After the group formation we incorporate the use of hash byte function over the group tuples along with owner id to form a group string. Similarly in detection process, a string will be formed to verify the suspicious watermarked database. The proposed scheme does not require any additional storage to store the verification information in the original database so the approach is robust against the any distortion in original database. Cast comparison experiments are conducted over watermark embedding and detection algorithms with the group size variation and the results are analyzed to show the running cost of the proposed scheme.

Keywords: ownership protection; group wise partition; tuple hash value; hashbyte; robustness

I. INTRODUCTION

Now a days, more and more data are sold and transmitted on the Internet. Copyright protection of owners of database is becoming more and more necessary with the rapid growth of the Internet and multi-dimensional development of digital multimedia contents, distribution of data become easier and easier day-by-day but along with this data may be compromised after unsafe transmission through the unsecure network. Digital watermarking provides an effective method of protecting digital data from illegal copying, and tempering by embed or delete operation directly into the data. Digital watermarking empowered the original owners to have some protection for their digital database contents by identifying copyright ownership by verification capability of genuineness [1]. Now the research of digital watermark technique focuses on the research on relational databases watermark technique. Increasing use of databases in applications is creating a need for protecting data copyright in databases [2, 3, 4], and the owners of relational databases worry about their data being pirated and about their ownership protection.

1.1 Technical challenges of database watermarking [6]

There are many differences between the structures of multimedia data and relational databases. Therefore, the watermarking process on relational database is challenged by the following factors:

i. **lack of redundant data**: A relational database is made up of tuples, each indicating an independent object. Therefore, watermarks basically have no places to hide or embed whereas multimedia object consists of a large number of bits with considerable redundancy. Thus, the watermark has a large cover in which to hide.

ii. **Out-of-order relational data**: Tuples of a relational database have no fixed location. This makes building a corresponding relation is very difficult in relational databases. However relative spatial/temporal positioning of various pieces of a multimedia object typically does not change.

iii. **Frequent updating**: Insertion, dropping, updating of operation of relational database is very frequent. Without malicious intention, users often casually drop some tuples or attributes. On the other hand, the pirate can add or substitute the tuples and attributes whereas, multimedia objects typically remain intact; portions of an object cannot be dropped or replaced arbitrarily without causing perceptual changes in the object. Because of these differences, techniques developed for multimedia data cannot be directly used for watermarking relations.

1.2 Requirements of database watermarking

Watermarking database has unique requirements that differ from those required for watermarking digital image and audio systems. The watermarked database must maintain the following properties [6]:

i. **Usability:** The amount of change in the database caused by the watermarking process should not result in degrading the database and making it useless. The amount of allowable change differs from one database to another, depending on the nature of stored records.

ii. **Robustness:** Watermarks embedded in the database should be robust against attacks to erase them. That is, the database watermarking algorithm must be developed in such a way to make it difficult for an adversary to remove or alter the watermark beyond detection without destroying usability of the database.

In this paper, we proposed an idea of protecting piracy of relational databases through group wise verification of specific strings during embedding watermark process group wise strings will be formed and during detection process, these strings will be verified.
The rest of the paper is organized as follows.
Section 2 simply describes the previous research schemes of this field. Section 3 specifies the watermarking algorithms based on owner id and tuple hash value. Section 4 gives a formal interpretation of the algorithms. Section 5 conducted some experiments and analyzed its robustness. Section 6 draw some conclusions.

II. PREVIOUS RESEARCH SCHEMES OF DATABASE WATERMARKING

In 2002, Agrawal and Kiernan first proposed a robust watermark for databases in. This method marks only numeric attributes and proposed the idea of watermarking using least significant bits (LSB). They do not account for multibit watermarks which make their technique vulnerable against simple attacks, for example, shifting of only one least significant bit results in loss of watermark [2]. It has some flaws that the embedded marks should be closely related to the primary key attribute of relational databases, the primary key attribute value could not be modified or replaced, or else, the scheme would have no meaning.

In 2006, a public key watermark was proposed in the aim of publicly ownership verification [7].

In 2008, Sun et al. introduced another technique for inserting an image into the database as watermark information. In this method, they converted one or more images into flow of bits. They used hash value of database tuple to find the location of each pixel and marked bit. They considered mod of hash value and watermark's length. If someone takes large image as watermark information, then length of watermark increases. And this method cannot insert all the pixels into the database. Therefore, this method is not efficient for small databases [5].

In 2010, a watermarking service system was presented to provide a secure framework for for data transaction and transfer via the internet [7].

In 2011, Min, Li and Wenyue, Zhao proposed an asymmetric watermarking scheme that employed the digital signature technology. A message can be signed by the owner with a private key. Anyone can verify this signature using the corresponding public key. Signature cannot be forged, that is signed message is indeed from the private key holder. [5`]
In 2012, U. Pratap et al. propose, a new technique of database which based on inserting the bits of a binary image in relational database. The proposed technique also minimizes the variation in watermarked database. Experimental results justify the feasibility of the proposed technique and its robustness against common database attacks [6].

We have already published a survey paper by summarizing and analyzing of above mentioned previous research work and after that by Combining the merits of above mentioned related research work, we have published an effective watermarking relational databases technique scheme based on embedded proportion and optimistic probability without using any specific database key. Now we are proposing a watermarking scheme that is not dependent on optimistic probability also we are not inserting any physical information into database for verification purpose so it will be free from any distortion.

III. PROPOSED WATERMARKING ALGORITHM

In our scheme, first decide how many groups depends on size of original database. Create groups based on tuple hash value in original database now each group has its own identity with set of tuples. Select tuples one by one in each group and generate watermark information and combined computed information and preserve it for verification purpose during detection process.
According to the idea, database grouping, the watermark embedding and detection algorithms are proposed in section 3.1 and section 3.2 respectively. And some notations which known only by the owners used in the algorithms are described in table 1.

Table 1 Abbreviations

Ai: the attribute, group information is based on A_i.value
h_i^r: tuple hash value used to manage group tupples

L: used to select the number of bits of A1. The length of the extracted characteristic bits
S: used to select the number of L bits of Ai. The L bits are chosen from the Sth position of the starting of the binary value of Ai
d: hash value with owner id for security purpose.
T_g: Total number of groups
m: group number
ES_m: m^{th} group embedding string
DS_m: m^{th} group detection string

3.1 Watermark Embedding Algorithm

The watermark embedding algorithm in details is described in Algorithm 1 and Algorithm 2. Here RDB denotes original relational database. The parameters L, S ,ES_m and OID are known only to the owner of the databases. In algorithm 1 describes that the first of all tuples are divided into tuple groups based on their tuple hash value. After that all tuples in each group are sorted according to their tuple has value. Algorithm 2 describes that how the each group is watermarked independently. If the value of numeric attribute *A*i is not null, we can compute watermark information. As step 5,6 and 7. Now to improve security attribute value and owner id hashed together and we convert the result into binary string and extract a predefined length of string for each tuple in each group and combined together to generate watermark information.

Algorithm1:
1) Input: (RDB);
2) for i =0 to n-1 do
3) $h_i^r = H(OID, A_i.value)$
4) $m = h_i^r \bmod T_g$
5) $r_i \epsilon g_m$
6) end for;

Figure 1: database-partitioning algorithm

Algorithm2:
1) Input: (RDB);
2) sort the group tuples as per tupple hash value;
3) while ($g_m \neq T_g$ -1) do
4) for each tuple ϵ group
5) if Ai .value with tuple is not null then
6) d = H (OID, hashbyte(tuple))
7) convert the integer value d to a binary sequence and note the value of sequence in Abi, i is an integer;
8) extract the L bits from Abi, Wi = BString(Abi, S, L);
9) convert Wi to IWi (integer value)
10) form a string by combining the digits of IW_i in ES_m
11) end;
12) End;

Figure 2: embedding algorithm

3.2 Watermark Detection Algorithm

The watermark detection algorithm is described in algorithm 3. Here, we verify the suspicious RDB group by group. The parameters L, S and OID have the same value used for the watermark embedding. In algorithm 2, first procedure is to create tuple wise groups as per their tuple hash value. Notes that the variable T_g must be equals the number of total groups in suspicious RDB. And the problem of watermark detection is judged by string comparison of ES_m and DS_m that are being generated during embedding and detection process.

Algorithm3:
1) Input: (suspicious RDB);
2) for i =0 to n-1 do
3) $h_i^r = H(OID, A_i.value)$
4) $m = h_i^r \bmod T_g$
5) $r_i \epsilon g_m$
6) end for;
7)) sort the group tuples as per tuple hash value;
8) while ($g_m \neq T_g$ -1) do
9) for each tuple ϵ group
10) if Ai .value with tuple is not null then
11) d = H (OID, hashbyte(tuple))
12) convert the integer value d to a binary sequence and note the value of sequence in Abi, i is an integer;
13) extract the L bits from Abi, Wi = BString(Abi, S, L);
14) convert Wi to IWi (integer value)
10) form the string by combining the digits of IWi in DS_m
11) if ES_m matched with DS_m
12) return true;
13) group has watermark
14) else
15) return false;
16) group has no watermark
17) end;
18) End;

Figure3: detection algorithm

The function BString(Abi, S, L) is used to extract a string as the watermark bits from *Abi*, the length of the section string is *L* and it is extracted from the Sth position of the string *Abi*., and noted as *Wi*. If the length of the string *Abi* is less than *L*, the empty position of *Abi* will be filled with '0'. For example, BString('11001010',5, 3), the extract function returns *Wi* equals '101'. And after that convert extracted bits to integer and embed with the identified weaker attribute [8].

IV. ALGORITHM ANALYSIS

In above proposed watermarking algorithms, the order of tuples in a group must be remain same as before and after watermarking. Also group formation strategy must remain same during the embedding and detection process. When DS_m is detected same as ES_m the detected ratio of the

matched group in RDB and suspicious group of RDB should remain 1 then we can predict that the watermark exists.

4.1 Robustness analysis

Robustness is the basic requirement of watermark technique. In this section we show that our watermarking scheme persist against moderation. There are four typical attacked modes include:

i. Subset deletion attack: In this type of attack, the attacker may take a subset of the tuples of the watermarked database and hope that the watermark will be removed.

ii. Subset addition attack: In this type of attack, the attacker adds a set of tuples to the original database. This is one of the most difficult attacks to defeat. The attacker may add some tuples to the watermarked table.

iii. Subset alteration attack: In this type of attack, the attacker alters the tuples of the database through operations such as linear transformation. The attacker hopes by doing so to erase the watermark from the database.

iv. Subset selection attack: In this type of attack, the attacker randomly selects and uses a subset of the original database that might still provide value for its intended purpose. The attacker hopes by doing so that the selected subset will not contain the watermark [4].

For our watermarking scheme, we considering subset alteration attack with different - different group sizes. In this experiments, we take size of relational databases be 1000, for the different size of groups, from table 2, we can see that the scheme is robust against subset modified attack.

Table 2 Detection of watermark under the subset modification attack

Group Size	Modified Proportion	Ratio of matched group	Existence of watermark
10	0	1	yes
	.05	0.15	no
	.25	0.14	no
	.5	0.12	no
	.75	0.10	no
20	0	1	yes
	.05	0.24	no
	.25	0.20	no
	.5	0.18	no
	.75	0.15	no
50	0	1	yes
	.05	0.39	no
	.25	0.35	no
	.5	0.27	no
	.75	0.18	no
75	0	1	yes
	.05	0.52	no
	.25	0.35	no
	.5	0.21	no
	.75	0.11	no

When the modified proportion is other than zero, the watermark cannot be detected from relational databases. However, when the embedded proportion is 0, the watermark can be detected from relational databases. Thus table 2 shows that the watermarking scheme is robust against any little change in original database.

4.2 Running Time Analysis

We test the running time when the average number of tuples in row groups is varied from 50 to 300.Experiment shows that when the average number of tuples in row groups is doubled, the time cost increases several seconds but feasible.. Similarly, in detection process, time is little bit higher than embedding process but running time during embedded and detection process is rational and acceptable. The comparison result is shown in figure 4. This experiment indicates that the proposed scheme is feasible for bigger size groups.

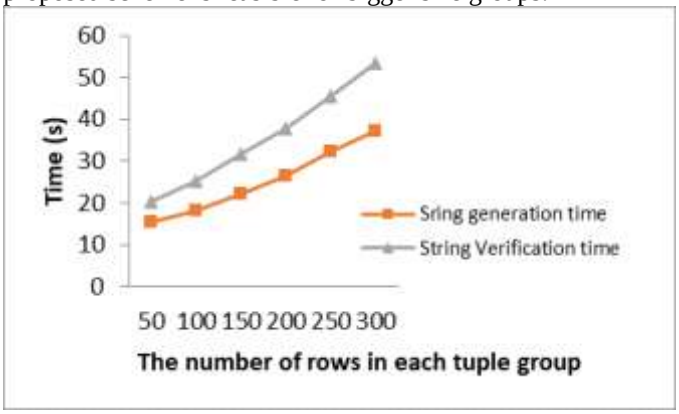

Figure 4 : running time comparison

V. CONCLUSIONS

In this paper, we have proposed a watermark technique for database authentication that is based on group-wise string verification. This paper is an extension of our previous work that was based on random embedded proportion value and optimistic probability. Proposed scheme involve owner id and topple hash value and hash byte function for improving security. In this scheme we have not use any specific key value from the database so it will difficult to correlate marked tuples and attribute. Based on the experimental analysis scheme is robust against any small change in original database. Running time experiment shows time growth in embedding and detection process is rational so our scheme can be used for large databases. Proposed scheme is distortion free and resilient because nothing is embedded in original database and can identify any slight modification in the database.

REFERENCES

[1] Saraju P. Mohanty , "*Digital Watermarking : A Tutorial Review*" Indian Institute of Science, Bangalore, 1999

[2] R. Agrawal, J. Kiernan. "Watermarking Relational Databases", *In: Proceeding of the 28th VLDB Conference.* Hong Kong, 2002: 155-166.

[3] G.H. Gamal, M.Z. Rashad and M.A. Mohamed *"A Simple Watermark Technique for Relational Databa"* Mansoura Journal for Computer Science and Information Systems Vol. 4, No.4, Jan2008.

[4] Raju Halder, Shantanu Pal, Agostino Cortesi *"Watermarking Techniques for Relational Databases: Survey, Classification and Comparison"* Journal of Universal Computer Science, Vol. 16, no.21 2010, pp.3165-3190

[5] Min, Li, Wenyue, Zhao, *"An Asymmetric Watermarking Scheme for Relational Database"*, Communication Software and Networks (ICCSN), IEEE 3rd International Conference. 2011,pp.180-184

[6] Udai Pratap Rao a, Dhiren R. Patel a, Punitkumar M. Vikani, **"***Relational Database Watermarking for Ownership Protection* 2nd International Conference on Communication, Computing & Security [ICCCS-2012] Science Direct pp.988-995.

[7] B. Wu, et aI., "Design and implementation of spatial data watermarking service system", Geo-spatial Information Science, vol.13, no. I, pp. 40-48, 2010.

[8] Anuj Kumar Dwivedi ,Dr. B. K. Sharma ,Dr. A. K. Vyas "Relational databases watermarking technique based on embedded proportion" International Education & Research Journal [IERJ] E-ISSN No : 2454-9916 | Volume : 3 | Issue : 6 | June 2017 pp 34-36

Author's profile

Mr.Anuj Kumar Dwivedi holding Master's Degree in IT and pursuing PhD in the field of watermarking technique for copyright protection of relational databases from Jodhpur National University, Jodhpur. He is working as Assistant Professor in the Department of Computer Science and Engineering in Ajay Kumar Garg Engineering College Ghaziabad. His area of specialization are data structure and algorithms.

Dr. B. K. Sharma is a Professor, HOD MCA and Dean Hostel of Ajay Kumar Garg Engineering College, Ghaziabad. He has obtained his MCA degree from JNU,New Delhi, M.Tech. from Guru Gobind Singh Indraprastha University, Delhiand Ph.D. from Shobhit University, Meerut. His areas of specialization are Software Watermarking, Discrete Mathematics, Theory of Computation and Compiler Design. During his career of more than decade in the teaching, he has published many Research papers in International/National Journals/Conferences. He has also published many books for engineering students

