International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol.3, No.1, February 2013

WATERMARKING RELATIONAL DATABASES

Mayuree K.Rathva[1],Prof.G.J.Sahani[2]

Dept.of Computer Engineering, SVIT,Vasad,
Anand,Gujarat, INDIA
[1]mayurirathva@gmail.com and [2]Gurcharan_sahani@yahoo.com

ABSTRACT

Uptil now most of the work is done on image, video, audio.but today the Database watermarking becomes the research topic because of the increasing the use of relational database systems. Which deals with the legal issue of copyright protection of database systems. .Watermarking is an information hiding technique which is used to embed a mark within some host content. In this paper,we have focused on the review of four relational database watermarking techniques proposed by researchers [R.Agarwal & Jerry Kiernan, ZHU Qin, Brijesh B. Mehta, A.Odeh and A. Al-Haj].

KEYWORDS

Database Security, Database Watermarking, Multi-place Watermarking, Copyright Protection.

1. INTRODUCTION

Watermarking was introduced for image processing and then it extended for security of text and multimedia data.Now days, it is also used for software and database.

There are two phases of the Database Watermarking Techniques.

Watermark Embedding and Watermark Verification. During *watermark embedding* phase, a private key K (known only to the owner) is used to embed the watermark W into the original database. The watermarked database is then made publicly available. To verify the ownership of a suspicious database, the verification process is performed where the suspicious database is taken as input and by using the private key K the embedded watermark is extracted and compared with the original watermark information[5].

The novel approach for the copyright protection for outsourced database for the watermarking,In which watermark is embedded into the secret key and the chaotic random series.for the detection process there is no need of original database and to judge the copyright,we can match the watermark rate.

Due to differences between multimedia and database we cannot directly use any of the technique as it is for database, which developed for multimedia data. These differences include[3][5] :

- A multimedia object consists of a large no. of bits hence,there is a large amount of space available to hide the watermark.while Database consist of the tuples,each tuple represents a separate object.so,the watermark is spread over these separate objects.

DOI : 10.5121/ijcsea.2013.3107

- The change in a relative spatial/temporal positioning of multimedia object remains unchanged.while in case of database ,updates in database may changes the tuples.
- Drop or Replace operation is not possible in multimedia object without causing perceptual changes in the object.While, tuples may simply be dropped by delete operation in database.

Watermark is an open problem that aimed to one goal. This goal is how to insert [error/ mark/ data/ formula/ evidence/ so on] associated with a secret key known only by the data owner in order to prove the ownership of the data without lossless of its quality.

Agrawal et al. introduce a watermarking technique for numerical data [1]. This technique dependent on a secret key, uses markers to locate tuples to hide watermark bits, hides watermark bits in the least significant bits.

2. LITERATURE SURVEY

The security of relational databases has been a great concern since the expanded use of these data over the Internet. Because data allow unlimited number of copies of an "original" without any quality loss and can also be easily distributed and forged[1]. Hence, Digital watermarking for relational databases emerged as a candidate solution to provide copyright protection, tamper detection, traitor tracing and maintaining integrity of relational data

The characteristics for watermarking relational databases are **:**

Detectability:-The owner of the database should be able to detect the watermark by examining the tuples from the suspicious database[6].

Robustness:**-**The capability of the watermarking scheme to survive deliberate (for example, modifying, adding, deleting part of the data) and unintentional attacks (for example, digital reproduction and photocopying). The watermark should be detectable even in an object modifed by the attacker.

Capacity*:***-**It is the maximum amount of data that can be embedded and the optimal way to embed and extract this information.Private key selection: According to Kirchhoff's, the method used for inserting the watermark is public. To protect the watermark from the intruder, the private key should be selected properly.

Updatability:-The watermark algorithm should be such that either the tuples of the relational database are inserted or deleted, the watermark value should not be changed.

The watermarked database may suffer from various types of intentional and unintentional attacks which may damage or erase the watermark, as described below**:**

1. Benign Update:- The tuples or data of any watermarked relation are processed as usual. As a result, the marked tuples may be added, deleted or updated which may remove the embedded watermark or may cause the embedded watermark undetectable[1].

2 Deletion attack:-The Attacker deletes marked tuples from the relational database which leads to synchronization errors .

3 Alteration attack:-Attacker alters the data values of the tuples which leads to disturbance in the watermark. Altering the data values violates the usability constraints and makes the data useless .

4 Insertion attack:-Attacker inserts tuples to the data set hoping to disturb the embedded watermark which results in synchronization errors.

Brijesh B. Mehta used the same watermark in two attributes using our proposed algorithms. Therefore, it will be difficult for attacker to remove both watermarks from the database, based on extracted bits from both algorithms we can generate original watermark very easily, and we can prove ownership of database[3].In this approach binary image is used as watermark.The whole procedure of embedding and extraction of watermark is performing in two phases, in first phase we insert watermark in the numeric field of the database and in second phase we insert a watermark in the seconds field of database same way at the time of extraction we follow the reverse order of above phases.

3. COMPARITIVE ANALYSIS

3.1. Algorithm proposed by R.Agarwal and J Kiernan for watermark insertion and detection[1]

Firstly, Rakesh Agrawal & Jerry Kiernan have proposed an algorithm for database watermarking based on primary key and private(secrete) key. They are inserting a single bit watermark into the numeric field of database and then detecting it with the help of detection algorithm.In this Database watermarking techniques Database relations that can be watermarked have attributes which are such that changes in a few values do not affect the application.

Following are the notation used in the algorithm:

Ƞ - Number of relations in the tuple
ʋ - Number of attributes available in the relation for marking
ζ – Number of least significant bits available for marking in an attribute
1/ϒ – Fraction of tuples marked
Ꮚ – Number of tuples marked
α - Significance level of test for detecting a watermark
Ʈ – Minimum number of correctly marked tuples needed for detection.
.
// the private key κ is known only to the owner of the database.
// the parameters ϒ, ʋ, and ζ are also private to the owner.
1) foreach tuple r ϵ R do
2) if (Ƒ(r.P)mod ϒ equals 0) then //mark this tuple
3) attribute_index i= Ƒ(r.P)mod ʋ //mark attribute A_i
4) bit_index j = Ƒ(r.P) mod ζ //mark j_{th} bit
5) r.A_i = mark(r.P, r.A_i,j)
6) mark(primary_key pk, number ʋ, bit_index j) return number
7) first_hash = Ꜧ(κ○pk)
8) if(first_hash is even) then
9) set the j_{th} least significant bit of ʋ to 0
10) else
11) set the j_{th} least significant bit of ʋ to 1
12) return ʋ

Fig.1 (a1) Watermark insertion algorithm

// κ, ϒ, ʋ, and ζ have the same values used for watermark insertion. // α is the test significance level that the detector preselects.

1) totalcount=matchcount=0
2) foreach tuple s ϵ S do
3) if (Ƒ(s.P)mod ϒ equals 0) then //this tuple was marked
4) attribute_index i= Ƒ(s.P)mod ʋ //attribute Ai was marked
5) bit_index j = Ƒ(s.P) mod ζ //jth bit was marked
6) totalcount=totalcount+1
7) matchcount=matchcount+match(s.P, s.Ai,j)
8) Ʈ = threshold(totalcount, α)
9) if (matchcount >= τ) then suspect piracy
10) match(primary_key pk, number ʋ, bit_index j) return int
11) first_hash = Ꜧ(κ○pk)
12) if(first_hash is even) then
13) return 1 if the jth least significant bit of ʋ is 0 else return 0
14) else
15) return 0 if the jth least significant bit of ʋ is 1 else return 1

Fig.1(a2) Watermark detection algorithm

Result:

Currency Code=P	Nation	Buying Rate	Selling Rate
GUJ	Gujarat	239	167
MHR	Maharastra	1678	1223
PUNE	Pune	2678	1199
JAMMU	JammuKashmir	4430	4210
UPD	Uttarpradesh	199	109

Table 1(a): Foreign exchange rates

Currency Code=P	Nation	Buying Rate	Selling Rate
GUJ	Gujarat	239	167
MHR	Maharastra	1678	1223
PUNE	Pune	2678	1199
JAMMU	JammuKashmir	4530	4310
UPD	Uttarpradesh	199	109

Table 1(b): Foreign exchange rates (watermarked)

Currency Code=P	Nation	Buying Rate	Selling Rate
guj	Gujarat	239	167
mhr	Maharastra	1678	1223
pune	Pune	2678	1199
jammu	JammuKashmir	4530	4310
udp	Uttarpradesh	199	109

Table 1(c): Table with modified primary key

3.2. Watermark embedding and detection technique proposed ZHU Qin, YANG Ying, LE Jia-jin, LUO Yishu[2]

- Method proposed by these authors is almost same as the method proposed by Rakesh Agrawal & Jerry Kiernan.
- The basic difference is about primary key attribute in first approach primary key attribute has been used to generate watermark where as in this approach chaotic random number is used which has been generated from that secrete (private) key and primary key of database.
- So, the capacity of the watermark in first approach is limited and the scheme is not suitable for database which needs frequent updating, because it is very expensive to re-embed watermark into the updated database.

The scheme of a database relation is $R(P, A0, \ldots, Av\text{-}1)$, where P is the primary key attribute
The steps of embedding watermark are outlined as follows:

(1) Input the value of gap γ, which is the interval number between two adjacent marked tuples;
(2) Compute the value of ξj, which determine the number of bits of LSB of Aj for embedding according to the data range and the precision of Aj;
(3) For each tuple ri, repeat:
(3.a) Compute the bits conjunction of the secret key and the primary key ($K \cdot Pi$), and normalize it, s.t. $0 <$ NRM ($K \cdot Pi$) < 1;
(3.b) if LGS (NRM ($K \cdot Pi$)) mod $\gamma = 0$, then
(3.b.1) j = (NXT (LGS (NRM ($K \cdot Pi$))) mod v) +1;
(3.b.2) k = NXT (LGS (NRM ($K \cdot Pi$))) mod ξj ;
(3.b.3) $MRKij$= (NXT (LGS (NRM ($K \cdot Pi$))) mod γ;
(3.b.4) $ri.Aj$ = WTR ($ri.Aj$, $MRKij$, k).

Fig.2 (b1) Watermark embedding algorithm

- The watermark detection is the reverse process of the embedment, and its steps are similar to the embedding algorithm. For each tuple, the algorithm firstly computes the mark value that determined by the algorithm of watermark embedment.
- Then compares the computed value with the read value, and counts the matched bits. *CNT* denotes the total matched bits.
- The count of all the marked bits of the database that has not suffered from watermark attack should be n/ γ , so the watermark match rate equals CNT/ (n/γ).

Fig.2 (b2) Watermark detection algorithm

3.3. Algorithm proposed by Brijesh B.Mehta ,Udai Pratap Rao for watermark insertion and detection[3]

In this paper author has introdued a new approach,in which watermark was embedded at two different places.They have used image as watermark ,the first watermark was embedded in the numeric attribute of the tuple and second in the date attribute's time field.

The algorithm to insert a watermark.

1.Concatenation of the value of watermark bit and k2

2.Find the decimal equivalent of the string
3.Embed the decimal number in tuples selected by the
pre-defined key k1 as follows:
3.1.For each selected tuple do
3.2. For each selected Time attribute do
3.3. If the ‘MM’ field of the ‘Time’ mode k1 = 0
3.4. Embed the decimal number in SS field
3.5. Else Next attribute
3.6. End if
3.7. End loop
3.8.End loop

Fig.3 (c1) Watermark insertion algorithm

Instead of generating binary image from that binary equivalent of the extracted watermark. We are taking only LSB of the binary data.

1.Extract the decimal number in tuple selected by the predefined
key k1 as follows:
1.1.For each selected tuple do
1.2. For each selected ‘Time’ attribute do
1.3. If the ‘MM’ field of the ‘time’ mode k1 = 0
1.4. Extract the decimal number from SS field
1.5. Else Next attribute
1.6. End if
1.7. End loop
1.8.End loop

2. Find the binary equivalent of the extracted decimal number
3. Extract last bit (LSB) from is which indicate our watermark.

Fig.3(c2) Watermark detection algorithm

They have considered few of the major attacks applied on Database Watermarking like:

- Subset addition attack
- Subset deletion attack
- Subset selection attack
- Subset alteration attack

They have inserted same watermark at different places so, there is a less chances of it to get attacked and if so, it is comparatively easy to extract the original watermark because of the watermark is embedded at two places. We can have one correctly extracted watermark from that two places.

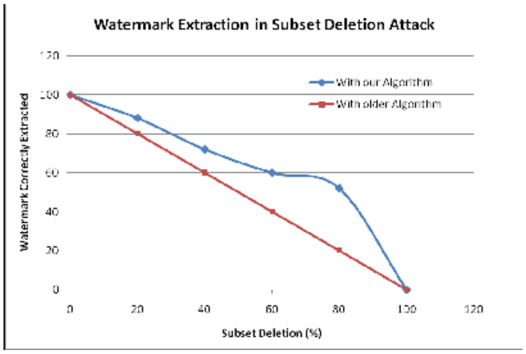

Fig.4:Robustness Results due to the Subset Deletion Attack

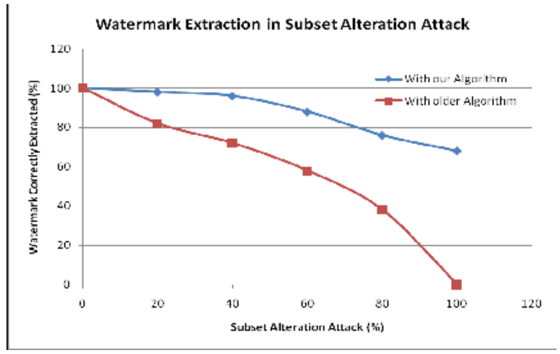

Fig.5:Robustness Results due to the Subset Alteration Attack

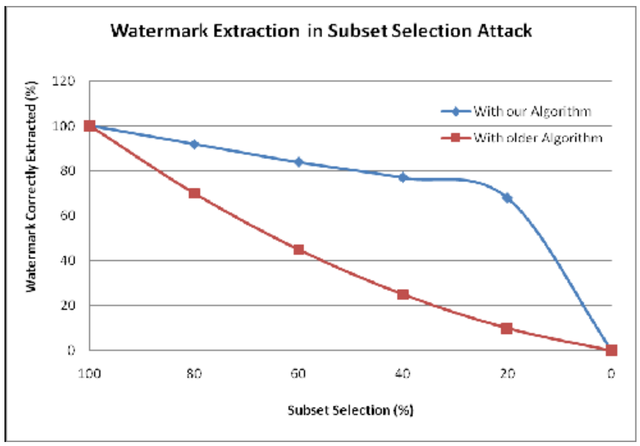

Fig. 6: Robustness Results due to the Subset selection Attack

Though it may be more costlier in a context of data correctness as it changes many attributes of a tuple in database but if you are thinking about robust copyright protection then you need to pay this price.

3.4. Algorithm proposed by Ashraf Odeh and Ali Al-Haj[4]

Approach given by the Ashraf Odeh and Ali Al-Haj, hide the watermark information (bits) in the mostly unnoticed 'time' attribute of the database tuples.

In database DATE field has two parts „Date" and „Time", here to hide binary information of watermark (SS) field is used from the time fields(HH:MM:SS).

- This approach is based on Image means a binary image has been used as watermark.
- A major advantage of using the time attribute is the large bit-capacity available for hiding the watermark, and thus large watermarks can be easily hidden, if required.

3.4.1. Watermark Embedding

- Transfer the image into a flow of bits.
- Group every 5 bits as a binary string.
- Find the decimal equivalent of the string
- Embed the decimal number in tuples selected by the pre-defined key '***K***' as follows:

```
for each selected tuple do
      for each selected 'Time' attribute do
            if the 'SS' field of the 'time' mode K = 0
                  embed the decimal number
            else Next attribute
            end if
      end loop
end loop
```

3.4.2. Watermark Extraction

- Watermark extraction requires neither the knowledge of the original un-watermarked database nor the watermark itself.
- The watermark extraction procedure is a direct reversal of the watermark embedding procedure as described below:

Extract the decimal number in tuples selected by the pre-defined key '***K***' as follows:

```
for each selected tuple do
      for each selected 'Time' attribute do
            if the 'SS' field of the 'time' mode K = 0
               extract the decimal number
            else Next attribute
            end if
      end loop
end loop
```

o Find the binary equivalent of the extracted decimal number.

o Group every 5 bits as a binary string.

o Reconstruct the binary image watermark from the binary strings.

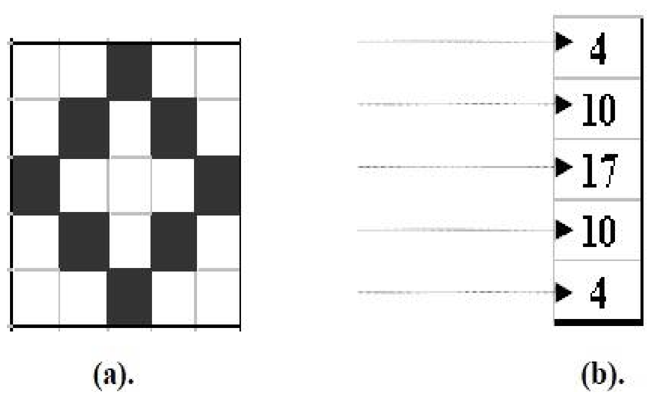

Fig.7: (a) Binary Image Watermark, and (b) its decimal equivalent Vector

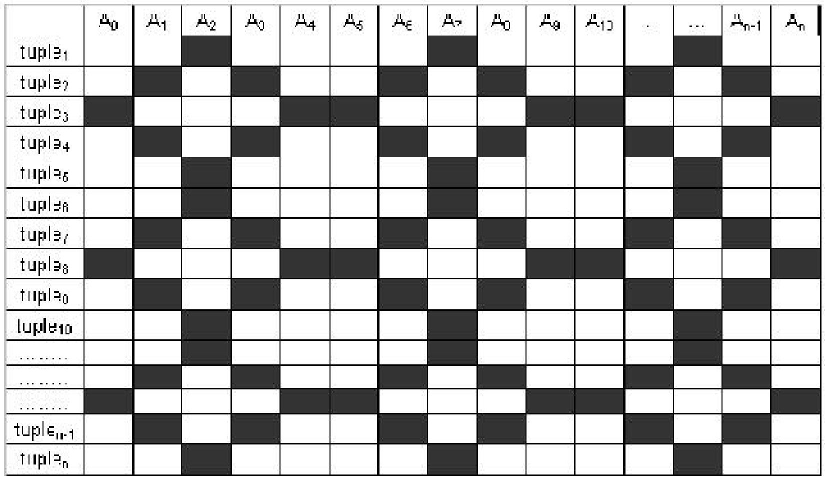

Fig.8: A snapshot of watermarked database

CONCLUSIONS

In this paper, we reviewed four papers proposed by different authors on watermarking relational databases that embeds the watermark bits in the database set by partitioning it. Every author worked for the robustness of the technique.*Agrawal & Kiernan*[1], *ZHU Qin, YANG Ying* [2] used watermarking for numeric data.while *Brijesh Mehta* [3] has applied watermarking for numerical and non numerical data.

It should be noted that the technique of watermarking [2]database is seldom applied for real-time protection of database copyright, neither used to prevent the pirates from illegal copy in server side directly The algorithm proposed by Brijesh Mehta which one is more secure approach because of same watermark was applied at two different places.

ACKNOWLEDGEMENTS

I am heartily thankful to my guide, *Prof. G.J.Sahani,* who encourage me and provide me necessary guidance.and I also want to thank my family and friends to help and support me.

REFERENCES

[1] R. Agrawal and J. Kiernan." Watermarking relational databases". In Proceedings of The 28th International Conference on Very Large Databases VLDB, 2002.
[2] ZHU Qin, YANG Ying, LE Jia-jin, LUO Yishu,(2006)"Watermark based Copyright Protection of Outsourced Database," IEEE, IDEAS, pp. 1-5.
[3] Brijesh B. Mehta, Udai Pratap Rao," A Novel approach as Multi-place Watermarking For Security in Database"Dept. of Computer Engineering, S. V. National Institute of Technology, Surat, Gujarat
[4] A. Odeh and A. Al-Haj "Watermarking Relational Database Systems", IEEE, pp. 270-274, 2008.
[5] Raju Halder, Shantanu Pal, Agostino Cortesi,"(2010) Watermarking Techniques for Relational Databases: Survey, Classification and Comparison" Journal of Universal Computer Science, vol. 16, no. 21, 3164-3190.
[6] Prof.Bhawana Ahire, Prof. Neeta Deshpande," Watermarking relational databases: A Review",IOSR Journal of Engineering (IOSRJEN) ISSN: 2250-3021 Volume 2, Issue 8.