34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Robust Lossless Watermarking based on Circular Interpretation of Bijective Transformations for the Protection of Medical Databases

J. Franco Contreras[1], G. Coatrieux[1], E. Chazard[2], F. Cuppens[3], N. Cuppens-Boulahia[3] and C. Roux[1]

***Abstract*— In this paper, we adapt the image lossless watermarking modulation proposed by De Vleeschouwer *et al.*, based on the circular interpretation of bijective modulations, to the protection of medical relational databases. Our scheme modulates the numerical attributes of the database. It is suited for either copyright protection, integrity control or traitor tracing, being robust to most common database attacks, such as the addition and removal of tuples and the modification of attributes' values. Conducted experiments on a medical database of inpatient hospital stay records illustrate the overall performance of our method and its suitability to the requirements of the medical domain.**

***Index Terms*— Watermarking, Medical Database, Health information security**

I. Introduction

Nowadays, medical databases are easily handled, accessed and shared by health professionals. They regroup electronic patient records for patient care and can support drug inventories, quarantine plannings as well as economical evaluation of medical activities and so on. At the same time, the more open the access to data, the more their security is compromised. Data can be extracted or rerouted from their final objective as well as modified. Derived from strict ethics and legislation, such as the HIPPA or the ISO/CEI 27001, one must ensure the security of medical data he or she possesses in terms of confidentiality, integrity and availability. Among existing protection mechanisms, watermarking has been recently proposed for the protection of databases.

The aim of watermarking is to allow the imperceptible embedding of a message into a host content without disrupting its normal use. Originally designed to offer copyright protection of digital multimedia content, watermarking has been successfully applied for other security objectives like in healthcare where more interest is given to data integrity and traceability [1].

Even though database watermarking is far from maturity, several methods have been proposed. We can distinguish them depending on the database information they modulate: numerical attributes, categorical attributes or tuple order. The last class of modulations refers to "distortion free" methods.

[1]J. Franco Contreras, G. Coatrieux and Christian Roux are with Institut TELECOM, TELECOM Bretagne, Brest 29238, France, and also with UMR1101 LaTIM, Brest 29609, France, {javier.francocontreras, gouenou.coatrieux}@telecom-bretagne.eu

[2]E. Chazard is with the Department of Medical Information and Archives, CHU Lille; UDSL EA 2694; Univ Lille Nord de France; F-59000 Lille, France

[3]F. Cuppens and N. Cuppens-Boulahia are with Institut TELECOM, TELECOM Bretagne, CS 17607, 35576 Cesson-Sévigné, France and also with UMR CNRS 3192 Labsticc

The first database watermarking method was proposed by Agrawal and Kiernan [2]. They mark secretely selected numerical attributes from secretely selected subsets of tuples by forcing some of their least significant bits to a specific value. The detection process stands then on the comparison of the number of correct matches between the extracted and the original sequences of bits with a decision threshold. Li *et al.* [3] extend this scheme, allowing to embed more than one bit. This extension makes it possible to embed a different sequence for each content user which has an interest for traitor tracing.

Categorical data differ from numerical data in the impossibility to modify or compare elements by means of mathematical operations. Considering for example the attribute "eye color", its value "blue" cannot be changed into "green" by adding a quantity δ. Sion *et al.* [4] overcome this issue by associating a numerical value to each possible categorical value, making possible the use of watermarking strategies for numerical attributes.

With the above methods, it is assumed that some distortion is tolerated for message embedding. In healthcare, such a distortion must be strictly controlled due to the fact that an undesired modification could cause a damage to patients or have economical consequences for the hospital. To our knowledge, only a few methods consider this issue. The Watermill method proposed by Lafaye *et al.* [5], searches for the maximum allowable distortion of a numerical attribute under aggregation query constraints, i.e. queries should return the same result before and after the watermarking process. In a different way, Shehab *el al.* [6], formulate the insertion process as a constrained optimisation problem and look for the appropiate watermark under statistical constraints.

Distortion-free solutions are also of interest. Li *et al.* [7] propose to embed the mark by switching or leaving unchanged the position of pairs of tuples according to the bit to embed. In [8], Guo extended this approach, by permutting both tuples and attributes. As defined, even though such a strategy preserves the values of the database attributes, embedded watermark will not resist common database update like the insertion or elimination of tuples, as well as the reorganization of the database.

In such a context, reversible watermarking provides a good alternative. The reversibility property allows the recovery of the original data from their watermarked version by inverting watermarking modifications. Reversible methods for database are very few and most are derived from strategies originally proposed for images. As an example, Farfoura *et al.* [9] presented a method which reversibly watermark

numerical data using the prediction-error-expansion modulation proposed by Alattar [10]. The first method adapted for categorical data has been proposed by Coatrieux *et al.* [11], who extend the histogram shifting modulation.

However, most if not all of the previous approaches are fragile. Embedded data will be lost after any database modifications, this is the reason why these methods are proposed for verifying the integrity of the database.

In this work, we present a new robust lossless watermarking method which modulates numerical attributes of relational database. It is based on the "Circular Interpretation of Bijective Transformations" modulation originally proposed by De Vleeschouwer *et al.* [12] for images. As we will show, our method is able to resist to most common manipulations like the insertion and elimination of tuples and the modification of tuple attribute values.

The rest of this paper is organized as follows. Section II presents the main steps of a common database watermarking system. We then describe our watermarking algorithm in section III. Before concluding, section IV gives and discusses the main results we obtained in terms of capacity and of robustness against different types of manipulations within the framework of an inpatient hospital record database.

II. Database Watermarking

Let us consider a relational database DB composed by a finite set of relations R_i. For the sake of simplicity, in the sequel we will consider a database with one single relation constitued of M unordered tuples $\{t_u\}_{u=1,...,M}$. In one relation, one tuple associates different attributes, each taking its value within a set of possible values. Herein, we will consider N attributes $\{A_1, A_2, ..., A_N\}$, $t_u.A_i$ corresponds to the value of the i^{th} attribute of the u^{th} tuple of the relation. In the table, each tuple t_u is uniquely identified by its primary key $t_u.PK$: one attribute or a set of attributes.

Most watermarking schemes proposed in the literature follow the embedding procedure depicted in figure 1. It starts by a clustering operation, creating non-overlapping groups of tuples. In this way, we ensure an independence between the watermark and the way the database is stored. In order to create N_g groups of tuples, a common strategy consists in deriving the group number of one tuple n_u from its primary key $t.PK$ combined with a cryptographic function H like the Secure Hash Algorithm (SHA), as shown in (1), where: $|$ represents the concatenation operator, K_W is the secret watermarking key.

$$n_u = H(K_W|H(K_W|t.PK))modN_g \quad (1)$$

The use of a cryptographic hash function allows the secure partitioning and the uniform distribution of the tuples into the groups, *i.e.* groups approximately contain the same number of tuples.

Once the groups are created, the watermark is embedded by modulating the values of one or several tuples' attributes or by changing the order of the tuples. Usually, one bit of message is embedded per group of tuples, spreading it over all the tuples and making it robust to database modifications.

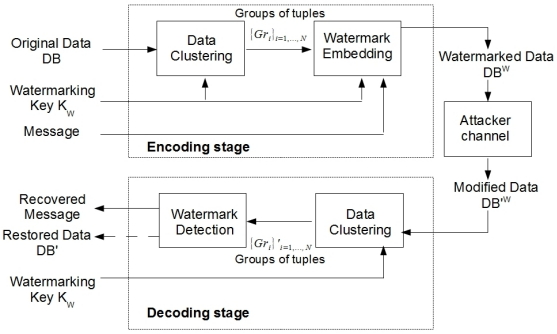

Fig. 1. Common embedding chain in database watermarking

The detection process works in a similar way. It re-identifies groups of tuples and next detects or extracts the watermark. Our algorithm follows the same strategy.

III. Proposed Method

The basic principle of the Circular Interpretation of Bijective Transformations modulation proposed by De Vleeschouwer *et al.* [12] consists in splitting a block of pixels into two groups and to modulate the difference between some of their features for reversibly embedding one bit into this pixel block. Herein, we propose to adapt their modulation to numerical attributes of groups of tuples in the following way. Let us assume that tuples of the database have been distributed in groups $\{G^i\}_{i=1,...,N_g}$ as exposed in section II. For one group G^i, we secretly assign each tuple to one subgroup: $G^{A,i}$ or $G^{B,i}$, making use of the rule in (2):

$$n_u = H(K_W|H(K_W|t.PK))mod2 \quad (2)$$

By next, if A_n is the attribute retained for the embedding, its histograms in each subgroup $G^{A,i}$ and $G^{B,i}$ are calculated and mapped into a circle as illustrated in figure 2a. The subgroup $G^{A,i}$ (resp. $G^{B,i}$) is then characterized by its center of mass $C^{A,i}$ (resp. $C^{B,i}$) to which we associate a vector $V^{A,i}$ (resp. $V^{B,i}$) (as illustrated in figure 2b). The embedding

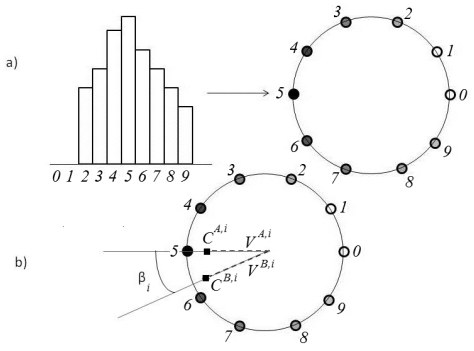

Fig. 2. a) Mapping process. For each bin of the histogram, a proportional virtual mass is placed in each position. b)Vectors associated to each center of mass and the angle β

process consists in modulating the relative angular position of vectors $V^{A,i}$ and $V^{B,i}$, *i.e* the value of the angle $\beta_i =$

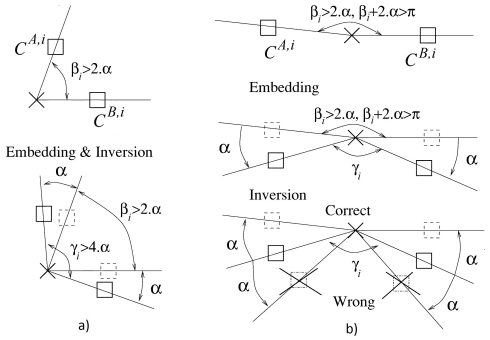

Fig. 3. Problematic Cases. a) $\beta_i > 2 * \alpha$, so α is too small for swapping vectors. Such a group is a "non-carrier". Attributes' values in the group are modified in order to increase the value of β_i into γ_i making possible for the reader to identify such non-carriers due to the fact that $\gamma_i > 4 * \alpha$. b) In this situation, the detector is not able to determine the original attributes' values in the groups after embedding, as they can come from two possible positions. The receiver needs additional information to recover original values.

$(\widehat{V^{A,i}, V^{B,i}})$. To do so, one bit $b = \{0, 1\}$ is embedded by modifying β_i according to (3)

$$\beta_i + 2 * \alpha \; if \; b = 0, \beta_i - 2 * \alpha \; if \; b = 1 \qquad (3)$$

where α is the amplitude of the angle variation. In our concern, α is such as $\alpha = 2\Delta * \pi / L$, where L is the number of bins in the attribute histogram and Δ the absolute distortion applied to each attribute considered for embedding. More precisely, modifying by α the angle β_i results in adding $(2b-1)\Delta$ to the attributes' values of $G^{A,i}$ or $G^{B,i}$.

It is important to notice that the groups where $\beta_i > 2 * \alpha$ cannot carry or cannot be watermarked with one bit of information, due to the fact that α is too small to swap $V^{A,i}$ and $V^{B,i}$. In order to avoid confusion between those groups, we refer as "non-carriers", with the previous ones, we refer as "carriers", at the reading stage as well as to allow watermark removal, β_i is increased by $2 * \alpha$. This procedure is depicted in figure 3a [12]. Blocks for which β_i exceeds $\pi - 2 * \alpha$ cannot carry information neither (see figure 3b). Moreover, it is not possible to find a rule that will permit the reader to recognize these groups. As a consequence, in order to inform the reader about their presence and position, some additional information should be embedded along with the message in order to allow the reader to restore the original database.

At the reading stage, groups and subgroups of tuples are reconstructed. For each group, the reader extracts the embedded bit according to the relative angular position of vectors (*i.e.* γ_i) and reconstructs the original database. Indeed, the sign of γ_i indicates the direction of rotation in order to reverse the embedding process. The magnitude of γ_i allows to detect non-carrier groups, being those with $\gamma_i > 4 * \alpha$.

Our scheme can be made robust or fragile. The latter case relies on the insertion of a sequence of bits, those of the message one wants to embed. As an example, one can insert a digital signature of the database for the purpose of protecting its integrity. At the reception, the recipient just has to compare the extracted signature to the one recomputed on the restored database. Any difference will indicate the database integrity is lost. Any other metadata such as the origin of the database or a short description only available to those who possess the watermarking key could also be inserted.

In order to make the previous scheme robust, we suggest the embedding of a secret pseudo-random sequence of symbols $S = \{s_i\}_{i=1...M}$, with $s_i \in \{+1, -1\}$, and to consider the following rules for carrier groups.

$$\beta_i + 2 * \alpha \; if \; s_i = -1, \beta_i - 2 * \alpha \; if \; s_i = +1 \qquad (4)$$

For non-carrier groups, insertion follows the same rules as before.

At the detection stage, our purpose is to detect the pseudo-random sequence by means of a correlation measure between the original and the extracted sequences. Notice that because of the presence of non-carrier groups, the embedded sequence will always be slightly different from the original one.

IV. Results

Performance analysis of our algorithm in terms of capacity and robustness was conducted on a real medical database containing pieces of information related to inpatient stays at the hospital. For sake of simplicity, we have only considered one relation of 10248 tuples (see sample view given in figure I). In this table, each tuple is represented by four attributes: stay identifier ("*id_stay*"), patient identifier ("*id_patient*"), patient age and the expected stay duration of the patient in the medical unit ("*d_exp*"). The attribute "*id_stay*" was considered as the primary key and used for tuple groups and subgroups constitution.

TABLE I
SAMPLE VIEW OF OUR MEDICAL DATABASE

id_stay	id_patient	age	d_exp
601433878	7892	29,44	13,50
601484325	28653	40,15	31,00
601527723	14552	65,87	4,78

A. Capacity Performance

In the fragile scheme, the embedding capacity is directly related to the number of carrier groups (see section III), *i.e.* groups for which $\beta_i < 2 * \alpha$. The capacity varies with the attribute probability distribution and in particular with its standard deviation σ_{A_n}. Notice that in each case we fixed the value of Δ (or equivalently of α) equals to the square root of the standard deviation which empirically gives a good trade-off between distortion and capacity. From figure 4a, it appears that, for a table with a fix number of tuples, the capacity is influenced by the size of the groups (equivalently by the number of groups). Furthermore, for an attribute with a high standard deviation, *e.g* "*age*", the increase of the number of groups does not improve the capacity as much as for an attribute whose values are concentrated around its mean, *e.g.* "*d_exp*".

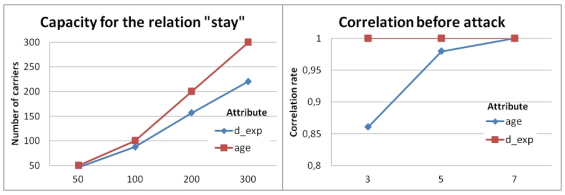

Fig. 4. a) Left side - Capacity or equivalently the number of carriers per number of groups for the attributes in the relation stay. b) Right side - Correlation measure before any attack.

B. Robustness

We tested the robustness of our scheme against different attacks: modification of attribute' values, elimination or insertion of tuples; with several degrees of intensity expressed in terms of percentage of modified tuples: 10%, 20%, 40% and 60%. For the "modification attack", a percentage of tuples is randomly selected and their attributes' values are changed. The "elimination attack" corresponds to a deletion of randomly selected tuples while the addition attack consists in the insertion of some tuples with random attributes' values but following the same distribution as in the original database. These "new" tuples are uniformly distributed among the different groups.

In this experiment, 50 groups and several values of Δ were considered. In figure 4b, we can see that the correlation measure between the secret pseudo-random sequence and the embedded one without any attack depends again on the probability density of the attribute selected for the embedding. Indeed, the correlation values obtained with the attribute "*d_exp*" are greater than with the attribute "*age*" of higher standard deviation.

Figure 5 illustrates the performance of our algorithm under different attacks for both attributes "*age*" and "*d_exp*". Indicated measures correspond to the ratio between the original correlation measure (*i.e.* before the attacks) and the correlation after the attack. As it can be seen, the robustness of our scheme depends on the value Δ and on the statistical properties of the attribute. If the standard deviation σ_{A_n} of the attribute A_n is small, our scheme will be robust even for small values of Δ. This make us think that the compromise between robustness and imperceptibility can be derived from the standard deviation of the attribute.

V. Conclusion

In this paper we have proposed a new robust and reversible database watermarking method for numerical data based on a circular representation of a bijective transformation. With our scheme one can insert a sequence of bits for verifying the database integrity or identify one person who rerouted it. Our scheme demonstrates also good performance against most common database manipulations. Future work will focus on increasing embedding capacity and on minimizing distortion introduced within the database. Indeed, there is an interest not having to remove the watermark before using the database, this will allow a continuous protection and interoperability with systems that are not watermarking compliant.

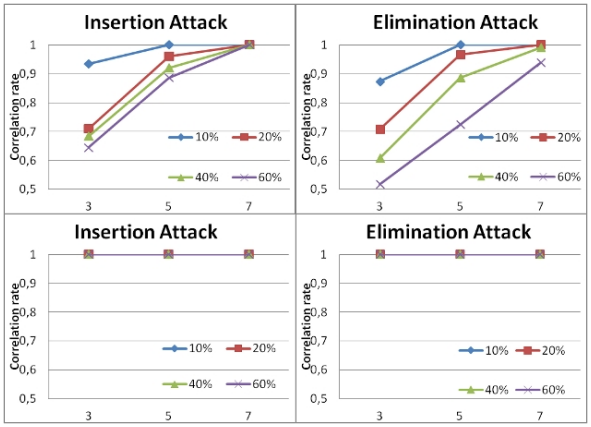

Fig. 5. Correlation measure for different values of Δ. Different attack intensities were considered. Upper results correspond to the attribute "*age*", whereas lower ones to the attribute "*d_exp*". From the figure we observe a high detection rate even in the case of a severe attack. Moreover, the selection of the correct value of Δ according to the attribute distribution can increase robustness

References

[1] G. Coatrieux, H. Maitre, B. Sankur, Y. Rolland, and R. Collorec, "Relevance of watermarking in medical imaging," in *Information Technology Applications in Biomedicine, 2000. Proceedings. 2000 IEEE EMBS International Conference on*, 2000, pp. 250 –255.

[2] R. Agrawal and J. Kiernan, "Chapter 15 - watermarking relational databases," in *VLDB '02: Proceedings of the 28th International Conference on Very Large Databases*, pp. 155 – 166. Morgan Kaufmann, San Francisco, 2002.

[3] Y. Li, V. Swarup, and S. Jajodia, "Fingerprinting relational databases: schemes and specialties," *Dependable and Secure Computing, IEEE Trans. on*, vol. 2, no. 1, pp. 34 –45, jan.-march 2005.

[4] R. Sion, "Proving ownership over categorical data," in *Data Engineering, 2004. Proceedings. 20th International Conference on*, march-2 april 2004, pp. 584 – 595.

[5] J. Lafaye, D. Gross-Amblard, C. Constantin, and M. Guerrouani, "Watermill: An optimized fingerprinting system for databases under constraints," *IEEE Trans. on Knowledge and Data Engineering*, vol. 20, pp. 532–546, Apr. 2008.

[6] M. Shehab, E. Bertino, and A. Ghafoor, "Watermarking relational databases using optimization-based techniques," *IEEE Trans. on Knowledge and Data Engineering*, vol. 20, pp. 116–129, 2008.

[7] Y. Li, H. Guo, and S. Jajodia, "Tamper detection and localization for categorical data using fragile watermarks," in *Proceedings of the 4th ACM workshop on Digital rights management*, New York, NY, USA, 2004, DRM '04, pp. 73–82, ACM.

[8] J. Guo, "Fragile watermarking scheme for tamper detection of relational database," in *Computer and Management (CAMAN), 2011 International Conference on*, may 2011, pp. 1 –4.

[9] M.E. Farfoura and S-J. Horng, "A novel blind reversible method for watermarking relational databases," in *Parallel and Distributed Processing with Applications (ISPA), 2010 International Symposium on*, sept. 2010, pp. 563 –569.

[10] A.M. Alattar, "Reversible watermark using the difference expansion of a generalized integer transform," *Image Processing, IEEE Trans. on*, vol. 13, no. 8, pp. 1147 –1156, aug. 2004.

[11] G. Coatrieux, E. Chazard, R. Beuscart, and C. Roux, "Lossless watermarking of categorical attributes for verifying medical data base integrity," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society,EMBC*. Sept. 2011, pp. 8195–8198, IEEE.

[12] C. De Vleeschouwer, J.-F. Delaigle, and B. Macq, "Circular interpretation of bijective transformations in lossless watermarking for media asset management," *Multimedia, IEEE Trans. on*, vol. 5, no. 1, pp. 97 – 105, march 2003.