Ontology-Guided Distortion Control for Robust-Lossless Database Watermarking: Application to Inpatient Hospital Stay Records

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

***Abstract*— In this paper, we propose a new semantic distortion control method for database watermarking. It is based on the identification of the semantic links that exist in-between attribute's values in tuples by means of an ontology. Such a database distortion control provides the capability for any watermarking scheme to avoid incoherent records and consequently ensures: i) the normal interpretation of watermarked data, i.e. introducing a watermark semantically imperceptible; ii) prevent the identification by an attacker of watermarked tuples. The solution we present herein successfully combines this semantic distortion control method with a robust lossless watermarking scheme. Experimental results conducted on a medical database of more than one half million of inpatient hospital stay records also show a non-negligible gain of performance in terms of robustness and database distortion.**

***Index Terms*— Watermarking, Medical Databases, Ontology**

I. Introduction

Nowadays, health professionals can easily share and access distant medical databases for different tasks ranging from remote and collaborative diagnosis to economical evaluation of medical activities and so on [1]. At the same time, this ease of access increases security risks. Data records may be redistributed or modified without authorization. The number of reported data leaks each year is still considerable, even in sensitive domain like healthcare [2]. Existing security mechanisms such as access control or encryption avoid non-authorized users to access information but once these mechanisms are bypassed or more simply when the access is granted, data are no longer protected.

In this context, watermarking appears as a complementary security mechanism that provides an "*a posteriori*" protection: data can be accessed while still being protected [3]. Basically, it consists in the embedding of a message or watermark into a multimedia host document (e.g. image or database) based on the principle of controlled distortion. Originally designed for copyright protection of digital multimedia content, watermarking has been successfully applied for other security objectives like integrity and traceability.

Database watermarking was introduced by Agrawal *et al*. [4]. Since 2002, several methods have been proposed [5], [6]. Among them, we can distinguish "attribute-distortion-based" methods that modify or modulate database attributes' values for message embedding from methods said as "attribute-distortion-free". Methods from the latter category play on the order of tuples within a relation for message insertion rather than modifying the attributes' values [7]. By doing so, they make the watermark dependent on the way the database is stored while inducing constrains on the database management system. Regarding methods of the former category, their authors assume some attribute value distortion can be carried out without perturbing any *a posteriori* uses of data.

In order to better take into account watermark imperceptibility in "attribute-distortion-based" methods, two different approaches have been considered. A first class of methods is based on the adjustment of the watermark under some distortion constraints. For instance, Shehab *et al.* consider attribute statistics constraints (e.g. mean, standard deviation) and adapt the watermark amplitude by means of optimization techniques [6]. In a recent work [8], Kamran *et al*. go one step further and propose to preserve data-mining process classification results. In [9], Lafaye *et al.* consider query-result constraints and look at preserving the response to *a priori* known aggregation queries and modulate pairs of tuples in consequence. The second class refers to reversible or lossless watermarking methods [10], [11]. The reversible property ensures that database attributes' values modulated for message embedding can be recovered by inverting the introduced distortion. Because the watermark can be removed, distortion constraints are alleviated, and the watermark can be updated without introducing more distortion. Nevertheless, for these methods too, there is an interest to control the induced distortion. Indeed, keeping the watermark into the database maintains it "*a posteriori*" protected. In [11], we proposed an original robust lossless watermarking scheme that reversibly modulates the relative angle of the center of mass of circular histograms associated to groups of values of one numerical attribute in a relation.

As exposed, all the above methods focus on preserving statistical properties of the database but they do not consider strong semantic links that exist in-between attributes of the database, semantic links that should be preserved as well. Indeed, after the embedding process, database tuples must remain semantically coherent in order to: 1) ensure the correct interpretation of the information contained in the database; 2) keep the watermark invisible to the eyes of an attacker. The previous methods fail to do so. In order to overcome this issue, we propose an ontology-guided distortion control method that allows the embedder to identify the limits of distortion for one numerical attribute in a tuple. Basically, an ontology is attached to a specific area of knowledge

[1]J. Franco-Contreras and G. Coatrieux are with Institut Mines-TELECOM, TELECOM Bretagne, Unité INSERM 1101 LaTIM, France {javier.francocontreras, gouenou.coatrieux}@telecom-bretagne.eu

[2]N. Cuppens-Boulahia and F. Cuppens are with Institut Mines-TELECOM, TELECOM Bretagne, UMR CNRS 3192 Lab-STICC, France

[3]C.Roux is with Institut Mines-TELECOM, Mines Saint-Étienne, France

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

and allows defining shared concepts and their relationships by means of a common vocabulary. Ontologies have been successfully applied in several domains for data extraction [12] and image annotation [13] as example, but to our knowledge, they have not been yet proposed to semantically control watermarking distortion. The information contained in an ontology can be helpful to distinguish attributes more or less prone to be watermarked. We demonstrate it in this work by integrating our semantic distortion control into the robust lossless scheme we proposed in [11]. Our method avoids the introduction of incoherent information in the database while not diminishing the robustness of this scheme against tuple insertion and deletion attacks, on the contrary.

The rest of this paper is organized as follows. In Section II, we present the main steps of a common chain of database watermarking as well as the basic principles of the robust lossless scheme we exploit, before explaining how ontologies can be used in order to control the database distortion in Section III. We provide some experimental results and point out the interest of our proposal in Sect. IV. Section V concludes this paper.

II. Database Watermarking

A. A general database watermarking chain

Formally, a database is a collection of data organized into a finite set of relations $\{R_i\}_{i=1,\ldots,N_R}$. For sake of simplicity, we will consider one database composed of one single relation constituted of N unordered tuples $\{t_u\}_{u=1,\ldots,N}$, each of M attributes $\{A_1,\ldots,A_M\}$, with $t_u.A_n$ refering to the value of the n^{th} attribute of the u^{th} tuple. Notice that the attribute A_n takes its values within a set called attribute domain. Futhermore, each tuple is uniquely identified by either one attribute or a set of attributes, we call its primary key $t_u.PK$.

The main stages considered in the majority of database watermarking schemes are presented in Fig.1. On the message embedding side, a pretreatment is first applied so as to make the watermark insertion/extraction independent from the database storage structure. Ordinarily, it consists in a group construction operation that creates a set of N_g non-intersecting groups of tuples $\{G^i\}_{i=1,\ldots,N_g}$.

Typically, the group number of one tuple is obtained from the result of a cryptographic hash function applied to its primary key $t_u.PK$, concatenated with a secret watermarking key K_S as exposed in (1) where '|' represents the concatenation operator [5], [6]. The use of a cryptographic hash function, e.g. Secure Hash Algorithm (SHA), ensures the secure and equal distribution of tuples into groups.

$$n_u = H(K_S|H(K_S|t_u.PK))modN_g \quad (1)$$

Then, if N is the total number of tuples in the database, each group will approximately contains $\frac{N}{N_g}$ tuples. By next, one bit or symbol of the message is embedded per group. To do so, the values of one or several attributes are modified accordingly to the retained watermarking modulation. Thus, one may expect to embed a message corresponding to a sequence of N_g symbols $S = \{s_i\}_{i=1,\ldots,N_g}$.

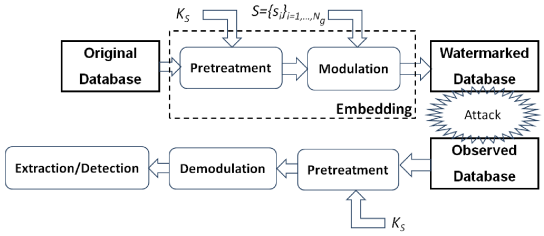

Fig. 1. A common database watermarking chain.

Watermark extraction works in a similar way. First, tuples are distributed into N_g groups. Depending on the watermarking modulation, one symbol is extracted from each of these groups. If tuple primary keys are not modified, the knowledge of the watermarking key ensures synchronization between watermark embedding and extraction stages.

B. Lossless Circular histogram center of mass modulation

In order to demonstrate the interest of the semantic distortion control method we detail in section III, we combined it with the robust lossless scheme we proposed in [11]. This scheme embeds one symbol of the sequence S per group of tuples. To do so, it modulates the relative angular position of the circular histogram center of mass of one numerical attribute. Due to space limitation, we sum-up its basic principles in the case S is a sequence of bits. We invite the reader to consult [11] for more details. Thus, let us assume the attribute A_t, the domain of which corresponds to the integer range $[0, L-1]$, is retained for embedding. Let us also consider the group of tuples G^i in which we are going to embed one bit $s_i = \{0, 1\}$. This task is conducted by equally dividing G^i into two sub-groups $G^{A,i}$ and $G^{B,i}$ following the same strategy exposed in sect. II-A and (1). Then and as illustrated in Fig.2, the histograms of A_t in each subgroup are calculated and mapped onto the unit circle. In order to embed s_i, the relative angle $\beta_i = (\widehat{V^{A,i}, V^{B,i}}) \simeq 0$ between the vectors $V^{A,i}$ and $V^{B,i}$ associated to the histograms' centers of mass is modulated. Depending on the value of s_i, the circular histograms of $G^{A,i}$ and $G^{B,i}$ are rotated in opposite directions with an angle step $\alpha = \frac{2\pi\Delta}{L}$, where Δ corresponds to the shift amplitude of the histogram. More precisely, modifying the angle β_i of $2\alpha(2s_i - 1)$ results in adding $(2s_i - 1)\Delta$ to the attributes of $G^{A,i}$ and $(1 - 2s_i)\Delta$ to those of $G^{B,i}$. At the reading stage, the sign of the watermarked angle β_i^w indicates the embedded symbol in G^i as well as the direction of rotation to follow so as to invert the insertion process. Notice that because α is a fixed value, not all of the groups of tuples can convey one bit of message, see [11] on how handling such a situation.

In the case this scheme is used so as to identify the owner or the recipient of the database, the detection will have to verify that the sequence S (an user identifier) is present in the database even if this one has been can be "attacked" (e.g. tuple addition/suppression). This can be achieved by means of a correlation measure $C_S =< S, \hat{S} >$, where $\hat{S}$ is the extracted sequence. If C_S is greater than a decision threshold Tr_S then S is said to be present in the database.

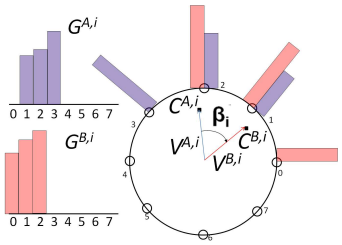

Fig. 2. Histogram mapping of each sub-group $G^{A,i}$ and $G^{B,i}$ onto a circle. The angle between the vectors pointing the centers of mass is modulated in order to embed one symbol of the message.

III. Ontology-Guided Distortion Control

A relational database aims at providing efficient storage and rapid access to large amounts of data. However, it lacks of semantic information about the meaning and links between different attributes' values in a tuple. An ontology can be herein useful offering additional semantic information about the database content. For an area of knowledge, an ontology provides a common vocabulary and define, with different levels of formality, the meaning of the terms and the relations between them [14]. It is composed of concepts, which represent objects or sets of objects within a domain. Concepts are linked by means of relations that specify hierarchical or associative interactions between them.

From this standpoint, each domain value, subset or range of values of an attribute A_t can be associated to one ontology concept. We depict in Fig. 3 such a mapping considering the following example. Let us consider a tuple with attributes "diagnosis", "age", ... The value "alzheimer" in the domain of the attribute "diagnosis" can be associated to a concept "alzheimer" in a medical ontology. This concept is related to another concept "more than 60 years old", which can be mapped into a range of possible values for the attribute "age". From a watermarking point of view, this semantic relations make us aware that one attribute age value should not be turned into a value smaller than 60 in a tuple where the "diagnosis" attribute value is "alzheimer". As exemplified, the value of the attribute A_t in the u^{th} tuple, i.e. $t_u.A_t$, semantically depends on the set $S_{t_u.A_t}$ of values of the other attributes of t_u, i.e. $t_u.\{A_i\}_{i=1,\ldots,M;i\neq t}$, or a subset of them.

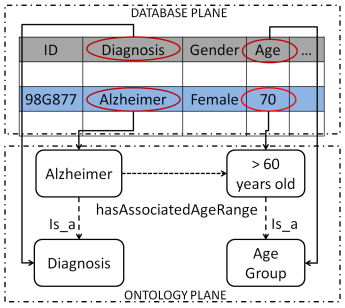

Fig. 3. Existing connection between a relational database and an ontology. Dotted and dashed arrows represent ontological relations between concepts in the ontology. Solid arrows represent connections between attributes or attributes values and ontological concepts.

As a consequence, we propose to use the concepts and relations of an ontology associated to the database in order to identify the maximum tolerated distortion of its attributes' values. For a numerical attribute A_t, this distortion limit in the tuple t_u corresponds to the range of allowable values $t_u.A_t$ can take: $Rg_{t_u.A_t}$, under the semantic constraints of $S_{t_u.A_t}$. If we come back to the previous example, A_t ="age" is an integer the value of which in a tuple $t_u.age$ belongs to an integer range $Rg_{t_u.age}$ imposed by the set $S_{t_u.age}$ ="alzheimer". In a more general way, if the attribute domain of A_t corresponds to the integer range $[A_{t,min}, A_{t,max}]$, the range $Rg_{t_u.A_t}$ can be defined as the union of N_{rg} different ranges such as: $Rg_{t_u.A_t} = [A_{tmin}, A_{tmax,1}] \bigcup \cdots \bigcup [A_{tmin,N_{rg}}, A_{tmax}]$; set of ranges than can be identified from the ontology by querying it considering the other attributes' values in t_u, i.e. $S_{t_u.A_t}$. The knowledge of $Rg_{t_u.A_t}$ will be used as reference to guide the watermark embedding process.

It is important to notice that the semantic distorsion control we propose is complementary to any other statistical distortion control method. They can be used simultaneously.

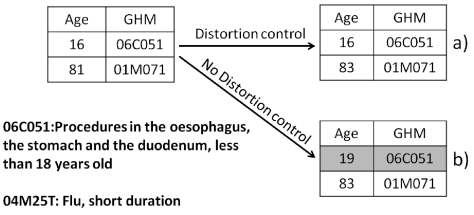

Fig. 4. Example of modification of two tuples taking or not into account semantic distortion limits. Semantically incorrect tuples are highlighted.

IV. Experimental Results

In this section, we evaluate the influence of our semantic distortion control over the performance of the scheme presented in [11] in terms of robustness against tuple insertion and deletion attacks. In order to maintain the reversibility of this scheme, only tuples free of semantic constraints were used for message embedding.

The following experiments have been conducted on a test database constituted of one relation of 536200 tuples issued from one real medical database containing pieces of information related to inpatient stays in French hospitals. In this table, each tuple associates fifteen attributes like the hospital identifier (*id_hospital*), the patient stay identifier (*id_stay*), the patient age (*age*), the attribute GHM (patient homogeneous group) and several other useful data for statistical analysis of hospital activities. *id_stay* and *age* are numerical attributes while GHM is a categorical attribute. GHM is the French equivalent of the the Diagnosis-Related Groups (DRG) of the Medicare system in the USA. Its attribute domain consists in a list of codes intended for treatment classification and reimbursement. A GHM code results from a function that takes as input the patient age, the ICD10 principal and associated diagnosis and several others element we can not detail herein due to space limitation. In this experiment, for sake of simplicity, we summed up the domain ontology to the relations between the attribute

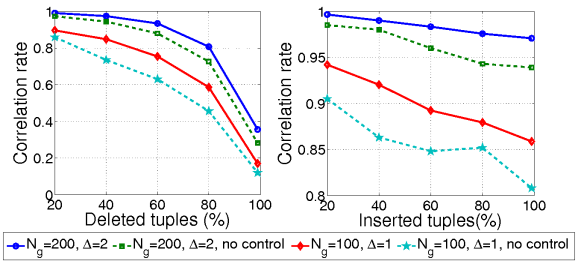

Fig. 5. Correlation rate for the attribute *Age* with $\Delta = 1, 2$ taking $N_g = 100$ and 200 groups with and without distortion control in the case of: a tuple deletion attack (left); a tuple insertion attack (right). Solid and dashed lines represent the results with and without semantic distortion control respectively.

GHM and age. Notice that in our implementation, the domain ontology was implemented in Protégé [15] and queried by means of the SPARQL query language. In order to constitute the groups of tuples (see Sect.II), the attributes *id_stay* and *id_hospital* were considered as the primary key. The attribute *age* was selected for message embedding.

A. Illustrative example of ontology interest

An example presenting the advantage of controlling semantically the database distortion by means of an ontology is given in Fig. 4. This latter shows an extract of the original database with only two tuples and the corresponding watermarked database extracts with and without semantic distortion constraints, i.e. tables a) and b) respectively. As it can be seen, taking into account the ontology avoids the apparition of incoherent tuples. Indeed, the GHM code $06C051$ corresponds to patients younger than 18 years old, if this constraint is satisfied in table a), this is not the case in table b) where the watermarked age value is 19 (see the shaded tuple). Such an incoherent value makes the tuple suspect to an attacker and can perturb the normal interpretation of data in a subsequent data-mining process.

B. Comparative Robustness Results

In this section, we evaluate the influence of our semantic distortion control onto the performance of the robust lossless scheme [11]. As depicted in Section II-B, the basic principle of this scheme stands in the detection of the presence in the database of a specific binary watermark by means of correlation. As a consequence, the robustness of our scheme can be established through the correlation measure for a given watermark. Results we provide here are given in average after 30 random simulations with the same parameterization but different group configuration and different watermarks.

Experiments were conducted with two attribute shift amplitudes $\Delta = 1, 2$ and two distinct numbers of groups $N_g = 100, 200$. The resulting watermarked databases were then attacked by means of tuple deletion or addition, the intensity of which is measured in percentage of tuples. As illustrated in Fig. 5, for the same parameterization, our method not only preserves the robustness of the considered scheme but it also increases its performance whatever the considered attack.

V. Conclusion

In this paper, we have proposed a database semantic distortion control which is ontology-guided. Our proposal aims at achieving two objectives: i) ensure the semantic of the database is preserved; ii) make the watermarking invisible to the attacker's eyes. To our knowledge, no other method has been proposed considering this issue. Furthermore, it can advantageously complete current statistical distortion control schemes. We have shown that, combined with a common robust lossless watermarking scheme, such a semantic distortion control reinforces or at least preserves its performance.

Acknowledgment

This work was supported by the French Armaments Procurement Agency (DGA) - Projet DGA RAPID FRAG&TAG - and by the Brittany Region Council. The authors are very grateful to the Department of Medical Information and Archives, CHU Lille; UDSL EA 2694; Univ Lille Nord de France; F-59000 Lille, France, for the experimental data used in this study.

References

[1] C. Quantin, D.O. Jaquet-Chiffelle, G. Coatrieux, E. Benzenine, and F.A. Allart, "Medical record search engines, using pseudonymised patient identity: An alternative to centralised medical records," *Int. Journal of Medical Informatics*, vol. 80, no. 2, pp. 6–11, 2011.

[2] M. McNickle, "Top 10 data security breaches in 2012.," In Healthcare Finance News. Accessed: 2014-03-27.

[3] G. Coatrieux, W. Puentes, L. Lecornu, C.C. Le Rest, and C. Roux, "Compliant secured specialized electronic patient record platform," in *Proc. Int. Conf. D2H2*, 2006, pp. 156–159.

[4] R. Agrawal and J. Kiernan, "Chapter 15 - watermarking relational databases," in *Proceedings of the 28th Int.Conf. on Very Large Databases*, pp. 155 – 166. 2002.

[5] R. Sion, M. Atallah, and S. Prabhakar, "Rights protection for relational data," *IEEE Trans. on Knowledge and Data Engineering*, vol. 16, no. 12, pp. 1509 – 1525, 2004.

[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] S. Bhattacharya and A. Cortesi, "A distortion free watermark framework for relational databases," in *Int.Conf. on Software and Data Technologies, Volume 2*, 2009, pp. 229–234.

[8] M. Kamran and M. Farooq, "A formal usability constraints model for watermarking of outsourced datasets," *Information Forensics and Security, IEEE Transactions on*, vol. 8, no. 6, pp. 1061–1072, 2013.

[9] 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, 2008.

[10] G. Coatrieux, E. Chazard, R. Beuscart, and C. Roux, "Lossless watermarking of categorical attributes for verifying medical data base integrity," in *Proc. of IEEE Int. Conf.of the Eng. in Med. and Biol. Society*, 2011, pp. 8195–8198.

[11] J. Franco-Contreras, G. Coatrieux, N. Cuppens-Boulahia, F. Cuppens, and C. Roux, "Robust lossless watermarking of relational databases based on circular histogram modulation," *IEEE Trans. on Information Forensics and Security*, vol. 9, no. 3, pp. 397–410, 2014.

[12] W. Su, J. Wang, and F. H. Lochovsky, "Ode: Ontology-assisted data extraction," *ACM Trans. Database Syst.*, vol. 34, no. 2, pp. 12:1–12:35, 2009.

[13] L. Hollink, G. Schreiber, J. Wielemaker, and B. Wielinga, "Semantic annotation of image collections," in *In proc. of Workshop on Knowledge Markup and Semantic Annotation*, 2003, pp. 0–3.

[14] A. Gomez-Perez and V. Benjamins, "Applications of ontologies and problem-solving methods," *AI Magazine*, vol. 20, no. 1, pp. 119–123, 1999.

[15] "Protégé ontology editor," protege.stanford.edu, 2013-12-09.