ISSN 2320 -2602

International Journal of Advances in Computer Science and Technology (IJACST), Vol.3 , No.11, Pages : 30-34 (2014)
Special Issue of ICCECT 2014 - Held during 01-02 December, 2014,Bangkok, Thailand

Database Copyright Protection using Watermarking Arabic-Characters

Shadi R. Masadeh, Ashraf Odeh, Ayad Zobaydi

Faculty of Information Technology
Isra University
P.O. Box 22, Al-Isra University Postoffice Amman, Jordan 11622
Shadi.almasadeh@iu.edu.jo, Ashraf.odeh@iu.edu.jo, Alzobaydi_ayad@iu.edu.jo

Abstract- In this paper we describe a database watermarking algorithm for Arabic character attributes. Many Arabic characters are expandable; and thuds watermark bits can be hidden in their extensions, without sacrificing the readability and appearance of the character. The experimental Result show that the proposed algorithm yield watermark which is robust to various Arabic database attack.

Keyword: watermark, Arabic-Characters attribute, Copyright

1. INTRODUCTION

Watermarking relational databases is an emerging research area that deals with the legal issue of copyright protection of relation databases. There are few published methods developed exclusively to the problem of watermarking relational databases Research literature in this area is very limited, and reported results are insufficient. All published methods have dealt with one type of relational database which is numeric databases. That is, there was always an assumption that the data base consisted of tuples that have only numeric attributes. However, from our experience, practical relational databases have numeric and nonnumeric attributes, and therefore we propose here the development of a relational database watermarking method that hides watermark information in the nonnumeric attributes; that is, the character-based attributes.

Watermarking database has sole condition that differ from those required for watermarking digital audio or video products such as Maintaining Watermarked Database, Preserving Database Semantics, Preserving Database Structure, Robustness, Updatability

Watermarking process on relational database is challenged such as Data redundancy fewness, Relational data out-of-order, Frequent updating Silberschaty et al.(2006).

The proposed algorithm based on character attribute due to embed watermark (WM) in all Arabic character attribute spread of whole tuples of database ,we used the space-based watermarking is the large bit-capacity available for hiding the watermark. This facilitates embedding large watermarks or multiple small watermarks, which can be used to hide bits without being subjected to removal or destruction
The experimental results show that new algorithm is robust against database attacks and can prove then ownership of database.

2. RELATED WORK

We described a watermarking algorithm based on hiding watermark bits in the extensions of expandable Arabic characters of non-numeric, multi-word, attributes of subsets of tuples. A major advantage of using this approach is the large bit-capacity available to hide large watermarks. This is opposite to the other proposed algorithms where watermark bits have limited potential bit-locations that can be used to hide them effectively without being subjected to removal or destruction. The robustness of the proposed algorithm was verified against a number of database attacks such subset deletion, subset addition, subset alteration and subset selection attacks.

Huang et al. (2004) studied digital watermarking technology of numeric attribute and analyzed current researches on this issue and purpose new watermarking mechanism using voting method which pays much attention to data's usability through relational structure as well as makes marked bits robust to various attacks.

As an embranchment of information hiding, the digital watermark techniques have been attracting more and more interests in both research and industrial fields. As a tool for storing and managing data, relational database is widely used in many information systems. It is a crucial issue to protect the copyright of relational data. In order to make watermarking information more intuitive and easier to identify, Zhang et al. (2003,2004) propose a novel watermarking method, which embeds an image watermark into relational database. Experimental results verify the effectiveness of the proposed method

ISSN 2320 -2602

International Journal of Advances in Computer Science and Technology (IJACST), Vol.3 , No.11, Pages : 30-34 (2014)
Special Issue of ICCECT 2014 - Held during 01-02 December, 2014,Bangkok, Thailand

Li et al. (2006,2003,2005) a robust scheme for watermarking relational databases. Compared with the watermarking scheme recently proposed by Agrawal and Kiernan "AK", it is more robust in the sense that it provides an upper bound for the probability that a valid watermark is detected from unmarked data, and for the probability that a fictitious secret key is discovered from pirated data. The authors in this paper applied an alternate scheme which as they say more robust against the attacks that change the size of data base relations.

Sion et al;. (2004, 2006, 2003) presents a robust scheme for watermarking relational databases. It is more robust in the sense that it provides an upper bound for the probability that a valid watermark is detected from unmarked data, and for the probability that a fictitious secret key is discovered from pirated data.
And applied an alternate scheme which as they say more robust against the attacks that change the size of data base relations.

Muhammad Addul Qadir (2006), Introduces a watermak-encodindg scheme. They developed and implement new watermark-encoding scheme and have implemented the system for plain text documents. They encode the information in the existing characters of the text in an intelligent way that does not change the document. Moreover, the hidden information is being preserved by the document.

3. THE PROPOSED ALGORITHM

In this paper describes a database watermarking algorithm for Arabic character attributes. Many Arabic characters are expandable; and thuds watermark bits can be hidden in their extensions, without sacrificing the readability and appearance of the character. For example (one) can be written (واحد) or in extended form (واحـــد) . Both have the same meaning but the second can carry binary information. In the algorithm, a binary image is used to watermark the relational database. The bits of the image are segmented into short binary strings that are encoded in non-numeric, multi-word attributes of selected tuples of the database. The embedding process of each short string is based on expanding the first character of a word whose location is determined by the decimal equivalent of the short string. Extraction of a short string is done locating the word in which one of its characters was expanded. The image watermark is then constructed by converting the decimals into binary strings. A major advantage of using the space-based watermarking is the large bit-capacity available for hiding the watermark. This facilitates embedding large watermarks or multiple small watermarks. This is in contrast to bit-based algorithms where watermark bits have limited potential locations that can be used to hide bits without being subjected to removal or destruction. The proposed algorithm has two procedures: watermark embedding procedure and watermark extraction procedure. The two procedures are described in the following sub-sections.

3.1 WATERMARK BITS INSERTION PROCEDURE

The watermark embedding procedure consists of the following operational steps:

Step 1: Arrange the watermark image into m strings each of n bits length.
Step 2: Divide the database logically into sub-sets of tuples. A sub-set has m tuples.
Step 3: Embed the m short stings of the watermark image into each m-tuple sub-set.
Step 4: Embed the n-bit binary string in the corresponding tuple of a sub-set as follows:
Find the decimal equivalent of the string. Let the decimal equivalent be d.
Embed the decimal number d in a pre-selected non-numeric, multi-word attribute by expanding the first expandable character of the d th word of the attribute.
Step 5: Repeat step 4 for each tuple in the subset.
Step 6: Repeat steps 4 and 5 for each subset of the database under watermarking.

An illustration of embedding the binary watermark into a sub-set of tuples is shown in Figure 6.1. The watermark is a of 3 x 3 binary image. Each of the three 3-bit binary strings is transformed into its decimal equivalent as shown in Figure1(a), and embedded in the 3-tuple sub-set, as shown in Figure 1(b),. The count of the word with the red-colored character-extension (-) indicates the decimal equivalent of the embedded short binary string. Extension is performed on the fist expandable character of the word.

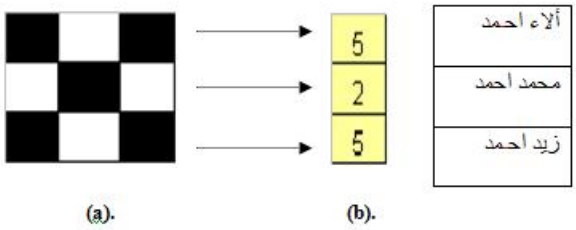

Figure 1 (a). Binary image watermark, and its decimal equivalent vector, (b).

Watermarking example; where subscripts represent space number, and (-) corresponds to expanded character.
A snapshot of the relational database after embedding the watermark throughout the database is shown in Figure 2. The tuples in the figure constitute the database, and the

ISSN 2320 -2602

International Journal of Advances in Computer Science and Technology (IJACST), Vol.3 , No.11, Pages : 30-34 (2014)
Special Issue of ICCECT 2014 - Held during 01-02 December, 2014,Bangkok, Thailand

A's are the watermarked non-numeric, multi-word attributes for each tuple.

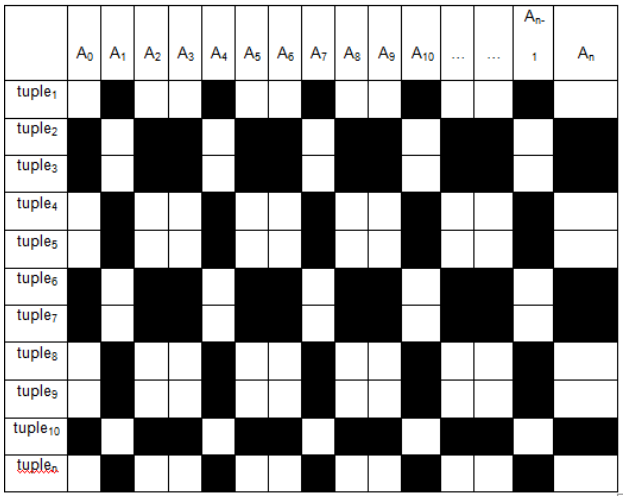

Figure 2. A snapshot of the watermarked database.

3.2 Watermark Bits Detection Procedure

The Watermark extraction procedure is blind. It requires neither the knowledge of the original un-watermarked database nor the watermark itself. This property is critical as it allows the watermark to be detected in a copy of the database relation, irrespective of later updates to the original relation. The watermark extraction procedure is a direct reversal of the watermark embedding procedure as described in the following steps:

Step 1: Locate the tuples of each sub-set in the database.
Step 2: Locate the non-numeric multi-word attribute of each tuple in the sub-set.
Step 3: In the selected attribute:
Find the word which has one of its characters expanded.
Count the number of the word starting from the begining.
Convert decimal equivalent of the count into a binary string.
Step 4: Repeat steps 2 and 3 for all tuples of the sub-set.
Step 5: Construct watermark by putting together extracted strings into an m x n image.
Step 6: Repeat steps 1 through 5 to extract all copies of the embedded watermark.

4. Experimental Results

The new database watermarking algorithm has been evaluated and tested on an experimental database that we have constructed. The database consists of 1000 tuples, and runs under the Oracle platform. We concentrated our performance evaluation on the robustness of the proposed algorithm by virtue of the fact that, database watermarking algorithms must be developed in such a way to make it difficult for an adversary to remove or alter the watermark beyond detection without destroying the value of the object. In particular, the database watermarking algorithm should make the watermarked database robust against the following types of attacks: subset deletion attack, subset addition attack, subset alteration attack, and finally subset selection attack.

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. The graph shown below in Figure 3 indicates that the watermark will be removed only, and only if, all the database was deleted! That is, removing more than 95 % of the database will not result in removing the watermark. This is due to the fact that the proposed algorithm embeds the same watermark everywhere in the database, making this type of attack ineffective.

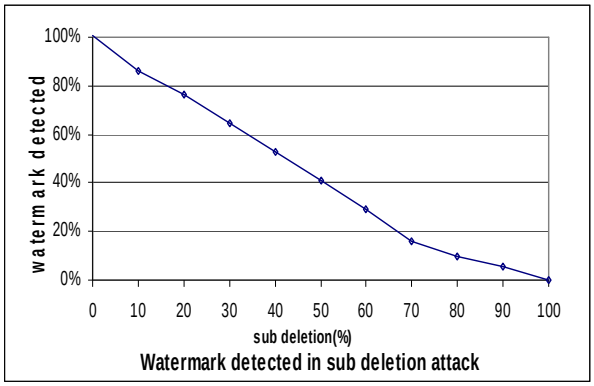

Figure3. Robustness results due to the 'subset deletion attack'

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. But this form of attack has little impact on the watermark embedded through our algorithm. The graph shown

ISSN 2320 -2602

International Journal of Advances in Computer Science and Technology (IJACST), Vol.3 , No.11, Pages : 30-34 (2014)
Special Issue of ICCECT 2014 - Held during 01-02 December, 2014,Bangkok, Thailand

below in Figure 4 indicates that the watermark will never be removed even if the added tuples are as many as the original tuples. That's, only the added tuples will not carry the watermark information.

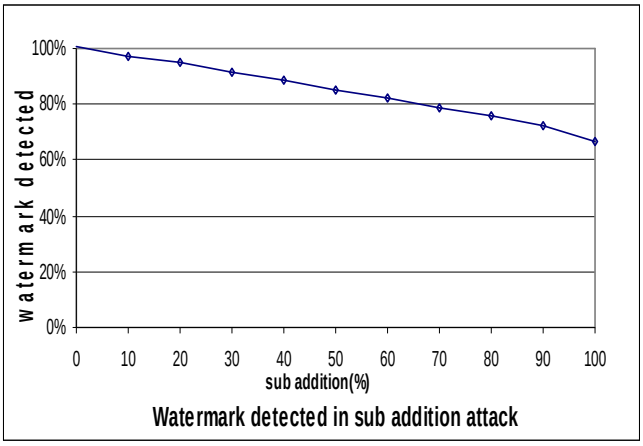

Figure 4. Robustness results due to the 'subset addition attack'.

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. The graph shown below in Figure 5 indicates that the watermark will remain even if 90 % of the tuples of the database were altered. This is due to the fact that the proposed algorithm embeds the same watermark everywhere in the database, making this type of attack ineffective.

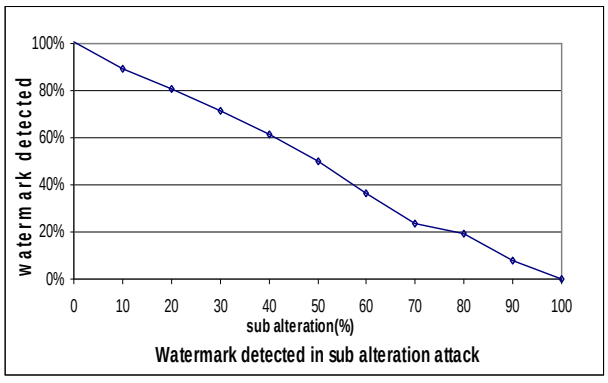

Figure 5. Robustness results due to the 'subset alteration attack'.

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. However, since the proposed algorithm embeds the watermark in the whole database, this attack is of little or no threat. The graph shown below in Figure 6 indicates that the watermark will remain even if the attacker selects a subset as small as 10% of the original database. That's no matter how the small subset he selects, the watermark will remain in the selected subset and thus providing the required copyright protection.

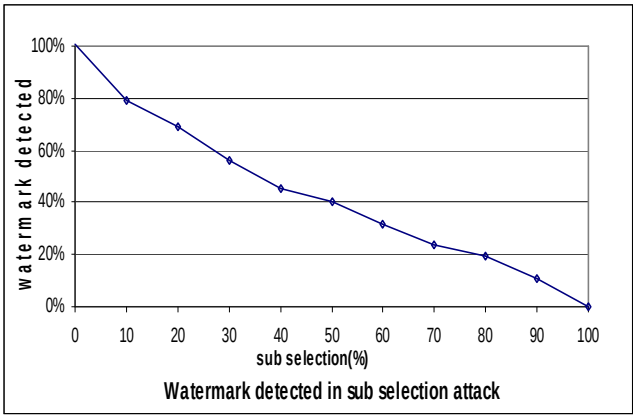

Figure 6. Robustness results due to the 'subset selection attack'.

5. CONCLUSION

In this paper we described a watermarking algorithm based on hiding watermark bits in the extensions of expandable Arabic characters of non-numeric, multi-word, attributes of subsets of tuples. A major advantage of using this approach is the large bit-capacity available to hide large watermarks. This is opposite to the other proposed algorithms where watermark bits have limited potential bit-locations that can be used to hide them effectively without being subjected to removal or destruction. The robustness of the proposed algorithm was verified against a number of database attacks such subset deletion, subset addition, subset alteration and subset selection attacks.

ISSN 2320 -2602

International Journal of Advances in Computer Science and Technology (IJACST), Vol.3 , No.11, Pages : 30-34 (2014)
Special Issue of ICCECT 2014 - Held during 01-02 December, 2014,Bangkok, Thailand

6. REFERENCES

Huang, M., Cao, J., Peng, Z., & Fang, Y. (2004). A new watermark mechanism for relational data. Presented at the 4th International Conference on Computer and Technology, China.

Li, Y., Swarup,V., & Jajodia, S. (2003). Constructing a Virtual Primary Key for Fingerprinting Relational Data. Presented at the 3rd ACM Workshop on Digital Rights Management, USA.

Li, Y., Swarup, V., & Jajodia, S. (2005). Fingerprinting Relational Databases: Schemes and Specialties. IEEE Trans. Dependable and secure Computing, 2(1), 34-45.

Li, Y. & Deng, R. (2006). Publicly Verifiable Ownership Protection of Relational Database. Presented at ASIACC, Taipei, Taiwan.

Sion, S., Atallah, M., & Prabhakar, S. (2003). Rights Protection for Relational Data. Paper Presented at the ACM International Conference on Management of Data, CA, USA.

Sion, S., Atallah, M., & Prabhakar, S. (2004a). Rights Protection for Relational Data. IEEE Trans. Knowledge and Data Engineering, 16(12), 912-926.

Sion, S., Atallah, M., & Prabhakar, S. (2004b). wmdb.*: Rights Protection for Numeric
Relational Data. Presented at the 20th International Conference on Data Engineering, USA.

Sion, S., Atallah, M., & Prabhakar, S. (2004c). Proving Ownership Over Categorical Data. Presented at the 20th International Conference on Data Engineering, USA.

Sion, S., Atallah, M., & Prabhakar, S. (2005). Rights Protection for Categorical Data. IEEE Trans. Knowledge and Data Engineering, 17(7), 1509-1525.

Silberschaty A and Korth H.F (2006), “Database System Concepts”, Fifth Edition, TMH

Zhang, Z., Jin, X., Wang, J., & Li, D. (2003). A Robust Watermarking Scheme for Relational Data. Presented at the 13th Workshop on Information Technology and Engineering, USA.

Zhang, Z., Jin. X., Wang, J., & Li, D. (2004). Watermarking Relational Database Using Image.