WATERMARKING MODEL IN DATABASE SYSTEMS

Tuan DoTrung
College of Science, Vietnam National University, Hanoi
Tao NguyenVan
University of Thai Nguyen

Abstract. The paper aims at a watermarking model in databases, for copyright management and embedding information in data. The model focuses on numeric data type and proposes solutions for the other multimedia data type, such as text, image and video.

Keywords : Model, watermark, database

1. Introduction

Watermarking process is studied in more than ten years, for the needs of copyright protect, data processing in the large data environment distributed in computer network. Watermarking is information embedding in data in keeping the data qualify and secret of embeded data [3, 4, 5, 6]. They apply watermarking techniques for (i) authoring of data; they base on embedding information in data in order to determine the owner of data; (ii) hiding information in data; they base on limited data modification, but not changing using qualify of data and not appearing embedded information.

Research results focused on the techniques on images, such as (i) Furie transformation of signals by fast Furie transform (FFT); discrete cosine transform (DCT); (ii) wavelet analysis; discrete wavelet transform (DWT); (iii) modification by singular value decomposition (SVD)... for embedding information on image data [4, 5]. The images have a lot of color levels, are represented under different compressed standards. The problem concerning watermarking on multimedia data have not studied yet, is proposed as a research one [1]. The working group of Institute of Information Technology, Vietnam National Institute on Technology and Sciences has obtained some results.

Watermarking process has characteristics [2]:

- Information quantity embedding in data assures that the modification of data is at permitted level. The level is in the accepted interval of data error. The information is so small that it does not influence to using quaify of data. Users can not differ attached information and may not take he information when not applying watermarking algorithm. Although, watermarked data allow users to copy data file in assuring the data owner. This character is called imperceptibility; the original data should be unnoticeable to the human observer, i.e. watermarks should not interfere with the protected media;

- Watermarked data are consistent, i.e. they keep these marks either in data manipulation operations or in attacks for taking watermarked information from data. This character is called robustness; unauthorized one should not be able to destroy the watermark without technique; watermark should be robust to common signal processing and intentional attacks;
- Watermarked information can be taken by watermark decomposition. It allows to obtain hidden information from data flow, is used in data security.

Concerning in watermarking are technique as follows :

- Embedding information into data;
- Detecting watermark in data;
- Watermark decomposition.

Watermarking research on image, sound in some years ago has some results, as in [5, 6]. However watermarking research in database systems was in beginning [2], proposed certain principal requirements about modeling and techniques. In [7] there is a hashing technique for the purpose.

The paper aims at a developing the technique proposed in [7], then presents a solution for integrating some watermarking techniques in a model allowing to watermark in multimedia data. When a particular multimedia data model does not existed yet, a relational database management system is used as replaced one. The solution in the paper focuses watermarking in (i) document; (ii) video data.

The rest of the paper : (i) the second part presents requirements and principal aspects of digital watermarking in the relational data model; (ii) the third part is proposed model and a solution for document and video data; (iii) the last part is conclusion and some remarks.

2. Watermarking in relational database systems

The relational data model allows to organize data by tuples, lines in two dimension table. Domain determines representation of attributes. Popular data types in the table are character and number. Some database management systems allow to use other types of data, as multimedia data (document, image, sound and video). Watermarking in relational data model is for copyright protection and information hiding.

Comparing the difficulties in watermarking process with conventional data type with multimedia data type, in [2] are some remarks (i) multimedia objects have a lot of data; there is data redundant. Users are not comfortable in selecting a location for information hiding; therefore a tuple is seen as an object in watermarking process; (ii) multimedia data modification is a difficult one because of complex structures of multimedia data; then decomposing hiding information from data is difficult. In relational data model, deleting a tuple may destroy watermark also.

Concerning digital data in a relational database system, [7, 9] proposed a watermarking algorithm. The algorithm bases on content characters of digital data [7] and metadata on these data [9].

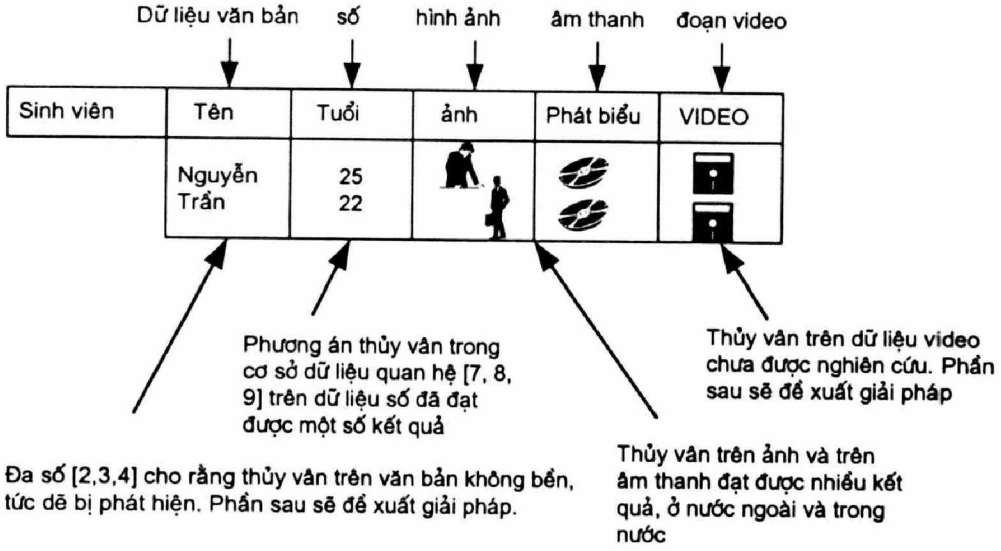

Figure 1. Watermarking with some kind of data type in relational data model

2.1. Algorithm for information embedding in data

The algorithm runs with the attribute Ai in which is watermarking. Embedding information affects to certain information bit among bits of binary representation of attribute value Ai. The process is repeated on tuples of the relational table.

Input to the relational table R(A1, A2,... An)

While not exists a tuple, do

If (random (0,1) < threshold a) AND (value Ai ≠ NULL) then

Transform (value Ai) to binary value

Select certain bit from the binary value

Attach selected bits into behind attribute value. The attribute is called watermarking attribute.

2.2. Algorithm for watermarking detecting

When suspecting some data attaching something wrong, they demand (i) to specify data having hidden information or not; (ii) decomposing hidden information. Because the hidden information does not affect data qualify in an application, it is not necessary to recover the original data. The watermark detection algorithm uses (i) threshold b, allowing to detect watermark; b is corresponding to threshold a; (ii) threshold c, allowing to refining experience values affecting to the threshold b.

With the suspicious table R (A1, A2,... An)

T1 = number of suspicious line

Init a counter T2 = 0

While exists a line of the table, do

If the value of Ai ≠ NULL, then

- *Transforming Ai value to binary value*
- *Extracting certain bit (as in the watermarking algorithm presented)*
- *If these bits matches the last part of attribute value of watermark attribute, let T2 = T2 + 1*
- *If T2/ T1 ≥ (a +c)/2 AND a > c then to inform "watermark exists"*

Else, if T2/ T1 ≥ c/ 2 AND a ≤ c then to inform □watermark exists□

Else to inform "watermark does not exist".

Two algorithms let us to determine the right owner of the data in the relational database. The model proposed in next part should develop the algorithms for digital watermarking on complex data type of database.

3. Model proposed

Below are some assumptions in the proposed model for watermarking in a database :

- Relational databases have tables R (A1, A2,... An), in which attributes accept number value, character value, sound, image or video. The table has N tuples (lines of table);
- Database composes of conventional data and image, sound and video data. This assumption allows watermarking is realized directly in data, but not in their representation on external files which do not belong to database;
- A video clip is decomposed into key frames. Some of them is demanded in image watermarking process; i.e. a problem of watermarking on video data is solved by watermarking on image;
- Concerning digital watermarking, some notations are (i) threshold a permits to select watermarking data; at cause of experiences, let the threshold a to set to approximate value of 0.25; (ii) threshold b permits to determine a watermarking tuple (line of table). With b = 0.12, i.e. 12 lines are selected among 100 lines. It corresponds to the probability to find a watermarking line is 0.12; (iii) a hashing function h(x) permits to select certain bit among the binary presentation of a number attribute;
- Knowledge on data, i.e. metadata is used to support information quantity hiding in data. The metadata is composed of characteristics about (i) attribute values of the relation (table of database); (ii) expectation value TB (average value); let to know information quantity hidden in data; (iii) information quantity TT, permits to know clarity level of attribute values

in corresponding to attribute semantic, i.e. the conception that the attribute has to present.

Watermarking in database should respond the effect as following :

- In secret; the users can not feel the existence of hiding information in data;
- Normal; data manipulation can realize normally;
- Robust; watermark information is consistent with data manipulation operations;
- Security; only authorized user can decompose hidden information from data;
- Watermarking detection permits authorized person to decompose watermarking information;
- Watermarking process allows updating operation on data;
- Watermarking communication; hidden information is transmitting in data flow, on data communication.

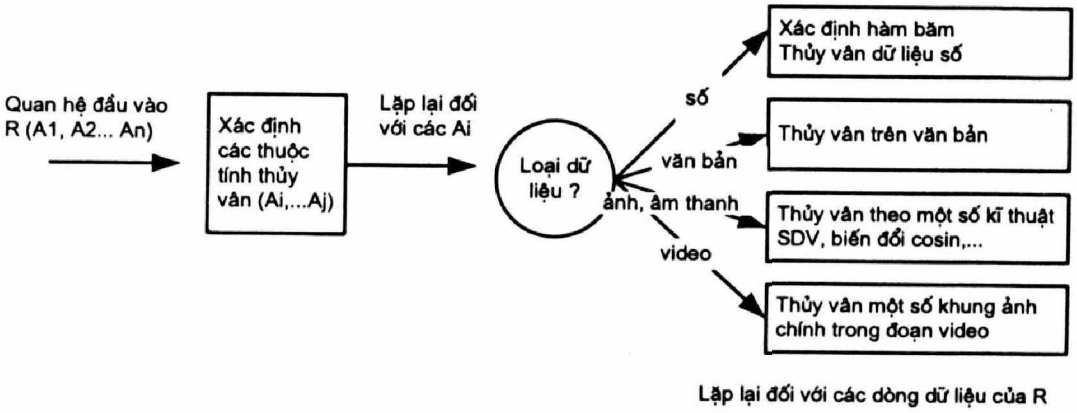

Figure 2. Watermarking model for different data types in database

3.1. Watermarking model

Watermarking process in database systems is realized after two levels : (i) watermarking for copyright protection; hidden information have not semantic content; (ii) watermarking for data authorizing and data hiding; hidden information is decomposed and it has semantic content; i.e. it has other using role.

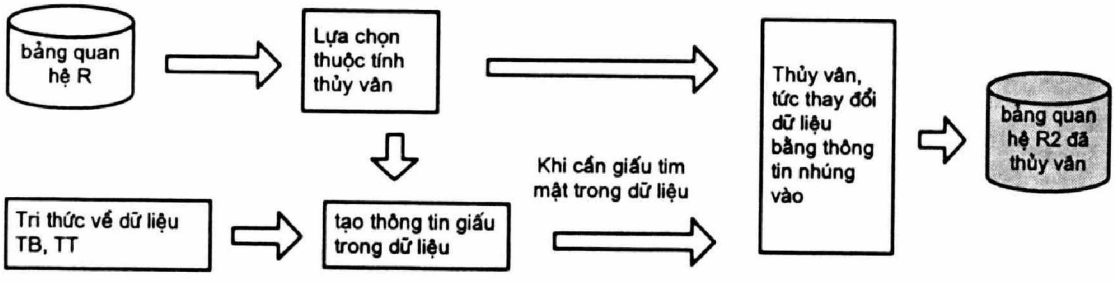

Figure 3. Watermarking in database with the relational one

After received data in database, users are informed that exists or not hidden information in these data. Watermark detection process bases on the knowledge about data, i.e. metadata via expectation value TB and information quantity TT.

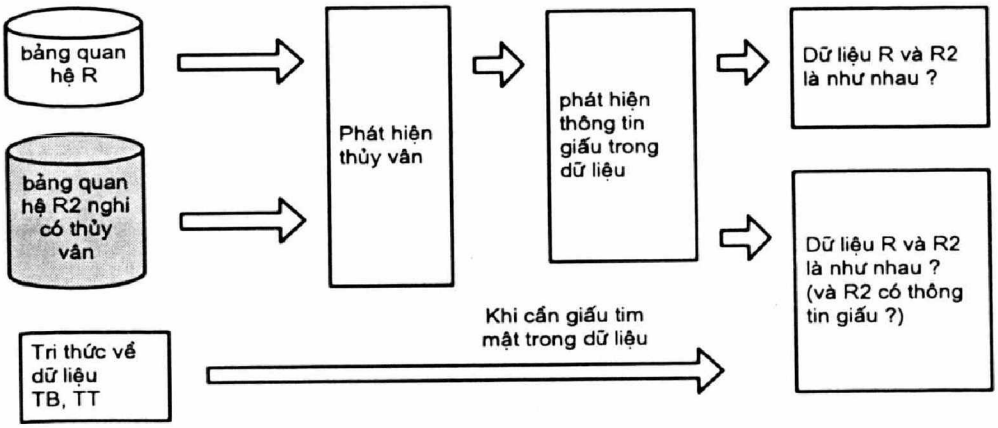

Figure 4. Watermark detection process

3.2. Watermarking on digital data

Embedding information into digital data demands some assumptions as follows:

- Database and knowledge about data (metadata), TB and TT, for embedding hidden information into data;
- Hashing function here is h(X, N, M) on bit string X; in which (i) X is binary presentation of an attribute value; N is number of the first selected; M is number of selected bit at the end of string of N bit. When there is not enough N bit form X, bit 0 can be inserted. M is a mark bit, is selected successively or after other selection manner;
- While watermarking with the attribute A1, select the next attribute A2 in the schema R as watermarking attribute. The selection is suitable with the requirement of only copyright protection. When needing to hide secret information, they use knowledge TB and TT in order to create a secret information. Data obtained with hidden information has to belong to the attribute domain; it assures that data have only accepted error. Hidden information is embedded in the attribute value A2;
- The threshold a, b and c used for watermarking and watermark detection are experience number. At the beginning, may be a=0.25, b=0.12 and c=0.12.
- The embedding function nh (Y, TBi, TTi) is for to create secret information. The function bases on the information Y the users want to embed into data. The knowledge TBi and TTi about the attribute value Ai are necessary to create a secret information corresponding to Ai value.

Input the table R(A1, A2,... An)

Input the knowledge about digital (TBi, TTi)

While exists a data tuple, do

If (random (0,1) < threshold a) AND (value Ai ≠ NULL) then

1. *Init the line counter k =1*
2. *Transform (value Ai) into the binary form X*
3. *N1 = length (X); N is created from N1; M =N/4;*
4. *Run the function h(X, N, M) having the mark TV*
5. *Determine a watermarking attribute Aj;*
6. *Marking the mark TV at the end of the attribute value Aj.*
7. *If they want to hide a secret information, then*

- *Generate a embedding information with the help of nh (Y, Tbi, Tti);*
- *Select a embedding information for the current data line; the line (tuple) is noted NHk;*
- *Insert NHk into the watermarking attribute value Aj.*

The watermark detection process uses the threshold c. Input of the process is original database, having the table R, and a suspicious database, having the table R2. Result of the process is yes/ no about the similar between two databases. When exists hidden information, let them to decompose the secret information.

Input the table R (A1, A2,... An)

Input the suspicious table R2 (A1, A2,... An)

Input the values N, M used in watermarking algorithm

Input the knowledge about data, TB, TT

Let T1 = number of suspicious data lines

Init the counter T2 = 0

While exists a line in the table R2, do

If value Ai ≠ NULL then

1. *Determine watermarking attribute Aj*
2. *Let X = binary form of the value Ai*
3. *Run the function (X, N, M), obtain TV*
4. *Compare TV with M last it of the attribute value Aj. If results is matching, then increase the counter T2 = T2 +1;*

If (T2/ T1 ≥ (a +c)/2 AND a > c) OR (T2/ T1 ≥ c/ 2 AND a ≤ c) then to inform "watermarking exists"

Else to inform "watermarking does not exist".

In the case of watermarking existing, they can decompose hidden information from the watermarking attribute.

3.3. Watermarking on image, sound and video

Basing on physical presentation of image and sound data, they have a bit matrix presentation, A. Watermarking process on these type of data have achieved a lot, permits to transform A into A'. In the database model, the method SVD [5] is proposed.

For video data, the model re-uses the technique of image watermarking, with only certain frames; i.e. with key frames of video clip.

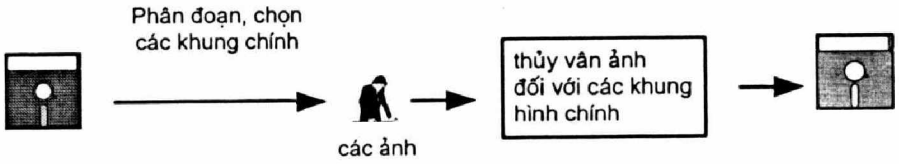

Figure 5. Re-use the image watermarking technique for video data

3.4. Watermarking on text

A large part of documents is with the presentation of characters. Very much effort on document watermarking, but there is not consistent and it is difficult to hide information. It is easy to find a wrong sentence in documents.

For the text type of data, the proposed model bases on knowledge of documents to generate (i) repeated characters; (ii) new characters. The location for watermarking in the documents is determined by a location function. A simple solution for the function is to select the fixed positions in the document. For example, positions in the document start at i, i+k, i+2k,... At the selected position, a new character is inserted or integrated into existing character.

4. Conclusion and remarks

The paper presents watermarking model in database systems. A relational database management is representative one. The model permits watermarking and watermark detection on different kinds of multimedia data. The image watermarking is worth, is presented in a lot of researches.

The threshold parameters, as a, b, c, N and M are experience parameters. In generally, with 1 millions data, they can chose appropriate parameters.

Watermarking model should be applied for database for training purpose. Then they need (i) parameters for watermarking; (ii) evaluation about consistency of watermarking process; (iii) watermarking pay-off. These remarks should be discussed in other paper.

Paper authors present their acknowledgement to the group on watermarking in Institute of Information Technology, Vietnam National Institute on Technology and Sciences for scientific materials.

References:

1. Tuan DoTrung, Cuong LuongXuan, Khun Piseth, Tao NguyenVan, About video data manipulation, *Natural sciences and technology Journal, VNU,* T.XIX, No.**3**(2003).
2. Agrawal R., Kiernan J., Watermarking relational databases, *Proc. of the 28th VLDB Conference*, 2002.
3. Mandal P., Thakral A., Verma S., Watermark based digital rights management, *Proc. of the int. conference on Information Technology : Coding and Computing,* 2005.
4. Paraskevi Bassia, Ioannis Pitas, Robust audio watermaking in the tine domain, *IEEE transaction on Multimedia*, Vol. 3, No. **2**(2001), p.232-241.
5. Ruizhen Liu, Tieniu Tan, An SDV-based watermaking scheme for protecting rightful ownership, *IEEE transaction on Multimedia*, Vol. 4, No.**1**(2002), p.121- 128.
6. Sanka Basu et. Al., Introduction to the special issue on multimedia database, *IEEE transaction on Multimedia*, Vol. 4, No.**2**(2002), p.141- 145.
7. Yong Zhang et al., Relational databases watermark technique based on content characteristic, *Proc. of the first Conference on Computing, Information and Control*, 2006.
8. Yong Zhang, Bian Yang, Xia-mu Niu, Reversible watermarking for reational database authentication, *Journal of Computers*, V. 17, N.**2**(2006), p. 59 - 66.
9. Yong Zhang, Xiamu Niu, Dongning Zhao, A method of protecting relational databases copyright with cloud watermark, *Transactions on engineering, computing and technology*, V.**3**(2004), p. 170 – 174.